České vysoké učení technické v Praze

Fakulta jaderná a fyzikálně inženýrská

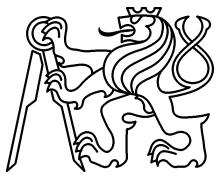
Akademie věd České republiky

Ústav informatiky a výpočetní techniky

Úvod do teorie neuronových sítí

Ing. František Hakl, CSc.

Ing. RNDr. Martin Holeňa, CSc.



1997

Ediční středisko ČVUT Praha

Předkládaná skripta shrnují látku prezentovanou v rámci přednášky „Úvod do teorie neuronových sítí“, která je určena pro studenty 5. ročníku matematického inženýrství na FJFI ČVUT Praha. Cílem této přednášky je seznámit posluchače se základními teoretickými poznatky a postupy, které jsou nezbytné k rigoróznímu studiu a efektivnímu použití paraelních výpočtů, označovaných jako „umělé neuronové sítě“.

Skriptum je rozděleno do dvou samostatných částí, z nichž každá byla psána s rozdílným přístupem k prezentaci učební látky. Přestože tyto dvě části na sebe navazují pouze volně, vzájemně jedna druhou vhodně doplňují a poskytují tak čtenáři možnost jak detailního studia, tak i možnost získání širokého přehledu.

První část skript (autor František Hakl) je psána pokud možno autonomně, tedy obsahuje důkazy veškerých lemm a tvrzení, s výjimkou několika široce známých tvrzení, se kterými byli posluchači 5. ročníku MI FJFI seznámeni již v průběhu svého studia, nebo jsou uvedeny v běžně dostupné literatuře.

Výklad v první části není zaměřen na detailní sumarizaci veškerých modelů neuronových sítí a jejich derivátů, ani na popis oblastí jejich aplikovatelnosti. Místo toho autor klade důraz na obecný popis paraelního zpracování informace neuronovými sítěmi, s cílem naznačit teoretické principy, na kterých je paraelismus neuronových sítí založen, se snahou zdůraznit souvislosti této teorie s poznatky standardních matematických disciplín. Důkladné zvládnutí těchto principů v prezentovaném rozsahu by mělo být postačující pro analýzu i některých jiných modelů neuronových sítí, které nejsou zde zmiňovány. Jednotlivé konkrétní modely neuronových sítí jsou uvedeny především na podporu a ilustraci vykládaných teoretických principů. Pokud bude mít student zájem o získání hlubšího přehledu vykládané látky, může využít literatury uvedené na konci skript.

Z důvodů omezeného rozsahu skript zcela opomíjíme v první části přístupy k neuronovým sítím založené na statistických metodách, teorii fuzzy množin, atd. Stejně tak vynecháváme výklad cyklických neuronových sítí, neuronových sítí odvozených od genetických algoritmů, a jiných výpočetních modelů, které s pojmem neuronových sítí nějak souvisí.

Druhá část skript (autor Martin Holeňa) má za cíl stručně sumarizovat současné poznatky o různých architekturách neuronových sítí a jejich dynamickém chování. Vzhledem k rozsahu předkládaných informací neobsahuje tato část důkazy jednotlivých tvrzení, eventuelně je obsahuje pouze v náznakové formě. Oproti předcházející první části se čtenář může seznámit s více modely neuronových sítí, pozornost je věnována souvislosti mezi neuronovými sítěmi a teorií fuzzy množin, dále pak rozvedení statistického pohledu na problematiku neuronových sítí.

Dovolte mi závěrem na tomto místě poděkovat slečně Ivaně Baranovské za přepis některých částí textu a celkovou jazykovou korekturu první části. Dále bych chtěl ocenit ochotu pana Arnošta Štedrého, který mi pomohl vyřešit mnohá úskalí a taje systémů TEX a LATEX 2ε, na která jsem při práci na těchto skriptech narazil.

Případné připomínky k textu, podněty či doporučení, směrujte prosím na níže uvedenou e-mailovou adresu.

V Praze, 14. prosince 1998 František Hakl

e-mail: hakl@uivt.cas.cz

Klasifikace moderních věd:

Je-li to zelené, nebo se to hýbe, patří to do biologie.
Páchne-li to, patří to do chemie.
Nefunguje-li to, patří to do fyziky.
Zní-li to jako nesmysl, patří to do ekonomie.
Je-li to nesrozumitelné, patří to do matematiky.

Obsah

Seznam obrázků

Část I

Teoretické základy umělých neuronových sítí

(autor František Hakl)

Věnováno památce mých rodičů.

Kapitola 1

Úvodní kapitola

Edingtonova teorie

Počet nejrůznějších hypotéz,
snažících se osvětlit určitý biologický jev,
je nepřímo úměrný dostupným vědomostem.

V této úvodní kapitole nastíníme základní předmět, kterému jsou tato skripta věnována. Velmi krátce pohovoříme obecně o základním popisu umělých neuronových sítí, uvedeme některé vybrané modely a jejich základní charakteristiku.

1.1 Použité značení

V celém textu budeme používat následujícího značení a používání fontů:

$\stackrel{\text{def}}{=}$	defice nově zavedeného pojmu, množiny či čísla
$K = \{1, \ldots, m\}$	množina přirozených čísel od 1 do m
B_α^N	koule v prostoru $\Re^N$ se středem v počátku a poloměrem α
$\Re^N$	reálný N-rozměrný prostor
Ω^m	m-násobný kartézský součin množiny Ω
$\breve{z}$	prvek v kartézském (vícenásobném) součinu dané množiny
$\bar{X}$	obecná (pod)množina v dané množině
X	systém podmnožin dané množiny
$\mathcal{X}$	množina systémů množin nad nějakou obecně danou množinou
$2^{\bar{X}}$	potenční množina množiny $\bar{X}$, $2^{\bar{X}} \stackrel{\text{def}}{=} \left\{ \bar{Z} \,\middle|\, \bar{Z} \subset \bar{X} \right\}$
$[\Omega]_\lambda$	lineární obal množiny Ω vektorů vektorového prostoru (=množina všech lineárních kombinací konečného počtu prvků z Ω)
$[\bar{B}]_\kappa$	konvexní obal množiny $\bar{B}$
$\mathcal{H}_n^*$	vektorový prostor (se specifikovanými parametry)
$C_{\bar{A}}$	prostor funkcí spojitých na dané množině $\bar{A}$
$\mathbf{NP}_k^\alpha$	množina objektů s danými vlastnostmi, zpravidla množina funkcí, zobrazení, atd., dle kontextu
B^*	obecný učicí algoritmus

$\bar{A} \triangle \bar{B}$	symetrická diference množin $\bar{A}$ a $\bar{B}$
$\bar{A} \dot{-} \bar{B}$	rozdíl množin $\bar{A}$ a $\bar{B}$
$o\left(\widetilde{f}\right)$	$\widetilde{g} = o\left(\widetilde{f}\right)$ vyjadřuje skutečnost, že funkce $\widetilde{g}$ je na asymptotickém okolí nuly zhora omezena v absolutní hodnotě funkcí $\widetilde{f}$
$O\left(\widetilde{f}\right)$	$\widetilde{g} = O\left(\widetilde{f}\right)$ vyjadřuje skutečnost, že funkce $\widetilde{g}$ je na asymptotickém okolí nekonečna zhora omezena v absolutní hodnotě funkcí $\widetilde{f}$
$\boldsymbol{A}$	matice reálných čísel, sloupce či řádky budeme označovat společným názvem ŘADY MATICE
$\vec{\boldsymbol{x}}$	vektor daného vektorového prostoru (vektory vždy budeme považovat za sloupce)
$\boldsymbol{A}^{(n)}$	čtvercová matice řádu 2^n
$\vec{\boldsymbol{x}}^{(n)}$	vektor dimenze 2^n
$\langle \vec{\boldsymbol{x}} \,\vert \vec{\boldsymbol{y}} \rangle$	skalární součin vektorů, $\langle \vec{\boldsymbol{x}} \,\vert \vec{\boldsymbol{y}} \rangle = \sum_{i=1}^{n} \vec{\boldsymbol{x}}_i \vec{\boldsymbol{y}}_i$
$\vec{\boldsymbol{x}} \odot \vec{\boldsymbol{y}}$	tensorový součin vektorů či matic
$\overset{\wedge}{\text{AND}}$	binární logická funkce AND
$\overset{\vee}{\text{OR}}$	binární logická funkce OR
$\overset{\sqcup}{\text{XOR}}$	binární logická funkce XOR
$\overset{\neg}{\text{NOT}}$	unární logická funkce NOT
$\lfloor M \rfloor$	největší celé číslo menší než M (=celá dolní část)
$\lceil M \rceil$	nejmenší celé číslo větší než M (=celá horní část)
$j^\flat$	$j^\flat$ představuje vektor z $\{-1,+1\}^n$, (resp. $\{0,1\}^n$, podle kontextu), jehož souřadnice korespondují s binárním zápisem čísla j (souřadnice vektoru $j^\flat$ odpovídající pozici 1 v binárním zápisu j, jsou rovny 1).
$\vec{\boldsymbol{v}}^\sharp$	celé číslo, odvozené od vektoru $\vec{\boldsymbol{v}} \in \{-1,+1\}^n$, (resp. $\vec{\boldsymbol{v}} \in \{0,1\}^n$, podle kontextu), kde jedničkové složky vektoru $\vec{\boldsymbol{v}}^\sharp$ odpovídají 1 v binárním zápisu čísla $\vec{\boldsymbol{v}}^\sharp$, ostatní složky odpovídají pozicím 0 v tomto zápisu
$\binom{a}{i}$	binomický koeficient
$\widetilde{f}$	obecné zobrazení mezi dvěma množinami
$\left(\widetilde{f} * \widetilde{g}\right)$	konvoluce funkcí $\widetilde{f}$ a $\widetilde{g}$
$\widetilde{A}\,(b)$	hodnota funkce (zobrazení) $\widetilde{A}$ v argumentu b
$\log_e(x)$	přirozený logaritmus
$\log_2(x)$	logaritmus o základu 2
$x \equiv y \mod r$	$x \equiv y \mod r$ vyjadřuje, že číslo x je dělitelné číslem y se zbytkem r (děleno moduló r)
$Prob_{\widetilde{\Pi}}\left(\bar{A}\right)$	pravděpodobnost množiny $\bar{A}$ při hustotě pravděpodobnosti $\widetilde{\Pi}$
$\left\vert \bar{V} \right\vert$	mohutnost množiny $\bar{V}$
$\vert\delta\vert$	absolutní hodnota z čísla δ
$\left\Vert \vec{\boldsymbol{z}} \right\Vert_{max}$	maximová norma vektoru
$\left\Vert \vec{\boldsymbol{z}} \right\Vert_{E}$	Eukleidovská norma vektoru
$\left\Vert \vec{\boldsymbol{z}} \right\Vert_{\mathcal{L}_2^{\bar{K}}}$	specifikovaná norma vektoru

Uveďme několik poznámek k použité terminologii. Nově definované pojmy jsou zvý-

razněny v textu použitím jiného fontu, např. NOVĚ ZAVEDENÝ POJEM . Co se týče české terminologie, v mnohých případech se jedná pouze o autorův překlad původního anglického termínu. Přestože snahou autora bylo těmito překlady vystihnout co nejpřesněji daný pojem (při co možná nejmenším pokřivení českého jazyka), je možné předpokládat nesoulad námi užívané terminologie s terminologií používanou jinými českými autory. V případech, kdy je tento nesoulad velmi pravděpodobný, uvádíme v příslušných definicích i původní anglický termím, např. (anglický termín: PAC learning) , mimo jiné i proto, že ani anglická terminologie není ještě plně ustálena.

1.2 Motivace pro výzkum umělých neuronových sítí

Obecně lze chápat každý počítač jako zařízení, které zobrazuje množinu vstupních dat (např. koeficienty lineární soustavy) na množinu dat jiných (např. řešení této soustavy). Nebudeme-li na takovýto počítač klást další zužující podmínky, lze konstatovat, že i nervový systém individuálního živočicha je jakýmsi prototypem (každý je jiný) počítače, a mnohou duševní činnost či vegetativní aktivitu živočicha lze považovat za realizaci jakéhosi výpočtu (např. více či méně smysluplná odpověď na více či méně smysluplnou otázku, reakce na daný podnět prostředí, atd.). Mezi těmito dvěma paralelními světy „živých počítačů“ a klasických počítačů elektronických existují však zásadní rozdíly. Pokusme se sumarizovat ty z nich, o nichž se domníváme, že mají vztah k principu jejich činnosti:

klasický počítač	**nervový systém**
• existence centrálního procesoru provádějícího veškerou činnost	• takovýto centrální procesor nebyl v nervových tkáních dosud objeven a nic nepotvrzuje jeho existenci
• nízká paralelizace, vyjímečně více procesorů ($\leq$ 1000) komunikujících mezi sebou	• velmi vysoký počet jednodušších (z hlediska přenosu informace) procesorů-neuronů ($\simeq 10^{12}$), velmi vysoká hustota propojení mezi nimi (10^3 pro každý neuron)
• přesná znalost způsobu vyhodnocení a zpracování informace (stroje typu RAM - Random Acces Machine, RASP - Random Acces Stored Program, Thuringův stroj)	• velmi matná představa o činnosti jednotlivých elementů, téměř žádná představa o způsobu komunikace mezi elementy
• nutnost detailní algoritmizace výpočtů, podložená hlubokým teoretickým pozadím (např. nutnost existence poznatků lineární algebry pro řešení soustav lineárních rovnic)	• zdá se, že schopnost vyhodnocování různých situací je integrální vlastností nervových systémů bez potřeby pochopení a uvědomění si způsobu zpracování informace
• velmi vysoká rychlost početních elementárních operací (dnes $\simeq$ 300 MHz)	• pomalý přenos informace (řádu milisekund)
• přesně definovaná architektura procesoru	• velký počet lokálních elementů vzájemně propojených, s velkou variabilitou hustoty a způsobu spojení.

Výše uvedené rozdíly ospravedlňují názor, že schopnosti živých organismů efektivně zpracovávat informace ze svého okolí (v mnohých ohledech a případech výrazně efektivněji než nejvýkonnější počítače) jsou dány především způsobem komunikace mezi jednotlivými neurony, a že tyto neurony lze považovat za jakési elementární, principem shodné, stavební kameny nervových center.

Dá se očekávat, že analýza struktur, které jsou velmi jednoduchým modelem nervových center živočichů (na základě dosud známých poznatků z oblasti anatomie, biochemie a neurologie), může být přínosem jak pro výzkum živých organismů, tak i opačně – pro výzkum a implementaci nových typů vysoce paralelních počítačových architektur, které umožní řešit problémy, pro které na počítačích odvozených od principu Thuringova stroje neexistuje efektivní algoritmizace.

V následující podkapitole zavedeme základní modely umělých neuronových sítí a popíšeme stručně jejich dynamické chování.

1.2.1 Základní popis architektury a dynamiky neuronových sítí

Nyní popišme obecnou architekturu a dynamiku neuronových sítí za účelem definování předmětu analýzy pro následující kapitoly těchto skript.

Neuronovou síť budeme chápat jako systém sestávající se z jakési *architektury*, *dynamiky* a *úlohy*.

Úlohou budeme rozumět zobrazení, které by měla daná neuronová síť realizovat mezi množinou svých vstupů (= vstupní prostor) a množinou svých výstupů (= výstupní prostor). Interpretace této úlohy může být rozličná, nejčastěji hovoříme o separaci tříd, kdy vstupní prostor je tvořen množinou $\bar{A} \in \Re^n$, vystupní prostor je množina $\{-1, +1\}$, a naším cílem je, aby výstupní hodnota neuronové sítě byla rovna charakteristické funkci množiny $\bar{A}$. Další častá úloha je dána jako problém aproximace. Pro dané zobrazení $\widetilde{Z} : \Re^n \to \Re^m$ požadujeme, aby výstupy z neuronové sítě co nejlépe aproximovaly zobrazení $\widetilde{Z}$ v nějaké metrice prostoru $\Re^m$. Poslední úlohu, kterou zde zmíníme, je problém interpolace a/nebo extrapolace (= též predikce), kdy pro zadané zobrazení $\widetilde{Z} : \Re^n \to \Re^m$ a množiny $\bar{A} \in \Re^n$, $\bar{B} \in \Re^n$ požadujeme, aby síť na základě množiny $\widetilde{Z}\left(\bar{A}\right)$ interpolovala či extrapolovala množinu $\widetilde{Z}\left(\bar{B}\right)$.

Poznamenejme, že nejčastější způsob zadání množin $\bar{A}$ a $\bar{B}$ je konečná množina jejich prvků a nejčastější způsob zadání zobrazení je dán jako konečná množina uspořádaných dvojic $\left\{\vec{\boldsymbol{x}}, \widetilde{Z}\left(\vec{\boldsymbol{x}}\right)\right\}$.

Architektura neuronové sítě je daná orientovaným grafem, jehož uzly jsou tvořeny neurony a který popisuje způsob propojení jednotlivých neuronů, dále pak popisem tzv. *přenosové funkce* neuronů. Množiny uzlů (= neuronů), do nichž nevede hrana, nazveme vstupními neurony, ty uzly, z nichž nevede hrana, nazveme výstupními neurony, ostatní uzly nazveme vnitřní (někdy též skryté) neurony. Podle druhu použitého grafu je zvykem rozlišovat následující základní typy architektur:

- dvouvrstvá neuronová síť tvořená bipartitním grafem
- neuronová síť tvořená úplným grafem
- vícevrstvá neuronová síť, kdy množina uzlů je rozdělena do disjunktivních množin $U_1, \cdots, U_n$ a hrany grafu jsou pouze mezi množinami U_i a U_{i+1}, $i \in \{1, \ldots, n-1\}$.

Přenosová funkce neuronu je zobrazení $\widetilde{P} : \bar{W} \times \bar{X} \to \Re$, kde $\bar{X}$ je množina všech možných vstupních vektorů (odpovídajících vstupním vrcholům) a $\bar{W}$ je parametrický prostor daného neuronu v grafu sítě. Každému vektoru z tohoto prostoru budeme říkat váhový vektor neuronu. Nejčastější přenosová funkce neuronu je rovna funkci

$$\widetilde{y}\left(\vec{\boldsymbol{x}}\right) \stackrel{\text{def}}{=} \widetilde{\sigma}\left(\langle \vec{\boldsymbol{w}} \,|\vec{\boldsymbol{x}}\rangle - t\right),$$

kde $\vec{\boldsymbol{w}}$ je vektor tvořený jednotlivými váhami, t je prahová hodnota daného neuronu, $\widetilde{y}\left(\vec{\boldsymbol{x}}\right)$ je výstupní hodnota neuronu, $\vec{\boldsymbol{x}}$ je vstupní vektor neuronu. $\widetilde{\sigma}$ bývá nejčastěji buď binární (tzv. tvrdá nelinearita):

$$\widetilde{\sigma}\left(z\right) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{pro} \quad x \geq 0 \\ -1 & \qquad\; x < 0 \end{cases},$$

nebo tzv. sigmoidální funkce.

Definice 1.2.1 *Nechť $\widetilde{\sigma}$ je reálná funkce reálné proměnné, monotónně rostoucí v celém definičním oboru $\Re$, vyhovující podmínkám*

$$\lim_{z \to +\infty} \widetilde{\sigma}\left(z\right) = 1, \quad \lim_{z \to -\infty} \widetilde{\sigma}\left(z\right) = -1.$$

Potom takovouto funkci nazveme SIGMOIDÁLNÍ FUNKCE .

Dynamikou neuronové sítě budeme rozumět změny parametrů přenosových funkcí neuronů (eventuelně i samotného grafu sítě) v čase. Dynamika sestává ze dvou částí

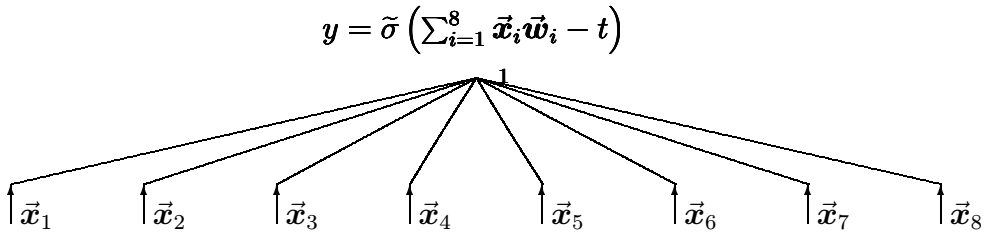
- aktivní dynamika (proces vybavování), kdy na vstupní neurony vložíme vstupní hodnoty sítě, tyto vstupní hodnoty po hranách probíhají k dalším neuronům, přes ně přechází modifikované podle jejich přenosových funkcí k dalším neuronům, až se modifikované hodnoty objeví na výstupních neuronech, kde jsou odečítány a interpretovány jako výstupní hodnoty sítě.

- adaptivní dynamika (proces učení), při které se hledají a nastavují váhy spojnic (= hran grafu) jednotlivých neuronů a ostatní parametry přenosových funkcí jednotlivých neuronů, eventuelně se upravuje celý graf tak, aby síť řešila danou konkrétní úlohu.

Nejdůležitější přístupy k adaptivní dynamice jsou založeny na minimalizaci hodnoty celkové chybové funkce

$$\min \left\{ \sum_{i=1}^{t} \left\| \widetilde{N}\left(\langle \vec{\boldsymbol{x}}_t \,|\vec{\boldsymbol{w}}\rangle\right) - y_t \right\|^2 \Big| \vec{\boldsymbol{w}} \in \bar{W} \right\},$$

kde $\{(\vec{\boldsymbol{x}}_1, y_1), \cdots, (\vec{\boldsymbol{x}}_t, y_t)\}$ je úloha, $\widetilde{N}\left(\langle \vec{\boldsymbol{x}} \,|\vec{\boldsymbol{w}}\rangle\right)$ je zobrazení realizované neuronovou sítí, $\vec{\boldsymbol{w}}$ je vektor všech vah v síti. Tato minimalizace se provádí nejčastěji gradientními metodami, např. učení metodou zpětného šíření (anglický termín: back-propagation) , v mnoha případech ale také různými heuristickými algoritmy, ze kterých zmíníme významnější v následujících příkladech.



Obrázek 1.1: Schéma architektury perceptronu.

1.2.2 Příklady základních neuronových sítí

Jako prvotní zdroj informací o základních modelech neuronových sítí, jejich dynamickém chování a aplikacích lze využít např. [ŠJ96]. Mezi nejznámější modely neuronových sítí patří následující dva, které jsou patrně nejlépe prozkoumány z teoretického hlediska a z hlediska praktických aplikací nejvíce rozšířené.

Perceptron

Perceptron je dvouvrstvá síť, kde výstupní vrstva je tvořena právě jedním neuronem, spojeným se všemi neurony vstupní vrstvy. Výstupní hodnota perceptronu je rovna

$$y \stackrel{\text{def}}{=} \widetilde{f}\left(\sum_{i=1}^{n} \vec{x}_i \vec{w}_i - t\right),$$

kde dvojice $\vec{w}, t$ představuje parametry přenosové funkce výstupního neuronu a $\widetilde{f}$ je buď tzv. tvrdá nelinearita nebo libovolná sigmoidální funkce. Pro danou úlohu

$$\{(\vec{x}_1, y_1), \cdots, (\vec{x}_t, y_t)\}$$

definujme proces učení jako iterační algoritmus popsaný vztahem

$$\vec{w}_{i+1} \stackrel{\text{def}}{=} \vec{w}_i + y\left(y - \vec{w}_i^T \vec{x}\right)\vec{x}, \tag{1.1}$$

pro nějaký výchozí vektor vah $\vec{w}_0$. Vlastnosti této posloupnosti budeme studovat v části 6.1.

Vrstevnaté sítě, učení metodou back propagation

Velmi často používaným typem neuronové sítě je vrstevnatý model, kdy neurony jsou rozděleny do vrstev a každý neuron v dané vrstvě je spojen se všemi neurony ve vrstvě předcházející a ve vrstvě následující (tak, že příslušný graf nemá smyčky, hrany vedou pouze mezi vrstvami).

Předpokládejme přechodovou funkci ve tvaru 1.1, kde ale funkce $\widetilde{f}$ je spojitá, dvakrát spojitě diferencovatelná, sigmoidální funkce.

Předpokládejme, že máme opět zadanou úlohu $\{(\vec{x}_1, \vec{y}_1), \cdots, (\vec{x}_t, \vec{y}_t)\}$, kde $\vec{x}_i \in \Re^n$, $\vec{y}_i \in \Re^m$, $i \in \{1, \ldots, t\}$, a že pomocí gradientní metody největšího spádu minimalizujeme hodnotu chybové funkce

$$\widetilde{E}(\vec{w}) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{i=1}^{t} \left\| \widetilde{N}(\vec{x}_i, \vec{w}) - \vec{y}_i \right\|_E^2.$$

Abychom pochopili princip metody zpětného šíření, vyslovíme následující lemmu, která dává explicitní vyjádření derivace vícenásobně zanořené složené funkce.

Lemma 1.2.1 *Předpokládejme, že pro pevné přirozené k je dána posloupnost reálných čísel $w_1, \ldots, w_k$, dále posloupnost $\widetilde{\sigma}_1, \ldots, \widetilde{\sigma}_k$ reálných diferencovatelných funkcí jedné reálné proměnné, a nechť reálná posloupnost $y_1, \ldots, y_{k+1}$ splňuje rekurzivní vztah*

$$y_{i+1} \stackrel{def}{=} \widetilde{\sigma_i}\,(w_i y_i), \quad i \in \{1, \ldots, k\}.$$

Potom platí, že

$$\frac{\partial y_{k+1}}{\partial w_k} = \widetilde{\sigma_k'}\,(w_k y_k) \cdot y_k$$

a

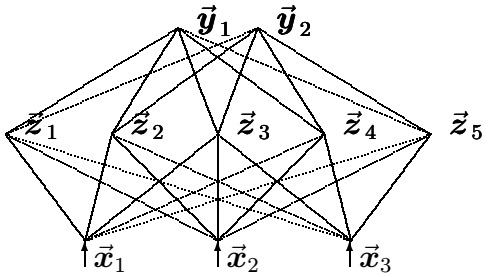
$$\frac{\partial y_{k+1}}{\partial w_j} = \left(\prod_{m=j+1}^{k} \widetilde{\sigma_m'}\,(w_m y_m) \cdot w_m \right) \widetilde{\sigma_j'}\,(w_j y_j) \cdot y_j,$$

pro $j \in \{1, \ldots, k-1\}$.

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■



Obrázek 1.2: Schéma architektury vrstevnaté sítě.

Uvědomíme-li si skutečnost, že formální pravidla pro derivaci složené funkce a pro derivaci součinu platí i v případě, kdy jednotlivé součinitele jsou matice a argumenty funkcí jsou vektory (zde předpokládáme, že $\widetilde{\sigma}$ je vektorová funkce, aplikovaná na jednotlivé složky svého argumentu stejně jako v jednorozměrném případě), je zřejmé, že lemma 1.2.1 popisuje i formální zápis derivace zobrazení, které je realizováno vícevrstvou neuronovou sítí. Nyní je ale zřejmý význam termínu zpětné šíření. Chceme-li spočítat hodnotu derivace výstupu z nějakého neuronu podle váhy mezi j-tou a $(j+1)$-ní vrstvou, postupujeme tak, že nejdříve spočítáme hodnotu všech neuronů pro daný vstupní vektor $\vec{\boldsymbol{x}}$, a potom podle analogických vzorců jako v lemmě 1.2.1 spočítáme derivace výstupních hodnot nejdříve pro váhy mezi poslední a předposlední vrstvou, dále pak pro váhy mezi vrstvami o jedničku níže a takto postupujeme až do konce pro váhy mezi první a druhou vrstvou.

Různé způsoby definování architektury sítě a její dynamiky umožňují postulovat nesčetné množství různých typů neuronových sítí. V současné době se hovoří o řádově několika desítkách mnohdy dosti odlišných neuronových sítí. Pro ilustraci a podání základní informace uvedeme pouze následující modely neuronových sítí, a to ve velmi sumarizované formě.

Hopfieldova síť

Tato síť je tvořena úplným grafem na n vrcholech. Každá váha mezi dvěma vrcholy má hodnotu ± 1 a pro danou posloupnost $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t \in \{-1, +1\}^n$ je váha mezi i-tým a j-tým neuronem dána vzorcem

$$\boldsymbol{W}_{ij} \stackrel{\text{def}}{=} \sum_{s=1}^{t} (\vec{\boldsymbol{x}}_i)_s \, (\vec{\boldsymbol{x}}_j)_s \,,$$

tedy výsledná matice všech vah $\boldsymbol{W}$ má tvar

$$\boldsymbol{W} = \sum_{i=1}^{t} \vec{\boldsymbol{x}}_i \vec{\boldsymbol{x}}_i^T. \tag{1.2}$$

Adaptace vah podle vzorce 1.2 umožňuje postupné předkládání jednotlivých vzorů po sobě. Aktivní dynamika, tedy vybavování, je popsána iteračním předpisem

$$\vec{\boldsymbol{y}}_{k+1} \stackrel{\text{def}}{=} \widetilde{f}\left(\boldsymbol{W}\vec{\boldsymbol{y}}_k\right), \quad \vec{\boldsymbol{y}}_k, \vec{\boldsymbol{y}}_{k+1} \in \{-1,+1\}^n, \tag{1.3}$$

kde hodnoty vektoru $\vec{\boldsymbol{y}}$ můžeme chápat jako výstupní hodnoty jednotlivých neuronů a $\widetilde{f}$ je již výše zmíněná tvrdá nelinearita. Iterace podle vzorce 1.3 se provádí, dokud není dosažen stabilní stav, kdy $\vec{\boldsymbol{y}}_{k+1} = \vec{\boldsymbol{y}}_k$.

Smyslem této neuronové sítě je zapamatovat si předkládané vzory. Dá se ukázat, že iterační posloupnost $\vec{\boldsymbol{y}}_k$ snižuje hodnoty energetické funkce definované jako

$$\widetilde{H}\left(\vec{\boldsymbol{y}}\right) \stackrel{\text{def}}{=} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{W}_{ij}\vec{\boldsymbol{y}}_i\vec{\boldsymbol{y}}_j \tag{1.4}$$

a že vede vždy do lokálního minima této funkce.

Kohonenova síť

Poměrně dosti odlišným typem neuronové sítě je tzv. Kohonenova síť (někdy též Kohonenova mapa). Graf této sítě má tu vlastnost, že počet vstupních hran každého neuronu je roven počtu vstupních neuronů celé sítě a k zakódování vstupních vektorů slouží váhy neuronů. Neurony nemají přenosovou funkci, ale pouze počítají v dané metrice vstupního prostoru vzdálenost předloženého vzoru od vektoru vstupních vah, např. Eukleidovskou vzdálenost ($\vec{\boldsymbol{w}}$ je vektor vah j-tého neuronu)

$$d \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n}\left(\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{w}}_i\right)^2}.$$

Tyto hodnoty lze považovat za výstupní hodnoty Kohonenovy sítě. Učení Kohonenovy sítě probíhá podle následujícího schématu.

- Nejdříve zvolíme parametr rychlosti učení $0 \leq \widetilde{\eta}\left(k\right) \leq 1$, který pro k-tý vzor rozhoduje o míře změny vah. Dále pro každý neuron zvolíme jeho okolí v grafu a zvolíme stategii zmenšování tohoto okolí v závislosti na výstupní hodnotě neuronu. Pro předložený vstupní vektor $\vec{\boldsymbol{x}} \in \Re^n$ spočtěme pro každý neuron sítě hodnotu

  $$d_j \stackrel{\text{def}}{=} \left\|\vec{\boldsymbol{x}} - \vec{\boldsymbol{w}}_j^k\right\|,$$

  kde $\vec{\boldsymbol{w}}_j^k$ je vektor vah pro k-tý neuron v čase k.

- Pro každý neuron j, pro který platí $d_j = \min_i\left\{d_i\right\}$, změníme váhy tohoto neuronu a všech neuronů v jeho aktuálním okolí podle formule

  $$\vec{\boldsymbol{w}}_{n+1}^{k+1} \stackrel{\text{def}}{=} \vec{\boldsymbol{w}}_n^k + \widetilde{\eta}\left(k\right)\left(\vec{\boldsymbol{x}} - \vec{\boldsymbol{w}}_n^k\right).$$

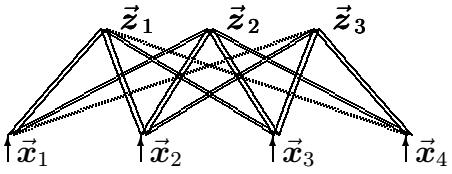
- Podle hodnoty d_j zredukujeme podle zvolené strategie okolí j-tého neuronu.
- Pokračujeme s tímto procesem pro každý další vstupní vektor.

Aktivní dynamika Kohonenovy sítě spočívá ve výpočtu hodnot d_j pro předložený vstupní vektor a všechny neurony v síti. Odpovědí sítě na tento vstup je pak pořadí neuronu s minimální hodnotou d_j.

Princip této sítě spočívá ve skutečnosti, že v průběhu učení se v síti vytvářejí jakési shluky neuronů, jejichž váhy jsou blízké předkládaným vzorům. Tyto shluky odpovídají třídám vstupních vektorů, které se pomocí Kohonenovy sítě vzájemně separují od sebe.

ART – Carpenter–Grossbergův klasifikátor

Návrhem této sítě se její autoři snažili potlačit nežádoucí jev projevující se u různých typů neuronových sítí, a to tzv. „problém proměnné stability“. Pod tímto pojmem se myslí skutečnost, kdy při předložení vzoru, který se má síť naučit, se poruší již dříve uložené informace, získané na základě dosavadního učení. Síť pak zpravidla již není konzistentní s dříve předloženými vzory pro učení. Proto byl rozpracován následující algoritmus Adaptivní Rezonanční Teorie.



Obrázek 1.3: Schéma architektury ART sítě ($N = 4$, $M = 3$).

Graf ART sítě je podobný dvouvrstvému modelu, ve kterém jsou spojeny všechny neurony vstupní vrstvy se všemi neurony vrstvy druhé, výstupní. Na rozdíl od vrstevnatých sítí, všechny hrany v grafu jsou obousměrné a každému směru je přiřazena váha. Hlavním rozdílem vzhledem k vrstvenatým sítím je opakovaná výměna vstupního vektoru mezi vstupní a výstupní vrstvou, tedy se dá říci, že vstupní vektor v této síti „rezonuje“. Dynamické chování sítě typu ART je popsáno následujícím schématem:

1. **Inicializace** Počáteční nastavení dopředných a zpětných vah:

$$\widetilde{\boldsymbol{W}}_{ij}(0) \stackrel{\text{def}}{=} \frac{1}{1+N}, \quad \widetilde{\boldsymbol{T}}_{ij}(0) \stackrel{\text{def}}{=} 1,$$

   kde $1 \leq i \leq N$, $1 \leq j \leq M$, a nastavení tzv. PRAHU OSTRAŽITOSTI ρ v rozmezí $0 \leq \rho \leq 1$ (symbol $\widetilde{\boldsymbol{W}}_{ij}(0)$ představuje dopřednou váhu, vedoucí z i-tého vstupu na j-tý výstup a symbol $\widetilde{\boldsymbol{T}}_{ij}(0)$ představuje zpětnou váhu, vedoucí opačným směrem, tj. z výstupů ke vstupům). Ve zpětných vahách jsou zakódovány všechny naučené vzory. Práh ostražitosti ρ určuje, jak moc je vstupní vzor vzdálen od nějakého již naučeného vzoru.

2. **Předložení nového vzoru na vstupy sítě** Přiložíme nový vzor $\vec{\boldsymbol{x}} \in \{-1, +1\}^N$ na vstup sítě.

3. **Výpočet výstupních hodnot**

$$\mu_j \stackrel{\text{def}}{=} \sum_{i=1}^{N} \widetilde{\boldsymbol{W}}_{ij}(t)\vec{\boldsymbol{x}}_i \ , \quad 1 \leq j \leq M$$

   (v této rovnici představuje μ_j výstup z j-tého neuronu a $\vec{\boldsymbol{x}}_i$ je i-tý element vstupního vzoru $\vec{\boldsymbol{x}}$, jehož hodnota může být pouze -1 nebo 1).

4. **Výběr nejlepšího výstupu**

$$\mu_{j^*} \stackrel{\text{def}}{=} \max_j \mu_j.$$

Tento vztah představuje postranní inhibici ve výstupní vrstvě sítě.

5. **Test ostražitosti** Porovnáme vstupní a vybraný vektor s prahem ostražitosti. Jestliže je

$$S \stackrel{\text{def}}{=} \frac{\sum_{i=1}^{N} \boldsymbol{T}_{ij^*} \vec{\boldsymbol{x}}_i}{\sum_{i=1}^{N} \vec{\boldsymbol{x}}_i} > \rho,$$

potom přejdeme ke Kroku 7, jinak pokračujeme Krokem 6 (S zřejmě představuje podobnost vstupního vzoru s vektorem zpětných vah pro j-tý neuron výstupní vrstvy).

6. **Potlačení vybraného vzoru** Výstup neuronu, který doposavad představoval nejlepší vybraný vzor nastavíme na nulu a tím se nebude dále uplatňovat při výběru dalšího výstupu (v Kroku 4).

Pokud jsme nevyčerpali všechny neurony ve výstupní vrstvě, tak pokračujeme Krokem 4. V opačném případě, máme-li dostatek volných výstupních neuronů, přiřadíme vstupní vzor k nějakému volnému neuronu a vytvoříme tak novou třídu.

7. **Přizpůsobení vah podle vybraného vzoru**

$$\widetilde{\boldsymbol{T}_{ij^*}}\,(t+1) \stackrel{\text{def}}{=} \widetilde{\boldsymbol{T}_{ij^*}}\,(t)\vec{\boldsymbol{x}}_i,$$

$$\widetilde{\boldsymbol{W}_{ij^*}}\,(t+1) \stackrel{\text{def}}{=} \frac{\widetilde{\boldsymbol{T}_{ij^*}}\,(t)\vec{\boldsymbol{x}}_i}{\frac{1}{2} + \sum_{i=1}^{N} \widetilde{\boldsymbol{T}_{ij^*}}\,(t)\vec{\boldsymbol{x}}_i}.$$

8. **Opakování algoritmu pro další vzor** Pokud chceme pokračovat v klasifikaci, povolíme všechny výstupy, které byly potlačeny v Kroku 6 a přejdeme ke Kroku 2. V opačném případě skončíme.

Předchozím modelem uzavřeme náš ilustrativní přehled nejužívanějších neuronových sítí a dále se budeme věnovat teoretické analýze vlastností zejména vrstevnatých sítí založených na myšlence perceptronu, důraz položíme na vlastnosti sítí s binárními vstupy a výstupy. Pro spojité sítě obdobného typu ukážeme základní aproximační vlastnosti a v dalších kapitolách naznačíme souvislost neuronových sítí a obecné teorie učení. Dále pak ukážeme některé numerické aspekty gradientních algoritmů učení.

Kapitola 2

Analýza booleovských sítí

Cooperův zákon

Pokud v matematickém textu nerozumíte nějakému slovu,
nevěšte hlavu.
Text dává smysl i bez něho.
(Jestli-že nedává, nedával by smysl ani s ním.)

V celé této kapitole se budeme věnovat analýze booleovských obvodů, to jest neuronových sítí, jejichž vstupy i výstupy, stejně jako stavy všech neuronů, jsou binární. Budeme zkoumat základní otázku, jaké booleovské funkce definované v definičním oboru $\{-1,+1\}^n$ je daná booleovská síť schopna spočítat. Vyjdeme z nejjednodušší booleovské sítě, booleovského perceptronu, který počítá tzv. prahový vektor svých vstupů, jak bude vidět dále.

V analýze booleovských funkcí hrají velmi důležitou roli vlastnosti Hadamardových matic, analyzované v následující části. Tyto vlastnosti nám umožní ostře separovat množiny booleovských funkcí, realizovaných obvody různé hloubky. Význam analýzy booleovských obvodů má význam i z hlediska praktického, v technické praxi je problém realizace dané logické funkce běžný, zejména v aplikacích řízení a optimalizace procesů. Kromě základní charakterizace funkcí počítaných různými obvody, se budeme zajímat i o horní a dolní odhad počtu lineárně separovatelných booleovských funkcí a horními a dolními odhady velikosti vah nezbytných k výpočtu dané funkce. V celé kapitole budeme budovat teorii pro binární vstupy ± 1, ale zřejmě všechna tvrzení zůstávají v platnosti pro libovolné binární vstupy $a, b \in \Re, a \neq b$.

2.1 Sylvestrova konstrukce Hadamardových matic

Tuto úvodní část věnujeme zavedení a popisu Hadamardových matic, hrajících klíčovou roli při analýze vlastností booleovských obvodů hloubky 1 a 2, které budou studovány v kapitole 2. Pro studium Hadamardových matic je vhodné zavést pojem tenzorového součinu podle následující definice.

Definice 2.1.1 *Nechť* $\boldsymbol{A}$ *je reálná matice typu* $r \times s$ *a* $\boldsymbol{B}$ *je reálná matice typu* $m \times n$ *. Potom* tenzorový součin matic $\boldsymbol{A}$ *a* $\boldsymbol{B}$ *je matice* $\boldsymbol{A} \odot \boldsymbol{B}$ *typu* $mr \times ns$ *,*

definovaná jako

$$\boldsymbol{A} \odot \boldsymbol{B} \stackrel{def}{=} \begin{pmatrix} \boldsymbol{A}b_{1,1} & \boldsymbol{A}b_{1,2} & \dots & \boldsymbol{A}b_{1,n} \\ \boldsymbol{A}b_{2,1} & \boldsymbol{A}b_{2,2} & \dots & \boldsymbol{A}b_{2,n} \\ & \vdots & & \vdots \\ \boldsymbol{A}b_{m,1} & \boldsymbol{A}b_{m,2} & \dots & \boldsymbol{A}b_{m,n} \end{pmatrix}.$$

Pro operaci tenzorového součinu platí asociativní zákon, což se dá velmi snadno ověřit přímo z definice této operace, a neplatí zákon komutativní. Definujme nyní Hadamardovy matice.

Definice 2.1.2 *Nechť* $\boldsymbol{A}$ *je čtvercová matice řádu* n*, jejíž prvky jsou z množiny* $\{-1, +1\}$ *. Potom* $\boldsymbol{A}$ *je* HADAMARDOVA MATICE *, jestli-že platí*

$$\boldsymbol{A} \cdot \boldsymbol{A}^T \stackrel{def}{=} n\boldsymbol{I}.$$

Souvislost Hadamardových matic a tenzorového součinu ozřejmuje následující lemma.

Lemma 2.1.1 *Tenzorový součin Hadamardových matic je opět Hadamardova matice.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Pro analýzu booleovských neuronových sítí jsou relevantní matice, vzniklé následující rekurzí z Hadamardovy matice řádu 2.

Definice 2.1.3 *Nechť* $\boldsymbol{A} \stackrel{def}{=} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ *. Potom matici* $\boldsymbol{B}^{(n)}$ *řádu* $2^n \times 2^n$ *vzniklou* n*-násobným tenzorovým násobením matice* $\boldsymbol{A}$ *nazveme* MATICE PARITY ŘÁDU n *.*

Způsob vytvoření matice $\boldsymbol{B}^{(n)}$ se v literatuře označuje jako Sylvestrova konstrukce. Matice $\boldsymbol{B}^{(n)}$ mají zajímavou algebraickou strukturu. Nejdůležitější vlastností z hlediska neuronových sítí je fakt, že sloupce matice $\boldsymbol{B}^{(n)}$ jsou tvořeny hodnotami booleovských funkcí $\{-1, +1\}^n \times \{-1, +1\}^n \to \{-1, +1\}$, které se nazývají FUNKCEMI PARITY , tj. funkcemi jejichž hodnota v argumentu $\left(\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}}\right)$, $\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}} \in \{-1, +1\}^n$ je rovna 1, právě když počet jedniček v $\vec{\boldsymbol{a}}$ na pozicích, kde v $\vec{\boldsymbol{b}}$ je jednička, se od počtu -1 na těchže pozicích liší o sudé číslo. Základní vlastnosti matic parity shrnuje následující pomocné tvrzení.

Lemma 2.1.2 *Pro matici parity* $\boldsymbol{B}^{(n)}$ *platí následující:*

1. $\boldsymbol{B}^{(n)}$ *je symetrická Hadamardova matice,*
2. $\boldsymbol{B}^{(n-1)}$ *je hlavní vedoucí submaticí matice* $\boldsymbol{B}^{(n)}$ *,*
3. *vynecháním sudých řádků a sloupců matice* $\boldsymbol{B}^{(n)}$ *dostaneme matici* $\boldsymbol{B}^{(n-1)}$*,*
4. *vynecháním lichých řádků a sloupců matice* $\boldsymbol{B}^{(n)}$ *dostaneme matici* $-\boldsymbol{B}^{(n-1)}$*,*
5. *počet jedniček v matici* $\boldsymbol{B}^{(n)}$ *se od počtu -1 v téže matici liší o hodnotu* 2^n*,*

6. *nechť* $\vec{\boldsymbol{a}}, \vec{\boldsymbol{x}} \in \{0,1\}^j$, $\vec{\boldsymbol{b}}, \vec{\boldsymbol{y}} \in \{0,1\}^{n-j}$, $0 < j < n$ *a nechť dále zápis* $(\vec{\boldsymbol{z}}, \vec{\boldsymbol{y}})$ *představuje blokový vektor s bloky* $\vec{\boldsymbol{z}}$ *a* $\vec{\boldsymbol{y}}$ *. Potom platí následující vztah:*

$$\boldsymbol{B}^{(n)}{}_{(\vec{\boldsymbol{a}},\vec{\boldsymbol{b}})^\sharp,(\vec{\boldsymbol{x}},\vec{\boldsymbol{y}})^\sharp} = \boldsymbol{B}^{(j)}{}_{\vec{\boldsymbol{a}}^\sharp,\vec{\boldsymbol{x}}^\sharp} \cdot \boldsymbol{B}^{(n-j)}{}_{\vec{\boldsymbol{b}}^\sharp,\vec{\boldsymbol{y}}^\sharp},$$

(řady matice $\boldsymbol{B}^{(n)}$ *číslujeme od 0 do* $2^n - 1$*),*

7. *pro všechna* $\vec{\boldsymbol{i}}, \vec{\boldsymbol{j}} \in \{0,1\}^n$ *platí rovnost*

$$\boldsymbol{B}^{(n)}{}_{\vec{\boldsymbol{i}}^\sharp,\vec{\boldsymbol{j}}^\sharp} = \left(-1\right)^{\sum_{\alpha=1}^{n} \vec{\boldsymbol{i}}_\alpha \cdot \vec{\boldsymbol{j}}_\alpha},$$

(řady matice $\boldsymbol{B}^{(n)}$ *číslujeme od 0 do* $2^n - 1$*).*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Dalším nosným pojmem pro analýzu booleovských neuronových sítí je pojem rostoucí matice.

Definice 2.1.4 *Čtvercová matice* $\boldsymbol{A}$ *je rostoucí, tvoří-li její řádky a sloupce neklesající posloupnosti. Čtvercová matice* $\boldsymbol{A}$ *je potenciálně rostoucí, existují-li permutační matice* $\boldsymbol{P}$ *a* $\boldsymbol{Q}$ *tak, že matice* $\boldsymbol{P} \cdot \boldsymbol{A} \cdot \boldsymbol{Q}$ *je rostoucí.*

Rostoucí matice musí mít hlavní diagonálu a všechny „diagonály“ rovnoběžné s hlavní diagonálou nerostoucí, což triviálně plyne z definice rostoucí matice. Stejně tak z definice triviálně plyne, že pro potenciálně rostoucí matici $\boldsymbol{A}$ je matice $-\boldsymbol{A}$ také potenciálně rostoucí a součet matice $\boldsymbol{A}$ s libovolnou konstantní maticí je opět potenciálně rostoucí matice. Existuje jednoduchá ekvivalentní podmínka pro to, aby matice byla potenciálně rostoucí.

Lemma 2.1.3 *Čtvercová matice* $\boldsymbol{A}$ *s prvky* ± 1 *je potenciálně rostoucí, právě když neobsahuje jako svojí podmatici matici* $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ *nebo matici* $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ *.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Počet submatic řádu 2 matice řádu n je zřejmě $\binom{n}{2}^2$ a tedy ověření, zda matice řádu n je potenciálně rostoucí, lze provést v čase $o\left(n^4\right)$. Další dvě pomocná tvrzení 2.1.4, 2.1.5 zavádí pomocné vektory $\vec{\boldsymbol{p}}$, $\vec{\boldsymbol{v}}$ a posloupnosti $\{z_i\}_1^\omega$ a $\{w_i\}_1^\omega$, jejichž algebraická struktura umožňuje dokázat stěžejní tvrzení této kapitoly.

Lemma 2.1.4 *Nechť* $\vec{\boldsymbol{p}}^{(0)} \stackrel{def}{=} \vec{\boldsymbol{v}}^{(0)} \stackrel{def}{=} 1$ *a nechť pro* $n \geq 1$ *platí*

$$\vec{\boldsymbol{p}}^{(n)} \stackrel{def}{=} \vec{\boldsymbol{p}}^{(n-1)} \odot \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad a \quad \vec{\boldsymbol{v}}^{(n)} \stackrel{def}{=} \vec{\boldsymbol{v}}^{(n-1)} \odot \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Nechť s_k *(resp.* t_k*) je součet prvních* k *členů číselné posloupnosti* $\{(\vec{\boldsymbol{p}}^{(n)})_i\}_1^{2^n}$ *(resp.* $\{(\vec{\boldsymbol{v}}^{(n)})_i\}_1^{2^n}$ *). Potom platí:*

1. $\forall l \in \{1,\dots,2^{n-1}\}$ *je* $s_{2l} = 0$
2. $\forall l \in \{1,\dots,2^{n-1}\}$ *je* $s_{2l-1} = (\vec{\boldsymbol{p}}^{(n)})_{2l-1} = (-1)^{\kappa(2l-1)}$, *kde symbol* $\kappa(i)$ *označuje počet jedniček v binárním rozvoji čísla* i.
3. $\forall l \in \{1,\dots,2^n - 1\}$ *platí* $s_l = t_{2^n-l}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Vektor $\vec{\boldsymbol{p}}^{(n)}$ definovaný v lemmatu je roven poslednímu sloupci matice $\boldsymbol{B}^{(n)}$, což přímo plyne z definice $\boldsymbol{B}^{(n)}$.

Lemma 2.1.5 *Nechť* $d_2,\dots,d_k$ *jsou přirozená čísla a* d_1 *je celé nezáporné číslo. Definujme posloupnost vektorů* $\vec{\boldsymbol{q}}_j$ *a* $\vec{\boldsymbol{u}}_j$ *rekurzivním předpisem:*

$$\begin{array}{lll} \vec{\boldsymbol{q}}_1 \stackrel{def}{=} \vec{\boldsymbol{p}}^{(d_1)} & & \vec{\boldsymbol{u}}_1 \stackrel{def}{=} \vec{\boldsymbol{v}}^{(d_1)} \\ \vec{\boldsymbol{q}}_2 \stackrel{def}{=} -\left(\vec{\boldsymbol{q}}_1^T, \vec{\boldsymbol{q}}_1^T, \dots, \vec{\boldsymbol{q}}_1^T\right)^T \quad d_2\times & & \vec{\boldsymbol{u}}_2 \stackrel{def}{=} -\left(\vec{\boldsymbol{q}}_1^T, \vec{\boldsymbol{q}}_1^T, \dots, \vec{\boldsymbol{q}}_1^T\right)^T \quad d_2\times \\ \vec{\boldsymbol{q}}_3 \stackrel{def}{=} \vec{\boldsymbol{q}}_2 \odot \vec{\boldsymbol{p}}^{(d_3)} & & \vec{\boldsymbol{u}}_3 \stackrel{def}{=} \vec{\boldsymbol{q}}_2 \odot \vec{\boldsymbol{v}}^{(d_3)} \\ \vec{\boldsymbol{q}}_4 \stackrel{def}{=} -\left(\vec{\boldsymbol{q}}_3^T, \vec{\boldsymbol{q}}_3^T, \dots, \vec{\boldsymbol{q}}_3^T\right)^T \quad d_4\times \quad a & & \vec{\boldsymbol{u}}_4 \stackrel{def}{=} -\left(\vec{\boldsymbol{q}}_3^T, \vec{\boldsymbol{q}}_3^T, \dots, \vec{\boldsymbol{q}}_3^T\right)^T \quad d_4\times \\ \vec{\boldsymbol{q}}_5 \stackrel{def}{=} \vec{\boldsymbol{q}}_4 \odot \vec{\boldsymbol{p}}^{(d_5)} & & \vec{\boldsymbol{u}}_5 \stackrel{def}{=} \vec{\boldsymbol{q}}_4 \odot \vec{\boldsymbol{v}}^{(d_5)} \\ \vec{\boldsymbol{q}}_6 \stackrel{def}{=} \dots & & \vec{\boldsymbol{u}}_6 \stackrel{def}{=} \dots \\ \dots & & \dots \end{array} \tag{2.1}$$

Nechť posloupnost $\{z_i\}_1$ *je definována jako zřetězení*

$$\{z_i\}_1 \stackrel{def}{=} \left(\vec{\boldsymbol{q}}_1^T, \vec{\boldsymbol{q}}_2^T, \dots, \vec{\boldsymbol{q}}_k^T\right),$$

a posloupnost $\{w_i\}_1$ *jako zřetězení*

$$\{w_i\}_1 \stackrel{def}{=} \left(\vec{\boldsymbol{u}}_1^T, \vec{\boldsymbol{u}}_2^T, \dots, \vec{\boldsymbol{u}}_k^T\right),$$

Potom platí:

1. *částečné součty posloupnosti* $\{z_i\}_1$ *jsou zhora omezeny číslem 1.*
2. *částečné součty posloupnosti* $\{w_i\}_1$ *jsou zdola omezeny číslem 0.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Pouze z důvodů stručného zápisu definujme skalární součin matic jako běžný skalární součin, kdy matici chápeme jako vektor obsahující její prvky.

Definice 2.1.5 *Nechť* $\boldsymbol{A}$ *a* $\boldsymbol{B}$ *jsou libovolné matice řádu* $m\times n$. *Potom pod pojmem* SKALÁRNÍ SOUČIN MATIC $\boldsymbol{A}$ *a* $\boldsymbol{B}$ *rozumíme číslo*

$$\langle \boldsymbol{A} \,|\, \boldsymbol{B} \rangle \stackrel{def}{=} \sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{A}_{i,j} \cdot \boldsymbol{B}_{i,j}.$$

Závěrem této kapitoly uveďme stěžejní tvrzení, umožňující zhora odhadnout skalární součin matice parity a libovolné rostoucí matice tvořené prvky ± 1. Jak uvidíme v kapitole o inkluzi tříd booleovských obvodů, tento odhad umožní jednotlivé třídy od sebe separovat.

Tvrzení 2.1.6 *Nechť* $\boldsymbol{C}$ *je čtvercová matice dimenze* 2^n, $n \geq 1$ *taková, že* $\boldsymbol{C}_{i,j} \stackrel{def}{=} 1$ *pro* $i + j \leq 2^n + 1$ *a* $\boldsymbol{C}_{i,j} \stackrel{def}{=} -1$ *pro* $i + j > 2^n + 1$ *(tedy matice* $\boldsymbol{C}$ *má na vedlejší diagonále a nad ní samé jedničky, pod vedlejší diagonálou jsou samé* -1*). Dále předpokládejme, že* $\boldsymbol{D}$ *je libovolná rostoucí matice dimenze* 2^n *s prvky z množiny* $\{-1, +1\}$ *. Potom platí, že*

$$2^n(n+1) = \left\langle \boldsymbol{B}^{(n)} \,|\boldsymbol{C}\right\rangle \geq \left\langle \boldsymbol{B}^{(n)} \,|\boldsymbol{D}\right\rangle.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Poznamenejme, že maximální vlastnost matice $\boldsymbol{C}$ vyslovená v předešlé větě platí pro širší třídu matic s prvky ± 1, které mají monotónní pouze sloupce, nebo mají monotónní pouze řádky (monotónní v tomto kontextu znamená, že daný sloupec (řádka) tvoří neklesající posloupnost).

Cvičení 2.1.0 Hadamardovy matice

1. *Dokažte, že je-li* n *dimenze libovolné Hadamardovy matice, potom* n *je 1,2 nebo násobek 4. (Stanovte podmínky, které musí splňovat první tři řádky.)*

2. *Dokažte následující Linsdeyovo lemma:*

   Lemma 2.1.7 (Linsdey) *Nechť* $\boldsymbol{X}$ *je matice řádu* $n \times m$ *, s prvky* $\boldsymbol{X}_{i,j} = \pm 1$*, jejíž sloupce jsou ortogonální. Nechť* $\bar{R}$ *je podmnožina řádkových indexů,* $\bar{S}$ *je podmnožina sloupcových indexů a* $\gamma \stackrel{def}{=} \sum_{s\in\bar{S}} \sum_{r\in\bar{R}} \boldsymbol{X}_{r,s}$*. Potom platí:*

   $$|\gamma| \leq \sqrt{n\left|\bar{S}\right|\left|\bar{R}\right|}.$$

   (Návod: Spočtěte hodnotu $\mathtt{T} \stackrel{def}{=} \sum_{j=1}^{m} \left(\sum_{r\in\bar{R}} \boldsymbol{X}_{r,j}\right)^2$*. Potom porovnejte tuto hodnotu s hodnotou* γ *. Využijte Cauchyho nerovnosti* $\sqrt{\sum_{k=1}^{q} 1}\sqrt{\sum_{k=1}^{q} \alpha_k^2} \geq \sum_{k=1}^{q} \alpha_k$*.)*

3. *V matici* $\boldsymbol{B}^{(n)}$ *nalezněte podmatici, která realizuje rovnost v Lindsdeyově nerovnosti (tedy ukažte, že odhad daný Linsdeyovou nerovností nelze pro matici* $\boldsymbol{B}^{(n)}$ *zlepšit).*

2.2 Práh množiny vektorů

Výchozím pojmem pro analýzu booleovských obvodů bude pojem prahového vektoru (viz. např. [RSO94]). Tento pojem nám umožní zúžit naši pozornost nejenom pouze na booleovské funkce definované na celé krychli $\{-1, +1\}^n$, ale i na booleovské funkce definované na libovolné podmnožině diskrétní množiny $\{-1, +1\}^n$. Každou booleovskou funkci můžeme popsat klasicky jako zobrazení $\widetilde{b} : \{-1, +1\}^n \to \{-1, +1\}$. Toto zobrazení lze

popsat buď nějakým předpisem, nebo jako posloupnost uspořádaných dvojic argumentů $\vec{z} \in \{-1,+1\}^n$ a funkčních hodnot $\widetilde{b}(\vec{z})$, což máme k dispozici i v případě definice $\widetilde{b}$ pomocí nějakého předpisu. Booleovskou funkci $\widetilde{b'}$ definovanou pouze na nějaké podmnožině krychle $\{-1,+1\}^n$ můžeme opět popsat posloupností uspořádaných dvojic argumentů, v nichž je $\widetilde{b'}$ definována, a odpovídajících funkčních hodnot. Druhou možností, jak popsat $\widetilde{b'}$, je uvést všechny vektory $\vec{x}_1,\ldots,\vec{x}_S$ (v našem případě S vektorů), z nichž vektor $\vec{x}_1$ je tvořen prvními složkami argumentů, ve kterých je $\widetilde{b'}$ definována, $\vec{x}_2$ je tvořen druhými složkami argumentů, v nichž je $\widetilde{b'}$ definována (v pevně zvoleném pořadí argumentů), $\vec{x}_3$ třetími složkami, atd. K těmto vektorům $\vec{x}_1,\ldots,\vec{x}_S$ ještě přidáme vektor $\vec{y}$, který obsahuje odpovídající funkční hodnoty $\widetilde{b'}$ (opět ve stejném pořadí). Potom ale tedy systém vektorů $\vec{x}_1,\ldots,\vec{x}_S$ a $\vec{y}$ popisuje booleovskou funkci $\widetilde{b'}$.

Dimenze těchto vektorů je zřejmě rovna počtu vrcholů krychle $\{-1,+1\}^n$, ve kterých je $\widetilde{b'}$ definována. Jak uvidíme dále, tento – dalo by se říci – „transponovaný“ pohled na způsob definice booleovských funkcí umožňuje prostřednictvím geometrických vlastností vektorů $\vec{x}_1,\ldots,\vec{x}_S$ analyzovat vektory $\vec{y}$, které lze vyjádřit jako vážený součet $\vec{y} = \sum_{i=1}^{S} w_i\vec{x}_i$, tedy vektory $\vec{y}$ spočítatelné pomocí booleovského perceptronu. Na rozdíl od klasické definice perceptronu nebudeme brát v úvahu existenci prahu, neboť ten lze zavést v našem pojetí přidáním vektoru $\vec{x}_{S+1}$, který je konstantní a je tvořen buď $+1$ či -1. Výhodou tohoto postupu je kompaktnější a přehlednější zápis všech tvrzení a důkazů.

Definice 2.2.1 *Nechť funkce* $\widetilde{sgn} : \Re^n \to \{-1,+1\}$ *je definována jako*

$$\widetilde{sgn}(z) \stackrel{def}{=} \begin{cases} 1 & pro \quad x > 0 \\ -1 & \qquad x \leq 0 \end{cases}.$$

Definice 2.2.2 *Nechť* $\vec{x}_1,\ldots,\vec{x}_S$ *jsou vektory z prostoru z* $\{-1,+1\}^n$. *Pak vektor* $\vec{y}$ *je* PRÁH VEKTORŮ $\vec{x}_1,\ldots,\vec{x}_S$, *jestliže existují čísla* $w_1,\ldots,w_S$ *tak, že vektor* $\sum_{i=1}^{S} w_i \cdot \vec{x}_i$ *má pouze nenulové složky, a platí:*

$$\vec{y} \stackrel{def}{=} \widetilde{sgn}\left(\sum_{i=1}^{S} w_i \cdot \vec{x}_i\right).$$

(Funkce $\widetilde{sgn}$ *je aplikována na vektor po složkách).*

Základní vlastností prahového vektoru je fakt, že není kolmý na žádný vektor $\vec{x}_i$.

Lemma 2.2.1 *Nechť* $\vec{y}^T(\vec{x}_1,\ldots,\vec{x}_S) = \vec{0}^T$. *Potom* $\vec{y}$ *není prahem* $(\vec{x}_1,\ldots,\vec{x}_S)$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Následující tvrzení ukazuje, že pro libovolné různé prahy $\vec{y}$ a $\vec{z}$ systému $\vec{x}_1,\ldots,\vec{x}_S$ jsou vektory $\vec{y}^T(\vec{x}_1,\ldots,\vec{x}_S)$ a $\vec{z}^T(\vec{x}_1,\ldots,\vec{x}_S)$ nutně od sebe různé, což umožňuje horní odhad počtu prahových vektorů.

Lemma 2.2.2 *Nechť* $\vec{y}^T$ *je práh systému* $(\vec{x}_1,\ldots,\vec{x}_S)$, $\vec{x}_i \neq \vec{0}$, $i \in \{1,\ldots,S\}$ *a matice* $\boldsymbol{X}$ *má za své sloupce právě vektory* $\vec{x}_i$. *Potom pro všechny vektory* $\vec{z} \in \{-1,+1\}^n$, $\vec{y} \neq \vec{z}$, *platí, že*

$$\vec{y}^T \cdot \boldsymbol{X} \neq \vec{z}^T \cdot \boldsymbol{X}.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Opačná implikace však neplatí, což lze dokumentovat na následujícím příkladě.

Příklad 2.2.1 *Zaveďme následující blokové rozdělení matice parity* $\boldsymbol{B}^{(n)}$ *definované jako*

$$\boldsymbol{B}^{(n)} \stackrel{def}{=} \begin{pmatrix} 1 & \vec{\boldsymbol{j}}^T \\ \vec{\boldsymbol{j}} & \boldsymbol{A} \end{pmatrix},$$

kde $\boldsymbol{A}$ *je čtvercová matice řádu* $2^n - 1$ *a* $\vec{\boldsymbol{j}}$ *je vektor z jedniček dimenze taktéž* $2^n - 1$. *Z vlastností paritních matic bezprostředně vyplývá*

$$\begin{pmatrix} 1 & \vec{\boldsymbol{j}}^T \\ \vec{\boldsymbol{j}} & \boldsymbol{A} \end{pmatrix} \cdot \begin{pmatrix} 1 & \vec{\boldsymbol{j}}^T \\ \vec{\boldsymbol{j}} & \boldsymbol{A} \end{pmatrix} = \begin{pmatrix} 2^n & \vec{\boldsymbol{0}}^T \\ \vec{\boldsymbol{0}} & 2^n \boldsymbol{I} \end{pmatrix}.$$

Současně ale z vlastností Schurova doplňku matice $\mathbf{1}$ *v matici* $\boldsymbol{B}^{(n)}$ *plyne, (viz. [Fie81], str. 31), že tento Schurův doplněk je inverzní maticí k matici* $\boldsymbol{A}$ *, tedy*

$$\boldsymbol{A}^{-1} = \frac{1}{2^n}\left(\boldsymbol{A} - \vec{\boldsymbol{j}} \cdot \vec{\boldsymbol{j}}^T\right).$$

Odtud ale dostáváme, že $\boldsymbol{A}^{-1}$ *je nekladná matice (obsahuje buď nuly, nebo čísla* $-\frac{2}{2^n}$*).*

Předpokládejme nyní, že pro nějaký vektor $\vec{\boldsymbol{w}}$ *platí* $\boldsymbol{A}\vec{\boldsymbol{w}} = \vec{\boldsymbol{r}} > \vec{\boldsymbol{0}}$. *Potom ale* $\vec{\boldsymbol{w}} = \boldsymbol{A}^{-1} \cdot \vec{\boldsymbol{r}} > \vec{\boldsymbol{0}}$. *Poslední nerovnost zřejmě platí na základě nekladnosti matice* $\boldsymbol{A}^{-1}$. *Definujme dále matici*

$$\boldsymbol{V}^{(n)} \stackrel{def}{=} \begin{pmatrix} -1 & \vec{\boldsymbol{j}}^T \\ \vec{\boldsymbol{j}} & \boldsymbol{A} \end{pmatrix},$$

která se od matice $\boldsymbol{B}^{(n)}$ *liší pouze ve znaménku prvku vlevo nahoře. Z výše uvedeného rozboru vlastností matice* $\boldsymbol{A}$ *vyplývá, že první sloupec matice* $\boldsymbol{V}^{(n)}$ *nemůže být prahovým vektorem zbylých sloupců, protože kdyby tomu tak bylo, pak by všechny koeficienty* $\vec{\boldsymbol{w}}_i$ *byly kladné, ale tím pádem se dostáváme do sporu, protože první indexy sloupců matice* $\boldsymbol{V}^{(n)}$ *(vyjma prvního) jsou kladné, a nelze z nich tedy nakombinovat záporné číslo.*

Současně ale sloupce matice $\boldsymbol{V}^{(n)}$ *tvoří bazi prostoru* $\Re^{2^n}$ *, přičemž první sloupec neleží v ortogonálním doplňku ostatních. Proto pro libovolné různé vektory* $\vec{\boldsymbol{z}}, \vec{\boldsymbol{y}} \in \{-1, +1\}^{2^n}$ *se vektory*

$$\vec{\boldsymbol{z}}^T \boldsymbol{V}^{(n)} \quad a \quad \vec{\boldsymbol{y}}^T \boldsymbol{V}^{(n)}$$

musí lišit alespoň v jedné ze složek s indexem větším než 1. *To dokazuje, že tvrzení lemmy 2.2.2 nelze obrátit.*

Zřejmě na základě lemmy 2.2.2 lze vyslovit následující tvrzení.

Lemma 2.2.3 *Pro systém vektorů* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S \in \{-1, +1\}^n$ *existuje nanejvýše* $(n+1)^S$ *různých prahů.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Významnou roli v analýze aproximačních možností booleovských obvodů hrají souřadnice prahového vektoru $\vec{\boldsymbol{y}}$ vzhledem k systému $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$.

Definice 2.2.3 *Předpokládejme, že vektory* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ *jsou lineárně nezávislé a nechť*

$$\vec{\boldsymbol{y}} \stackrel{def}{=} \sum_{i=1}^{S} \beta_i \vec{\boldsymbol{x}}_i + \vec{\boldsymbol{y}}^{\perp},$$

kde $\vec{\boldsymbol{y}}^{\perp}$ *je z ortogonálního doplňku* $[\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S]_\lambda$. *Potom vektor* $\vec{\boldsymbol{\beta}} \stackrel{def}{=} (\beta_1, \ldots, \beta_S)$ *nazveme* ZOBECNĚNÉ SPEKTRUM VEKTORU $\vec{\boldsymbol{y}}$ *a číslo*

$$\beta_{max} \stackrel{def}{=} \max \{ |\beta_i| \, | i \in \{1, \ldots, S\} \}$$

nazveme MAXIMUM SPEKTRA *vektoru* $\vec{\boldsymbol{y}}$.

Čísla β_i lze spočítat na základě pojmu pseudoinverzní matice takto: Označme si $\boldsymbol{X} \stackrel{\text{def}}{=} (\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S)^T$. Potom platí

$$\vec{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \vec{\boldsymbol{y}}.$$

Základní vlastností vektoru $\vec{\boldsymbol{\beta}}$ je následující odhad součtu absolutních hodnot jeho složek zdola.

Lemma 2.2.4 *Nechť* $\vec{\boldsymbol{y}}$ *je práh systému* $(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S)$. *Potom platí*

$$\sum_{i=1}^{S} \left| \vec{\boldsymbol{\beta}}_i \right| \geq 1.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Předešlé tvrzení poskytuje základní vztah mezi hodnotami zobecněného spektra a počtem S vektorů $\vec{\boldsymbol{x}}_i$. Toto tvrzení, které je hlavním závěrem této části výkladu, lze využít k důkazu neexistence polynomiálního počtu vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ takových, že $\vec{\boldsymbol{y}}$ je jejich prahem. Princip tohoto důkazu je ten, že shora odhadneme hodnoty $\left| \vec{\boldsymbol{\beta}}_i \right|$, a budou-li polynomiálně malé vzhledem k dimenzi $\vec{\boldsymbol{y}}$, pak S nemůže být omezeno polynomem vzhledem k dimenzi $\vec{\boldsymbol{y}}$.

Tvrzení 2.2.5 *Nechť* $\vec{\boldsymbol{y}}$ *je práh ortogonálních vektorů* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S \in \{-1, +1\}^n$. *Pak*

$$S \geq \frac{n}{\max_i \langle \vec{\boldsymbol{y}} \, | \vec{\boldsymbol{x}}_i \rangle}. \tag{2.2}$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Ukážeme ještě jednu formu tvrzení předešlého, zahrnujícího v sobě i odhad na velikost maximální váhy prahového vektoru.

Lemma 2.2.6 *Předpokládejme, že* $\vec{\boldsymbol{y}}$ *je práh systému* $(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S)$ *vektorů z* $\{-1,+1\}^n$ *a že příslušné hodnoty* $w_1, \ldots, w_s$ *jsou celočíselné. Nechť* $\tilde{w} \stackrel{def}{=} max_i|w_i|$ *a* $\tilde{\tau} \stackrel{def}{=} max_i|\langle \vec{\boldsymbol{y}} \,|\vec{\boldsymbol{x}}_i\rangle|$. *Potom platí*

$$S \geq \frac{n}{\tilde{w}\tilde{\tau}}.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

2.2.1 Vlastnosti prahových vektorů

Nyní se pokusíme analyzovat základní vlastnosti prahových vektorů obecně, bez ohledu na konkrétní systém $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$. Následující příklad ilustruje dobře známou skutečnost, že práh dvou vektorů $\vec{\boldsymbol{x}}_1 \stackrel{\text{def}}{=} (1,-1,1,-1)^T$, $\vec{\boldsymbol{x}}_2 \stackrel{\text{def}}{=} (1,1,-1,-1)^T$ nemůže realizovat logickou funkci $\stackrel{\sqcup}{\text{XOR}}$, definovanou na sobě si odpovídajících složkách vektorů $\vec{\boldsymbol{x}}_1$ a $\vec{\boldsymbol{x}}_2$.

Příklad 2.2.2 *Jesliže přijmeme předpoklad, že platí rovnost*

$$\begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \widetilde{sgn}\left(\beta \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} + \gamma \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}\right),$$

potom musí být splněno

$$\left.\begin{array}{l} \beta+\gamma<0 \\ \beta-\gamma>0 \\ \beta+\gamma>0 \\ \beta-\gamma<0 \end{array}\right\} \rightarrow \left.\begin{array}{l} \beta+\gamma=0 \\ \beta-\gamma=0 \end{array}\right\} \rightarrow \begin{array}{l} \beta=0 \\ \gamma=0 \end{array},$$

což je ve sporu s definicí prahového vektoru. Alternativně lze tento závěr obdržet na základě skutečnosti, že všechny tři uvažované vektory jsou vzájemně ortogonální, tedy žádný z nich nemůže být prahem zbývajících dvou (stejná argumentace platí i pro případ, kdy přidáme navíc konstantní vektor).

Jinými slovy booleovský perceptron se dvěma vstupními neurony a jedním neuronem výstupním nemůže počítat logickou funkci $\stackrel{\sqcup}{\text{XOR}}$, definovanou na svých vstupech. Bohužel tento negativní závěr byl v šedesátých letech interpretován jako charakterizující vlastnost sítí tvořených více booleovskými perceptrony, čímž došlo k výraznému zpoždění vývoje v tomto oboru.

V dalším nyní ukážeme, že tato skutečnost, popsaná v příkladě 2.2.2, má mnohem hlubší podstatu a lze ji velmi zobecnit. Definujme proto následující matici $\boldsymbol{M}_{\vec{\boldsymbol{y}}}$.

Definice 2.2.4 *Nechť* $n \stackrel{def}{=} 2k$ *a* $\vec{\boldsymbol{y}} \in \{-1,+1\}^{2^n}$. *Nechť* $\boldsymbol{M}_{\vec{\boldsymbol{y}}}$ *je čtvercová matice řádu* $2^{\frac{n}{2}}$, *jejíž sloupce jsou tvořeny po sobě následujícími subvektory vektoru* $\vec{\boldsymbol{y}}$ *délky* $2^{\frac{n}{2}}$. *Matici* $\boldsymbol{M}_{\vec{\boldsymbol{y}}}$ *nazveme* SDRUŽENOU MATICÍ *pro vektor* $\vec{\boldsymbol{y}}$.

Například pro vektor $\vec{\boldsymbol{y}} \stackrel{\text{def}}{=} (a,b,c,d,)^T$ je $\boldsymbol{M}_{\vec{\boldsymbol{y}}} = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$. Při analýze sdružených matic bude užitečná lemma o vlastnostech rostoucích matic.

Lemma 2.2.7 *Platí následující tvrzení:*

1. *Je-li matice* $\boldsymbol{A}$ *rostoucí (potenciálně rostoucí), je i matice* $\widetilde{sgn}\,(\boldsymbol{A})$ *rostoucí (potenciálně rostoucí).*
2. *Nechť prvky matice* $\boldsymbol{A}$ *jsou tvořeny čísly* ± 1*. Potom je-li* $\widetilde{sgn}\,(\boldsymbol{A})$ *rostoucí (potenciálně rostoucí), je i matice* $\boldsymbol{A}$ *rostoucí (potenciálně rostoucí).*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Jak již bylo řečeno, při našem výkladu využíváme popisu booleovských funkcí pomocí pojmu prahových vektorů, které kromě jiného umožňují popis a analýzu booleovských funkcí definovaných pouze v nějaké podmnožině $\{-1,+1\}^n$. Pro popis booleovské funkce definované v celém oboru $\{-1,+1\}^n$, musíme používat takový systém vektorů $\vec{\boldsymbol{x}}_1,\ldots,\vec{\boldsymbol{x}}_S$, který má tu vlastnost, že libovolná s-členná posloupnost tvořená z ± 1 je obsažena mezi řádky matice $\boldsymbol{X} \stackrel{\text{def}}{=} (\vec{\boldsymbol{x}}_1,\ldots,\vec{\boldsymbol{x}}_S)$. Jedním (a nejpřirozenějším) z těchto systémů je systém základních vektorů parity podle následující definice.

Definice 2.2.5 *Předpokládejme, že sloupce paritní matice* $\boldsymbol{B}^{(n)}$ *jsou očíslovány počínaje nulou. Potom vektor tvořící sloupec matice* $\boldsymbol{B}^{(n)}$*, jehož index v binárním zápisu obsahuje pouze jednu jedničku na* i*-té pozici zleva, nazveme* i*-tý* ZÁKLADNÍ VEKTOR PARITY *a budeme ho značit* $\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_i$.

Příklad 2.2.3 *Pro lepší čitelnost předešlé definice uveďme následující příklad základních vektorů parity pro* $n = 4$*:*

$$\tilde{\vec{\boldsymbol{p}}}^{(4)}{}_1 = (\ \ 1,-1,\ \ 1,-1,\ \ 1,-1,\ \ 1,-1,\ \ 1,-1,\ \ 1,-1,\ \ 1,-1,\ \ 1,-1)^T,$$

$$\tilde{\vec{\boldsymbol{p}}}^{(4)}{}_2 = (\ \ 1,\ \ 1,-1,-1,\ \ 1,\ \ 1,-1,-1,\ \ 1,\ \ 1,-1,-1,\ \ 1,\ \ 1,-1,-1)^T,$$

$$\tilde{\vec{\boldsymbol{p}}}^{(4)}{}_3 = (\ \ 1,\ \ 1,\ \ 1,\ \ 1,-1,-1,-1,-1,\ \ 1,\ \ 1,\ \ 1,\ \ 1,-1,-1,-1,-1)^T,$$

$$\tilde{\vec{\boldsymbol{p}}}^{(4)}{}_4 = (\ \ 1,\ \ 1,\ \ 1,\ \ 1,\ \ 1,\ \ 1,\ \ 1,\ \ 1,-1,-1,-1,-1,-1,-1,-1,-1)^T,$$

(Srovnej s obrázkem 2.1).

Nechť $\boldsymbol{Z}$ je matice z čísel 0 a 1 o n sloupcích a 2^n řádcích a nechť j-tý řádek této matice je roven binárnímu zápisu čísla $j-1$. Dále předpokládejme, že matice $\boldsymbol{A}$ vznikla z matice $\boldsymbol{Z}$ záměnou 1 za -1 a 0 za 1. Potom $\boldsymbol{A} = \left(\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_n,\ldots,\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_1\right)$.

Definice 2.2.6 *Označme symbolem* $\boldsymbol{E}^{(n)}$ *čtvercovou matici řádu* 2^n *obsahující samé jedničky. Symbolem* $\boldsymbol{I}^{(n)}$ *označme jednotkovou matici řádu* 2^n*.*

Algebraickou strukturu matic sdružených k základním vektorům parity popisuje následující lemma.

Lemma 2.2.8 *Předpokládejme, že* $n \stackrel{def}{=} 2k$ *a* $\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_1,\ldots,\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_n$ *jsou odpovídající základní vektory parity. Potom platí:*

1. *pro* $1 \leq j \leq k$ *je*

$$\boldsymbol{M}_{\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_j} = \boldsymbol{E}^{(j-1)} \odot \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \odot \boldsymbol{E}^{(k-j)}.$$

a

$$\boldsymbol{M}_{\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_{k+j}} = \boldsymbol{E}^{(j-1)} \odot \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \odot \boldsymbol{E}^{(k-j)}.$$

2. *Pro* $1 \leq j \leq k$ *platí, že*

$$\boldsymbol{M}_{\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_{k+j}} = \boldsymbol{M}^T_{\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_j}.$$

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Pro aplikaci předchozího algebraického popisu sdružených matic základních vektorů parity bude užitečná následující identita, platná pro tenzorový součin matic.

Lemma 2.2.9 *Nechť matice* $\boldsymbol{A}$, $\boldsymbol{B}$ *a* $\boldsymbol{C}$, $\boldsymbol{D}$ *jdou spolu násobit v tomto pořadí. Potom platí identita*

$$(\boldsymbol{A} \odot \boldsymbol{B}) \cdot (\boldsymbol{C} \odot \boldsymbol{D}) = (\boldsymbol{A} \cdot \boldsymbol{C}) \odot (\boldsymbol{B} \cdot \boldsymbol{D}).$$

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Za hlavní závěr této podkapitoly lze považovat tvrzení, jehož obsahem je skutečnost, že každá lineární kombinace sdružených matic základních vektorů parity je potenciálně rostoucí.

Tvrzení 2.2.10 *Nechť* $n \stackrel{def}{=} 2k$ *a* $\vec{\boldsymbol{y}}$ *je libovolná lineární kombinace základních vektorů parity* $\tilde{\vec{\boldsymbol{p}}}^{(n)}{}_1, \ldots, \tilde{\vec{\boldsymbol{p}}}^{(n)}{}_n$. *Potom matice* $\boldsymbol{M}_{\vec{\boldsymbol{y}}}$ *je potenciálně rostoucí.*

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Evidentním důsledkem je následující tvrzení:

Tvrzení 2.2.11 *Nechť* $n \stackrel{def}{=} 2k$, $\vec{\boldsymbol{y}} \in \{-1, +1\}^{2^n}$ *je práh systému základních vektorů parity a konstantního vektoru. Potom matice* $\boldsymbol{M}_{\vec{\boldsymbol{y}}}$ *je potenciálně rostoucí.*

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Vezměme v úvahu booleovskou funkci $\overset{\sqcup}{\text{XOR}}$ definovanou na krychli $\{-1, +1\}^2$. Potom odpovídající základní vektory parity jsou $(1, -1, 1, -1)^T$ a $(1, 1, -1, -1)^T$. Vektor $\vec{\boldsymbol{y}}$ funkčních hodnot $\overset{\sqcup}{\text{XOR}}$ je přitom roven $\vec{\boldsymbol{y}} = (-1, 1, 1, -1)^T$. Potom ale $\boldsymbol{M}_{\vec{\boldsymbol{y}}} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$

není potenciálně rostoucí matice, tedy $\vec{y}$ nemůže být prahem základních vektorů parity a konstantního vektoru.

Algebraická struktura základních vektorů parity nám umožní přesně spočítat počet od sebe různých prahových vektorů tohoto systému. Zdůrazněme však, že tento počet není roven počtu booleovských funkcí realizovaných binárním perceptronem v klasickém slova smyslu (ten není znám, v dalším uvedeme pouze některé známé odhady), nýbrž počtu booleovských funkcí realizovaných perceptronem s nulovým prahem. Bohužel zavedení nenulového prahu podstatným způsobem tuto analýzu počtu prahů zkomplikuje. Vezmeme-li ale v úvahu pouze nulový práh, platí následující.

Tvrzení 2.2.12 *Pro $n \stackrel{def}{=} 2k$ je počet různých prahů systému $\vec{\tilde{p}}^{(n)}{}_1, \ldots, \vec{\tilde{p}}^{(n)}{}_n$ základních vektorů parity roven číslu $2^{k(k+1)}$.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

2.3 Inkluze základních tříd booleovských obvodů

V následujícím obrátíme pozornost na množiny booleovských funkcí, které lze vyjádřit booleovskými neuronovými sítěmi (obvody) různé hloubky. Zavedeme třídy booleovských funkcí, n-proměnných, které lze spočítat obvody, jejichž počet uzlů je polynomiálně omezen v n.

Uvidíme, že důležitou roli hrají tzv. α-symetrické booleovské vektory. Na základě výkladu o prahových vektorech ukážeme způsob separace jednotlivých tříd prahových vektorů počítaných obvody různé hloubky. Začneme s definicí booleovského obvodu.

Definice 2.3.1 BOOLEOVSKÝ OBVOD *je acyklický orientovaný graf s hranovým a vrcholovým ohodnocením a jedinným vrcholem, ze kterého nevede žádná hrana. Vrchol tohoto grafu nazýváme* VSTUPNÍ VRCHOL *, nevede-li do něj žádná hrana,* VÝSTUPNÍ VRCHOL *, nevede-li z něj žádná hrana, a* VNITŘNÍ VRCHOL *, není-li ani vstupní, ani výstupní. Hodnoty hranového ohodnocení jsou obecně reálná čísla, ohodnocení hrany z vrcholu i do vrcholu j nazveme* VÁHOU HRANY $i \to j$ *a budeme značit $w_{i\to j}$.*

Ohodnocení vrcholů jsou uspořádané dvojice $(v,t) \in \{-1,+1\} \times \Re$, které pro vrchol i budeme budeme značit jako (v_i, t_i). Hodnotu v_i u vstupních vrcholů nazýváme vstupními hodnotami, ohodnocení v_i u výstupních vrcholů, výstupními hodnotami. Dále, ohodnocení vnitřního a výstupního vrcholu i splňuje rovnost

$$v_i \stackrel{def}{=} \widetilde{sgn}\left(\sum_{\{j|j\to i\}} w_{j\to i}\cdot v_j - t_i\right).$$

Definice 2.3.2 VELIKOST BOOLEOVSKÉHO OBVODU *je počet vnitřních vrcholů.* HLOUBKA VRCHOLU *v je definována jako délka nejdelší cesty mezi libovolným vstupním vrcholem a vrcholem v.* HLOUBKA BOOLEOVSKÉHO OBVODU *je rovna maximu hloubek všech vnitřních vrcholů.* DIMENZE BOOLEOVSKÉHO OBVODU *je rovna počtu vstupních vrcholů.*

V této definici booleovského obvodu připouštíme existenci prahových hodnot t_i jednotlivých uzlů (neuronů) booleovského obvodu. Toto rozšíření ve srovnání s definicí prahového vektoru systému $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ ale stále umožňuje používat všech vlastností prahových vektorů při analýze booleovských obvodů, protože existenci prahových hodnot t_i v uzlech booleovského obvodu lze simulovat existencí konstantního vektoru v systému $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$.

Definice 2.3.3 *Booleovský obvod dimenze n, $n \geq 1$, počítá vektor $\vec{\boldsymbol{y}} \in \{-1,+1\}^{2^n}$, právě když existuje takové očíslování vstupních vrcholů $i_1, \ldots, i_n$ a takové hranové ohodnocení, že pro každé ohodnocení vstupních vrcholů čísly ± 1, je ohodnocení výstupního vrcholu rovno číslu $\vec{\boldsymbol{y}}_\gamma$, kde index γ má v binárním zápisu na pozici k nulu, je-li $i_k = -1$, a jedničku, je-li $i_k = +1$.*

Pro lepší pochopení přesného významu předešlé definice uveďme jednoduchý příklad. Vektor $\vec{\boldsymbol{y}} \stackrel{\text{def}}{=} (1,1,-1,1,-1,-1,1,-1)^T$ je počítán booleovským obvodem B, právě když pro zobrazení $\widetilde{B} : \{-1,+1\}^3 \to \{-1,+1\}$, které obvod realizuje platí

$$\begin{aligned}
\widetilde{B}\big((\ \ 1,\ \ 1,\ \ 1)^T\big) &= \ \ 1,\\
\widetilde{B}\big((\ \ 1,\ \ 1,-1)^T\big) &= \ \ 1,\\
\widetilde{B}\big((\ \ 1,-1,\ \ 1)^T\big) &= -1,\\
\widetilde{B}\big((\ \ 1,-1,-1)^T\big) &= \ \ 1,\\
\widetilde{B}\big((-1,\ \ 1,\ \ 1)^T\big) &= -1,\\
\widetilde{B}\big((-1,\ \ 1,-1)^T\big) &= -1,\\
\widetilde{B}\big((-1,-1,\ \ 1)^T\big) &= \ \ 1,\\
\widetilde{B}\big((-1,-1,-1)^T\big) &= -1.
\end{aligned}$$

Z tohoto příkladu je souvislost mezi hodnotami vstupních uzlů a hodnotami složek vektoru $\vec{\boldsymbol{y}}$ již zřejmá.

Dále definujme třídy prahových vektorů, počítaných polynomiálně velkými booleovskými obvody hloubky d.

Definice 2.3.4 *Předpokládejme, že máme zadánu posloupnost*

$$\bar{A} \stackrel{def}{=} \Big\{ \vec{\boldsymbol{y}}_n \in \{-1,+1\}^{2^n} \,|n \in \{1,\ldots,\infty\} \Big\},$$

a nechť $\widetilde{p}(x)$ je daný reálný polynom. Potom $\bar{A}$ leží ve třídě $\mathbf{LT}_{d,\widetilde{p}}$, jestliže pro každé $n \geq 1$ existuje booleovský obvod dimenze n, jehož hloubka je nejvýše d a velikost nejvýše $\widetilde{p}\,(n)$, takový, že počítá vektor $\vec{\boldsymbol{y}}_n$. Dále definujme třídu

$$\mathbf{LT}_d \stackrel{def}{=} \bigcup_{\{\widetilde{p}\,|\widetilde{p}\,je\ polynom\}} \mathbf{LT}_{d,\widetilde{p}}.$$

Nechť navíc pro příslušné booleovské obvody platí, že jejich váhy jsou omezeny zhora hodnotou $\widetilde{p}\,(n)$. Potom příslušnou třídu budeme značit $\widehat{\mathbf{LT}}_d$.

Definice 2.3.5 *Vektory tvořící sloupce paritní matice $\boldsymbol{B}^{(n)}$ budeme nazývat vektory parity.*

Definice 2.3.6 *Nechť* $\bar{A} \stackrel{def}{=} \left\{ \vec{\boldsymbol{y}}_n \in \{-1,+1\}^{2^n} \,|n \in \{1,\ldots,\infty\} \right\}$ *a nechť* $\widetilde{p}(x)$ *je daný reálný polynom. Potom řekneme, že* $\bar{A}$ *leží ve třídě* $\mathbf{PT}_{1,\widetilde{p}}$ *, existuje-li přirozené číslo* $k \leq \widetilde{p}\,(n)$ *a systém vektorů parity* $\vec{\boldsymbol{p}}^{\widetilde{(n)}}{}_1, \ldots, \vec{\boldsymbol{p}}^{\widetilde{(n)}}{}_k \in \{-1,+1\}^{2^n}$ *, tak, že vektor* $\vec{\boldsymbol{y}}_n$ *je jejich prahem. Dále definujme třídu*

$$\mathbf{PT}_1 \stackrel{def}{=} \bigcup_{\{\widetilde{p} | \widetilde{p} \, je \; polynom\}} \mathbf{PT}_{1,\widetilde{p}}.$$

Nechť navíc příslušné koeficienty w_i *jsou omezeny zhora hodnotou* $\widetilde{p}\,(n)$ *. Potom příslušnou třídu budeme značit* $\widehat{\mathbf{PT}}_1$ *.*

Jak již bylo řečeno v předešlé kapitole (viz. tvrzení 2.2.5), pro odhad počtu vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$, pro které je dané $\vec{\boldsymbol{y}}$ prahovým vektorem, lze zdola odhadnout na základě znalosti hodnot standardního skalárního součinu vektoru $\vec{\boldsymbol{y}}$ s vektory $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$.

Vlastnosti vektoru $\vec{\boldsymbol{g}}$, definovaného v následující lemmě, nám umožní od sebe odseparovat třídy $\mathbf{LT}_2$ a $\mathbf{PT}_1$.

Lemma 2.3.1 *Nechť* $n = 2k$, $i^\flat$ *je binární zápis čísla* i *a definujme vektor* $\vec{\boldsymbol{g}}^{(n)}$ *dimenze* 2^n *předpisem*

$$\vec{\boldsymbol{g}}_i^{(n)} \stackrel{def}{=} \begin{cases} -1 & \sum_{l=1}^{n} (i^\flat)_l \equiv 4 \mod 0,1 \\ \quad pro & \\ 1 & \sum_{l=1}^{n} (i^\flat)_l \equiv 4 \mod 2,3. \end{cases}$$

Potom pro libovolné vektory parity $\vec{\boldsymbol{p}}^{\widetilde{(n)}}$ *(t.j. sloupce matice* $\boldsymbol{B}^{(n)}$ *) platí*

$$\left\langle \vec{\boldsymbol{g}}^{(n)} \,\middle|\, \vec{\boldsymbol{p}}^{\widetilde{(n)}} \right\rangle \leq 2^{\frac{n}{2}}.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Nyní zavedeme další třídu prahových vektorů, kterou jak uvidíme, lze spočítat velmi snadno booleovským obvodem hloubky 2 lineární velikosti.

Definice 2.3.7 *Nechť* n *je přirozené číslo,* $\vec{\boldsymbol{y}} \in \{-1,+1\}^{2^n}$ *a* $\alpha \in \{0,\ldots,2^n-1\}$ *. Potom vektor* $\vec{\boldsymbol{y}}$ *je* α-SYMETRICKÝ VEKTOR *, právě když hodnota* $\vec{\boldsymbol{y}}_i$ *závisí pouze na čísle*

$$\sum_{k=1}^{n} (i^\flat)_k \cdot (\alpha^\flat)_k$$

(tedy pouze na součtu jedniček v binárním zápise i*, které jsou na těch pozicích, kde je v binárním zápisu čísla* α *hodnota 1). Dále budeme symbolem* $\mathbf{SYM}_2$ *označovat množinu všech* α*-symetrických vektorů pro* $\alpha \in \{0,\ldots,2^n-1\}$ *.*

Tvrzení 2.3.2 *Nechť* n *je přirozené číslo,* $\vec{\boldsymbol{y}} \in \{-1,+1\}^{2^n}$ *je* (2^n-1)*–symetrický vektor a nechť čísla* $d_1, \ldots, d_m$ *jsou délky po sobě jdoucích konstantních úseků v posloupnosti*

$$\vec{\boldsymbol{y}}_{2^1-1}, \vec{\boldsymbol{y}}_{2^2-1}, \ldots, \vec{\boldsymbol{y}}_{2^n-1}. \tag{2.3}$$

Dále pro $k \in \{1, \ldots, m\}$ *definujme funkce* $\widetilde{v_k} : \{-1, +1\}^n \to \{-1, +1\}$ *jako*

$$\widetilde{v_k}\,(\vec{\boldsymbol{x}}) \stackrel{def}{=} \widetilde{sgn}\left(\sum_{j=1}^{n} \vec{\boldsymbol{x}}_j + n - 2\sum_{j=1}^{k} d_j - \frac{1}{2}\right). \tag{2.4}$$

Potom platí, že

$$-\,\vec{\boldsymbol{y}}_1\left(1 - \sum_{j=1}^{m} (-1)^j\,(\widetilde{v_j}\,(\vec{\boldsymbol{x}}) - 1)\right) = \vec{\boldsymbol{y}}_{\vec{\boldsymbol{x}}^\sharp}. \tag{2.5}$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Toto tvrzení v podstatě popisuje booleovský obvod pro výpočet libovolného α-symetrického vektoru. Architektura tohoto vektoru je popsána na obrázku 2.3. Obsah předešlého tvrzení lze formulovat také následovně.

Důsledek 2.3.3 *Každý α-symetrický vektor je možno spočítat booleovským obvodem hloubky dva, který má lineárně omezené váhy a lineárně omezený počet vrcholů.*

Protože pro každé α je α-tý vektor parity α-symetrickým vektorem, lze vektory parity realizovat ve třídě $\mathbf{LT}_2$. Uvažujme booleovský obvod zobrazený na obrázku 2.3. Nechť $\bar{A}$ je množina těch vstupních vrcholů, pro které je $(\alpha^\flat)_i = 0$. Dále ohodnoťme veškeré hrany jdoucí z vrcholů v množině $\bar{A}$ hodnotou 0. Nechť všechny zbylé hrany jsou ohodnoceny způsobem popsaným v tvrzení 2.3.2. Potom takovýto obvod hloubky 2 zřejmě počítá α-tý vektor parity.

Následující dvě tvrzení popisují základní inkluze mezi třídami $\mathbf{LT}_2$, $\mathbf{PT}_1$, $\mathbf{LT}_3$, $\mathbf{SYM}_2$. Důkazy obou dvou tvrzení jsou analogické, založené na odhadu ve výrazu 2.2.

Jednotlivé inkluze jsou vesměs zřejmé z předešlého výkladu. Ostrost inkluzí se obecně dokáže tak, že pro vybraný vektor $\vec{\boldsymbol{y}}$ spočítáme jeho standardní skalární součin se základními vektory parity a tyto hodnoty použijeme v odhadu pro počet vektorů $\vec{\boldsymbol{x}}_i$ nezbytných k tomu, aby $\vec{\boldsymbol{y}}$ byl jejich prahem. Tímto způsobem vždy dokážeme, že počet vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ nemůže být polynomiálně pro daný vektor $\vec{\boldsymbol{y}}$ omezen.

Tvrzení 2.3.4 *Platí inkluze:*

1. $\mathbf{PT}_1 \subset \mathbf{LT}_3$ *a* $\widehat{\mathbf{PT}}_1 \subset \widehat{\mathbf{LT}}_3$,
2. $\mathbf{SYM}_2 \not\subset \mathbf{PT}_1$ *a* $\widehat{\mathbf{SYM}}_2 \not\subset \widehat{\mathbf{PT}}_1$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

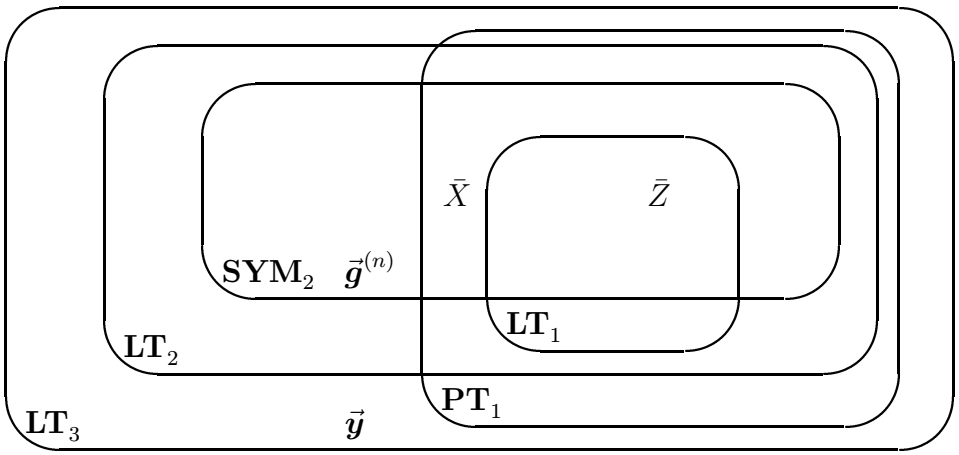
Tvrzení 2.3.5 $\mathbf{LT}_2 \subset \mathbf{LT}_3$ *a* $\mathbf{LT}_2 \neq \mathbf{LT}_3$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Nyní shrneme dosavadní poznatky o inkluzi tříd booleovských obvodů. Nechť $\vec{\boldsymbol{y}}$ je takový, že $\boldsymbol{M}_{\vec{\boldsymbol{y}}} = \boldsymbol{B}^{(n)}$, $n = 2k$, $\bar{Z}$ označuje množinu všech základních vektorů parity a $\bar{X}$ označuje množinu ostatních vektorů parity. Potom platí inkluze popsané následujícím diagramem:



Obrázek 2.3: Inkluze základních tříd booleovských obvodů.

Cvičení 2.3.0 Booleovské obvody

1. *Navrhněte booleovský obvod hloubky 2, počítající vektor $\vec{\boldsymbol{g}}^{(n)}$ definovaný v lemmě 2.3.1.*

2. *Navrhněte booleovský obvod hloubky 3, počítající vektor, jehož sdružená matice je rovna matici parity $\boldsymbol{B}^{(n)}$.*

2.4 Odhady velikosti vah prahového vektoru

Důležitou informací o booleovských obvodech je také odhad velikosti vah, nezbytných pro výpočet daného vektoru $\vec{\boldsymbol{y}}$. Dosud jsme se tímto problémem nezabývali, ukážeme ale, že úzce souvisí s horním a dolním odhadem počtu prahů systému vektoru $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$, kde alespoň jeden z vektorů $\vec{\boldsymbol{x}}_i$ je konstantní.

Protože pro daný váhový vektor prahu $\vec{\boldsymbol{y}}$ je jeho libovolný nenulový násobek opět váhovým vektorem pro práh $\vec{\boldsymbol{y}}$, mohou být velikosti vah libovolné. Nás ale zajímají velikosti vah především z hlediska posouzení toho, zda jsme schopni pro daný booleovský obvod jeho váhy uložit v paměti – zajímá nás tedy de fakto odhad na počet bitů nutných k uložení vah booleovského obvodu. Zřejmě vzhledem k hustotě racionálních čísel na reálné ose stačí brát v úvahu pouze racionální váhy. Proto můžeme bez újmy na obecnosti brát na zřetel pouze takové váhové vektory, které mají celočíselné složky, a zajímat se o

horní odhad velikosti těchto složek. Navíc většina tvrzení umožňuje zhora odhadnout tzv. celkovou váhu, tedy číslo $\sum_{i=1}^{S} |w_i|$, což je ale k verifikaci polynomiálního odhadu počtu bitů nutných k uložení vah postačující.

2.4.1 Dolní odhad – diskriminační lemma

Nejdříve vyslovíme tzv. diskriminační lemmu a na následujícím příkladě ukážeme její užitečnost pro odhad velikosti celkové váhy.

Lemma 2.4.1 *Nechť* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ *jsou z* $\{-1,+1\}^n$ *a vektor* $\vec{\boldsymbol{y}}$ *je prahem tohoto systému s celočíselnými koeficienty lineární kombinace* $w_1, \ldots, w_S$ *a* $w \stackrel{def}{=} \sum_{j=1}^{n} |w_j|$. *Nechť dále* $\boldsymbol{B}$ *je čtvercová matice řádu* n, *pro kterou platí, že*

$$(\forall k \in \{1, \ldots, n\})\,((\boldsymbol{B}\vec{\boldsymbol{y}})_k \cdot \vec{\boldsymbol{y}}_k \geq 0)$$

Potom platí:

$$w \cdot \max_{i \in \{1,\ldots,S\}} \{|\langle \vec{\boldsymbol{x}}_i \,|\boldsymbol{B}\vec{\boldsymbol{y}}\rangle|\} \geq \sum_{k=1}^{n} |(\boldsymbol{B}\vec{\boldsymbol{y}})_k|.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Význam diskriminační lemmy pro odhad velikosti vah spočívá v možnosti široké volby matice $\boldsymbol{B}$. Při vhodné volbě této matice můžeme pro daný systém vektorů dostat na základě tvrzení diskriminační lemmy dolní odhad na součet absolutních hodnot vah. Ilustrujme tuto skutečnost na následujících příkladech.

Příklad 2.4.1 *Nechť* $\boldsymbol{T}^{(k)}$ *je čtvercová matice řádu* 2^k *taková, že na hlavní diagonále a všude nad ní jsou jedničky a všude pod hlavní diagonálou jsou pouze* -1. *Potom lze velmi snadno ověřit (neboť* $\boldsymbol{M}_{\vec{\boldsymbol{p}}^{(\tilde{2}k)}{}_j} = \boldsymbol{E}^{(j-1)} \odot \left(\begin{smallmatrix} 1 & 1 \\ -1 & -1 \end{smallmatrix}\right) \odot \boldsymbol{E}^{(k-j)}$ *a* $\boldsymbol{M}_{\vec{\boldsymbol{p}}^{(\tilde{2}k)}{}_{k+j}} = \boldsymbol{M}^T_{\vec{\boldsymbol{p}}^{(\tilde{2}k)}{}_j}$*), že pro* i*-tý základní vektor parity* $\vec{\boldsymbol{p}}^{(\tilde{2}k)}{}_i$, $i \in \{1, \ldots, k\}$ *, platí*

$$\left\langle \boldsymbol{T}^{(k)} \middle| \boldsymbol{M}_{\vec{\boldsymbol{p}}^{(\tilde{2}k)}{}_i} \right\rangle = 2^{i+1}\,,\ \left\langle \boldsymbol{T}^{(k)} \middle| \boldsymbol{M}_{\vec{\boldsymbol{p}}^{(\tilde{2}k)}{}_{k+i}} \right\rangle = -2^{i+1}\,,\ \left\langle \boldsymbol{T}^{(k)} \middle| \boldsymbol{E}^{(k)} \right\rangle = 2^k.$$

Nechť $\vec{\boldsymbol{y}} \in \{-1,+1\}^{2^{2k}}$ *je takový vektor, že matice* $\boldsymbol{T}^{(k)}$ *je k němu sdružená. Na základě diskriminační lemmy a z předchozích rovností dostáváme při volbě matice* $\boldsymbol{B}$ *rovné identické matici, že celková váha* $\vec{\boldsymbol{y}}$ *, jakožto prahu základních vektorů parity, musí být alespoň rovna* 2^{k-1}.

Tento odhad ale zdaleka není optimální, ukážeme však v následujícím příkladě, že vhodnou volbou matice $\boldsymbol{B}$ ho lze podstatně zlepšit.

Příklad 2.4.2 *Nyní pro stejný vektor* $\vec{\boldsymbol{y}}$ *jako v příkladě předešlém provedeme odhad celkové váhy s využitím matice* $\boldsymbol{B}$ *, zkonstruované následovně. Nechť pomocná matice* $\boldsymbol{G}$ *je čtvercová řádu* 2^k, *mající těsně pod a nad hlavní diagonálou jedničky a na ostatních místech nuly s výjimkou prvku na pozici* $(1,1)$ *(při číslování řad od jedné), který je opět jednička. Nechť* $\vec{\boldsymbol{z}}$ *je vektor, pro který je matice* $\boldsymbol{G}$ *maticí sdruženou. Potom*

definujme matici $\boldsymbol{B}$ *jako diagonální matici, s vektorem* $\vec{\boldsymbol{z}}$ *na hlavní diagonále. Nyní spočítejme skalární součiny vektorů* $\tilde{\vec{\boldsymbol{p}}^{(n)}}_1, \ldots, \tilde{\vec{\boldsymbol{p}}^{(n)}}_n$, $n \stackrel{def}{=} 2k$, *s vektorem* $\boldsymbol{B}\vec{\boldsymbol{y}}$, *neboli spočteme skalární součiny matic* $\boldsymbol{M}_{\tilde{\vec{\boldsymbol{p}}^{(n)}}_i}$ *s maticí* $\boldsymbol{M}_{\boldsymbol{B}\vec{\boldsymbol{y}}}$. *Zřejmě ale matice* $\boldsymbol{M}_{\boldsymbol{B}\vec{\boldsymbol{y}}}$ *má těsně nad hlavní diagonálou 1 a těsně pod hlavní diagonálou* -1 *a na pozici* $(1,1)$ *jedničku. Odtud plyne, že všechny uvažované skalární součiny jsou rovny* -1, *protože matice sdružené k základním vektorům parity jsou po sloupcích (řádcích) konstantní, tedy skalární součiny jednotlivých řad těchto matic s odpovídajícími řadami matice* $\boldsymbol{M}_{\boldsymbol{B}\vec{\boldsymbol{y}}}$ *jsou nulové (v každém sloupci matice* $\boldsymbol{M}_{\boldsymbol{B}\vec{\boldsymbol{y}}}$ *je pouze jedna 1 a jedna* -1*) s výjimkou prvního a posledního sloupce (řádku). Skalární součin těchto dvou řad je roven* -1. *Stejně tak je* $\left\langle \boldsymbol{E}^{(k)} \middle| \boldsymbol{M}_{\boldsymbol{B}\vec{\boldsymbol{y}}} \right\rangle = 1$. *Navíc je zřejmé, že* $\sum_{k=1}^{2^n} |(\boldsymbol{B}\vec{\boldsymbol{y}})_k| = 2 \cdot 2^k - 1$, *tedy dostáváme z diskriminační lemmy nerovnost*

$$w \geq 2 \cdot 2^k - 1 = 2^{k+1} - 1.$$

Na složky vektoru $\vec{\boldsymbol{y}}$ z příkladu se můžeme dívat jako na výsledek porovnání velikosti indexů v matici $\vec{\boldsymbol{T}}^{(k)} = \boldsymbol{M}_{\vec{\boldsymbol{y}}}$. Lze tedy vyvodit závěr, že každý booleovský obvod hloubky 1, porovnávající dvě libovolná čísla s binárním zápisem délky nepřesahující hodnotu k, musí mít váhy se součtem absolutních hodnot alespoň $2^{k+1} - 1$. Tento odhad je optimální, neboť takovýto obvod lze snadno zkonstruovat. Vidíme tedy, že výsledky získané použitím diskriminační lemmy mohou být velmi rozdílné a jsou silně závislé na matici $\boldsymbol{B}$.

Cvičení 2.4.1 Diskriminační lemma

1. *Ověřte, že pro* $i \in \{1, \ldots, k\}$ *, platí*

$$\left\langle \boldsymbol{T}^{(k)} \middle| \boldsymbol{M}_{\tilde{\vec{\boldsymbol{p}}^{(2k)}}_i} \right\rangle = 2^{i+1}.$$

2. *Dokažte, že vektor* $\vec{\boldsymbol{y}}$ *definovaný v příkladě 2.4.1 není prahem systému základních vektorů parity (jinými slovy, k jeho výpočtu je nutno mít k dispozici nenulové prahy).*

3. *Navrhněte booleovský obvod hloubky 1, počítající vektor* $\vec{\boldsymbol{y}}$ *jako práh systému základních vektorů parity a konstantního vektoru, který má součet absolutních hodnot vah roven* $2^{k+1} - 1$.

2.4.2 Horní odhad velikosti vah a počtu prahových vektorů

Nyní obraťme pozornost na odhady počtu prahových vektorů klasického binárního perceptronu, t.j. perceptronu, jehož výstupem je hodnota $\widehat{sgn}\left(\sum_{i=1}^{S} \vec{\boldsymbol{w}}_i \vec{\boldsymbol{z}}_i - t\right)$, kde $\vec{\boldsymbol{z}}_i$ jsou vstupní hodnoty a t je prahová hodnota. Tento model perceptronu počítá prahové vektory systému základních vektorů parity a konstantního vektoru, jehož prostřednictvím je realizován práh. Výchozím pojmem k odhadu počtu takovýchto prahů je tzv. jádro prahu ([Hås94]), což je ve své podstatě v korespondenci s diskriminační lemmou minimální systém nerovnic,

$$\sum_{i=1}^{n} \vec{\boldsymbol{w}}_i \left(\vec{\boldsymbol{x}}_i\right)_j \leq \vec{\boldsymbol{y}}_j, \quad \text{pro} \quad \vec{\boldsymbol{y}}_j = -1,$$

$$\sum_{i=1}^{n} \vec{\boldsymbol{w}}_i \left(\vec{\boldsymbol{x}}_i\right)_j \geq \vec{\boldsymbol{y}}_j, \quad \text{pro} \quad \vec{\boldsymbol{y}}_j = 1,$$

který zaručuje, že jeho řešením jsou váhy pro práh $\vec{\boldsymbol{y}}$. Uvidíme, že pro práh $\vec{\boldsymbol{y}}$ vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ takovýto systém vždy existuje, nerovnice v něm nabývají rovnosti a počet

těchto nerovnic je roven právě číslu S. (Jinými slovy, matice $\boldsymbol{B}$ realizující přesný odhad velikosti vah v diskriminační lemmě musí obsahovat pro systém $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ alespoň S jedniček a na druhé straně takováto matice s právě S jedničkami existuje).

Definice 2.4.1 *Nechť $\vec{\boldsymbol{y}}$ je prahem systému vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ z $\{-1,+1\}^n$, $\boldsymbol{X}$ je matice, jejíž sloupce jsou tvořeny vektory $\vec{\boldsymbol{x}}_i$ a $\boldsymbol{\Omega}$ je matice o S sloupcích a 2^S řádcích, jejíž i-tý sloupec je roven i-tému základnímu vektoru parity $\vec{\boldsymbol{p}}^{\widetilde{(S)}}{}_i$. Nechť $\vec{\boldsymbol{\Theta}} \in \Re^S$ splňuje následující tři kritéria (zde funkce $\widetilde{sgn}$ je brána po složkách, $|\vec{\boldsymbol{z}}|$ je vektor tvořený absolutními hodnotami složek vektoru $\vec{\boldsymbol{z}}$ a $\vec{\mathbf{1}}$ označuje vektor tvořený jedničkami):*

1. $\widetilde{sgn}\left(\boldsymbol{X}\vec{\boldsymbol{\Theta}}\right) = \vec{\boldsymbol{y}}$
2. $\left|\boldsymbol{\Omega}\vec{\boldsymbol{\Theta}}\right| \geq \vec{\mathbf{1}}$
3. *počet složek vektoru $\left|\boldsymbol{\Omega}\vec{\boldsymbol{\Theta}}\right|$ rovných jedné je maximální přes všechny vektory $\vec{\boldsymbol{\Theta}}^*$, splňující předešlé dvě podmínky (je-li takovýchto $\vec{\boldsymbol{\Theta}}$ více, bereme v úvahu libovolný z nich).*

Každé $\vec{\boldsymbol{\Theta}}$ vyhovující předchozím podmínkám nazveme JÁDRO PRAHU *$\vec{\boldsymbol{y}}$.*

Existence vektoru $\vec{\boldsymbol{\Theta}}$ plyne z předpokladu, že $\vec{\boldsymbol{y}}$ je prahem systému $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$. Dále připomeňme, že i-tý řádek matice $\boldsymbol{\Omega}$ odpovídá binárnímu zápisu čísla i. Protože počet těchto řádků je roven 2^S, obsahují řádky matice $\boldsymbol{\Omega}$ všechny možné S–členné posloupnosti 1 a -1.

Lemma 2.4.2 *Předpokládejme, že $\vec{\boldsymbol{\Theta}}$ je jádrem prahového vektoru $\vec{\boldsymbol{y}}$ systému $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ a že matice $\boldsymbol{A}$ je podmatice matice $\boldsymbol{\Omega}$, tvořená těmi řádky $\vec{\boldsymbol{r}}^T$ matice $\boldsymbol{\Omega}$, pro které platí $\left|\vec{\boldsymbol{r}}^T\vec{\boldsymbol{\Theta}}\right| = 1$. Potom hodnost matice $\boldsymbol{A}$ je rovna číslu S.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

K odhadu počtu různých prahů systému základních vektorů parity a konstantního vektoru bude užitečná následující elementární lemma o determinantu matic obsahujících pouze ± 1.

Lemma 2.4.3 *Nechť $\boldsymbol{A}$ je čtvercová matice řádu m s prvky z množiny $\{-1,+1\}$. Potom* $\det \boldsymbol{A}$ *je dělitelný číslem 2^{m-1}.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Tvrzení 2.4.4 *Nechť $\vec{\boldsymbol{y}}$ je prahem systému vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ z $\{-1,+1\}^n$. Potom existují celočíselné váhy w_i pro $\vec{\boldsymbol{y}}$ tak, že pro všechna $i \in \{1, \ldots, n\}$ platí odhad*

$$|w_i| \leq \frac{S^{\frac{S}{2}}}{2^{S-1}}. \tag{2.6}$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Skutečnost, že váhy jsou dány jako řešení soustavy rovnic s koeficienty ± 1, umožňuje horní odhad počtu prahů.

Tvrzení 2.4.5 *Počet prahových vektorů systému* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ *je zhora omezen číslem* 2^{S^2+S}.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Ještě poznamenejme, že počet různých prahů pro systém základních vektorů parity o dimenzi 2^S a konstantního vektoru je roven počtu regulárních soustav $(S+1)$–lineárních rovnic s koeficienty ± 1 a pravou stranou tvořenou také ± 1.

2.4.3 Dolní odhad velikosti vah a počtu prahových vektorů

Nyní se krátce budeme věnovat dolnímu odhadu velikosti vah a počtu prahových vektorů. Uvidíme úzkou souvislost těchto dvou odhadů. K důkazu dolního odhadu počtu prahů je nutné tvrzení lemmy následující.

Lemma 2.4.6 *Nechť* $\vec{\boldsymbol{y}}$ *je prahem systému vektorů* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ *dimenze* n*, takových, že matice jejíž sloupce jsou rovny vektorům* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ *, má všechny řádky po dvou od sebe různé. Potom existují váhy* $\vec{\boldsymbol{w}}$ *takové, že*

$$(\forall i, j \in \{1, \ldots, n\}) \left(i \neq j \Rightarrow \left(\sum_{k=1}^{n} \vec{\boldsymbol{x}}_k \vec{\boldsymbol{w}}_k \right)_i \neq \left(\sum_{k=1}^{n} \vec{\boldsymbol{x}}_k \vec{\boldsymbol{w}}_k \right)_j \right).$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Na podkladě této lemmy lze provést dolní odhad počtu prahových vektorů ([YI65], [Smi65]). Tento dolní odhad provedeme indukcí podle počtu prahových vektorů, a to tak, že pro libovolný práh systému k-základních vektorů parity a konstantního vektoru (vše dimenze 2^k) zkonstruujeme 2^k+1 od sebe různých prahových vektorů systému základních vektorů parity a konstantního vektoru dimenze 2^{k+1}. Navíc pro různé prahové vektory dimenze 2^k budou takto zkonstruované prahy dimenze 2^{k+1} od sebe navzájem různé.

Z geometrického pohledu je podstata zmíněné konstrukce následující (berme v úvahu $k = 2$). Představme si v rovině čtverec (např. $\langle -1, 1\rangle \times \langle -1, 1\rangle$) a libovolnou přímku p. Dále předpokládejme situaci, kdy s přímkou p pohybujeme v jejím normálovém směru. Potom v každé poloze tato přímka rozděluje všechny vrcholy čtverce na dvě množiny. Zřejmě můžeme bez újmy na obecnosti předpokládat (viz. lemma 2.4.6), že nikdy nenastane situace, aby přímka p protínala současně dva vrcholy čtverce. Tedy pohybem takové přímky p ve směru jejího normálového vektoru dostaneme celkem $2^k + 1 = 5$ různých rozdělení

vrcholů čtverce. Nechť uvažovaný čtverec leží v pevně zvolené rovině trojrozměrného prostoru a tvoří jednu stěnu trojrozměrné krychle. Ponechme přímku p fixovanou v nějaké poloze. Rovinou obsahující opačnou stěnu krychle (označme tuto stěnu z) rovnoběžnou s původním čtvercem veďme přímku q, která je rovnoběžná s přímkou p a pohybujme přímkou q ve směru jejího normálového vektoru, který je rovnoběžný se stěnou z.

Zřejmě přímka q rozděluje vrcholy ležící na stěně z stejným způsobem jako přímka p v původním čtverci před svojí fixací. Evidentně rovina, obsahující obě přímky p a q rozděluje vrcholy krychle do dvou množin, a to právě $2^k + 1$ způsoby podle aktuální polohy přímky q. Zřejmě jsme pro pevně zvolenou přímku p dostali $2^k + 1$ různých rozdělení vrcholů krychle do množin $\bar{A}$, $\bar{B}$, $\bar{A} \cap \bar{B} = \emptyset$. Přitom je evidentní, že $\bar{A}$ a $\bar{B}$ jsou lineárně separovatelné právě rovinou obsahující přímky p a q.

Na druhé straně použijeme-li přímku $p^{'}$ takovou, že množiny vrcholů původního čtverce vymezené přímkou p jsou rozdílné od množin vrcholů původního čtverce vymezené přímkou $p^{'}$, dostáváme, že pro žádné dvě takto vzniklé množiny $\bar{A}^{'}$ a $\bar{B}^{'}$ nemůže platit $\bar{A} = \bar{A}^{'}$, neboť se musí lišit ve vrcholech krychle ležících v původním čtverci. Tato konstrukce je obsahem důkazu následujícího tvrzení a je ilustrována na obrázku 2.4.

Tvrzení 2.4.7 *Počet prahových vektorů systému* $\vec{\boldsymbol{p}}^{\tilde{(S)}}{}_1, \ldots, \vec{\boldsymbol{p}}^{\tilde{(S)}}{}_S, \vec{\boldsymbol{e}}^{(S)}$ *základních vektorů parity a jednotkového vektoru* $\vec{\boldsymbol{e}}^{(S)}$ *je zdola omezen číslem* $2^{\frac{S(S-1)}{2}}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Tohoto tvrzení lze přímo využít pro následující tvrzení, které říká, že existuje prahový vektor, jehož celková váha je exponenciální v S, což znamená, že k zápisu jeho vah je potřeba lineárního počtu bitů.

Tvrzení 2.4.8 *Existuje práh systému* $\vec{\boldsymbol{p}}^{\tilde{(S)}}{}_1, \ldots, \vec{\boldsymbol{p}}^{\tilde{(S)}}{}_S, \vec{\boldsymbol{e}}^{(S)}$ *základních vektorů parity a jednotkového vektoru* $\vec{\boldsymbol{e}}^{(S)}$ *, pro jehož všechny celočíselné váhové vektory* $\vec{\boldsymbol{w}}$ *platí odhad*

$$2^{\frac{S-2}{2}} \leq \sum_{k=1}^{S} |\vec{\boldsymbol{w}}_k|.$$

Navíc $\vec{\boldsymbol{w}}$ *lze volit tak, že platí*

$$\sum_{k=1}^{S} |\vec{\boldsymbol{w}}_k| \leq \frac{S^{\frac{S}{2}}}{2^S}.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

2.5 Otázka existence prahového vektoru

V této podkapitole budeme analyzovat problém, zda pro daný vektor $\vec{\boldsymbol{y}}$ a pro daný systém vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ lze efektivně rozhodnout otázku, zda $\vec{\boldsymbol{y}}$ je prahem zmíněného systému vektorů.

Pro tuto analýzu budeme potřebovat tvrzení Farkašovy věty, které zde nebudeme dokazovat a které vyslovíme v následujícím tvaru (důkaz lze nalézt v každé učebnici lineárního programování, pro studenty MI FJFI ČVUT doporučuji přednášku J. Pytlíček– Lineární programování):

Lemma 2.5.1 (Farkašovo) *Z následujících dvou soustav lineárních nerovnic*

1. $\boldsymbol{A}\vec{\boldsymbol{w}} \geq \vec{\boldsymbol{b}}$, $\vec{\boldsymbol{b}} \neq \vec{\boldsymbol{0}}$
2. $\vec{\boldsymbol{z}}^T\boldsymbol{A} < 0$, $\vec{\boldsymbol{z}^T}\vec{\boldsymbol{b}} > 0$,

má právě jedna nezáporné řešení.

Nyní vyslovíme nutnou a postačující podmínku pro to, aby vektor $\vec{\boldsymbol{y}}$ nebyl prahovým vektorem systému vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ ([Mur71]).

Tvrzení 2.5.2 *Vektor $\vec{\boldsymbol{y}}$ není prahem systému vektorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$ právě když soustava nerovnic*

$$\vec{\boldsymbol{z}}^T diag(\vec{\boldsymbol{y}})\boldsymbol{X} < \vec{\boldsymbol{0}}, \qquad \vec{\boldsymbol{z}}^T\vec{\boldsymbol{1}} > 0, \tag{2.7}$$

(kde $diag(\vec{\boldsymbol{y}})$ je diagonální matice s vektorem $\vec{\boldsymbol{y}}$ na hlavní diagonále) má nezáporné řešení.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Ekvivalence dokázaná v předchozím tvrzení má zásadní důsledek pro aplikaci booleovských obvodů, bohužel velmi negativní, protože následující problém

INSTANCE:LINEÁRNÍ PROGRAMOVÁNÍ
Celočíselná matice $\boldsymbol{A}$, celočíselné vektory $\vec{\boldsymbol{c}}$, $\vec{\boldsymbol{b}}$, odpovídajících dimenzí.

PROBLÉM:
Existuje racionální vektor $\vec{\boldsymbol{z}}$ tak, že platí $\vec{\boldsymbol{z}}^T\boldsymbol{A} \leq \vec{\boldsymbol{c}}, \quad \vec{\boldsymbol{z}}^T\vec{\boldsymbol{b}} \geq 0$?

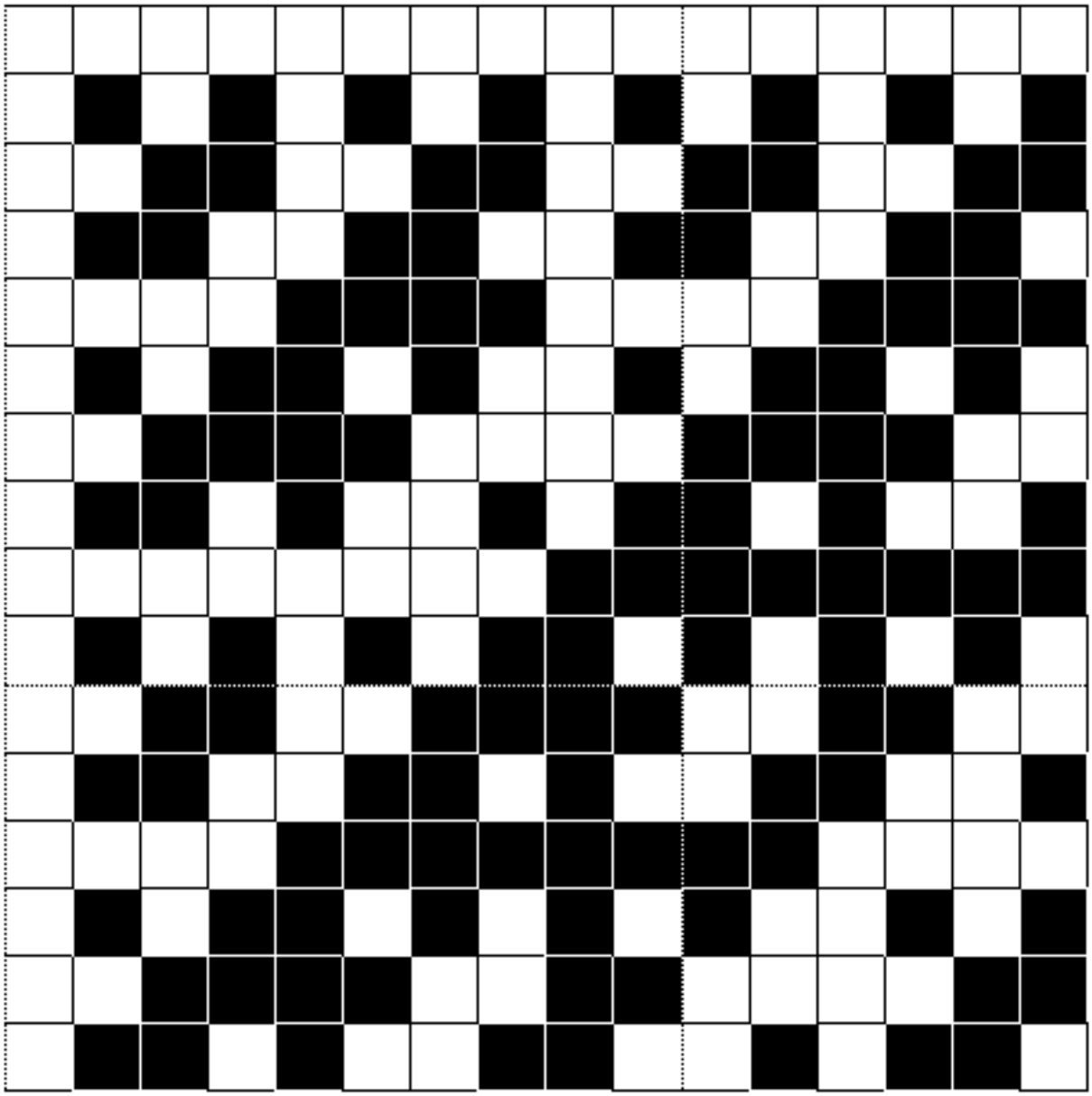
leží ve třídě NP–úplných problémů (viz [GJ79], str. 287–288). Tuto skutečnost lze na základě tvrzení 2.5.2 přeformulovat následovně:

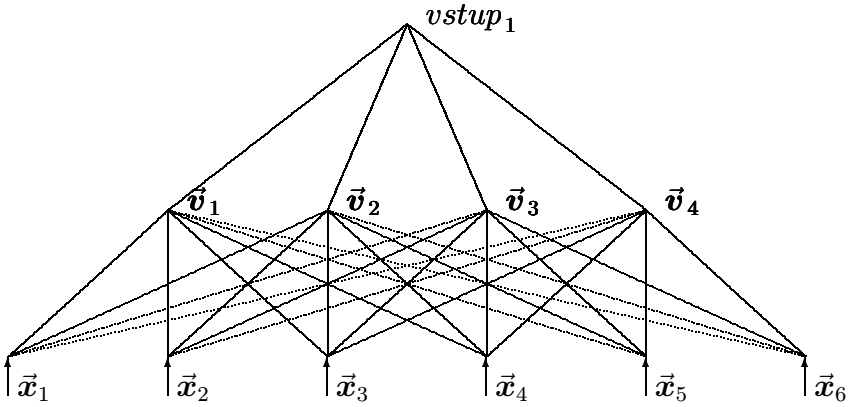
Tvrzení 2.5.3 *Následující problém*

INSTANCE:PRÁH SYSTÉMU VEKTORŮ
$\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S, \vec{\boldsymbol{y}} \in \{-1, +1\}^n$.

PROBLÉM:
Je vektor $\vec{\boldsymbol{y}}$ prahem systému $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_S$?

leží ve třídě NP–úplných problémů.





Obrázek 2.2: Booleovský obvod pro výpočet α-symetrických vektorů (pro $n = 5$ a $m = 3$). Uzly $\vec{x}_6$ a $\vec{v}_4$ slouží pouze k přenosu konstantního vstupu (prahové váhy). Váhy hran mezi vstupní a prostřední vrstvou jsou popsány vzorci 2.4, váhy mezi uzly prostřední vrstvy a výstupního uzlu jsou popsány vzorci 2.5.

Obrázek 2.4: Konstrukce 2^k+1 prahových vektorů odvozených z daného prahového vektoru pro $k = 2$.

Kapitola 3

Aproximační možnosti neuronových sítí

Finmanovo pravidlo

Nikomu se nechce pracovat se vzorci,
které vymyslel někdo jiný.

Pod aproximačními vlastnostmi umělých neuronových sítí budeme rozumět především schopnost neuronových sítí odpovídat na zadané vstupní hodnoty takovou výstupní hodnotou, která bude blízko hodnotě předem dané funkce v argumentu, který obsahuje vstupní hodnoty sítě.

Ukážeme, že je-li $\bar{K}$ kompaktní množina v $\Re^n$, pak neuronová síť s jednou skrytou vrstvou může uniformě aproximovat libovolnou spojitou funkci definovanou na $\bar{K}$. Tento problém je zajímavý zejména z toho důvodu, že nejčastější způsob aplikování neuronových sítí, t.j. oddělení (separace) více množin vzájemně od sebe, lze formulovat jako problém aproximace charakteristických funkcí těchto množin.

Jak bylo popsáno v úvodní kapitole, jediný perceptron dokáže rozdělit vstupní prostor $\Re^n$ do dvou poloprostorů. Zřejmě tedy síť sestávající z více perceptronů, které jsou paralelně vedle sebe a jejichž výstupy jsou vstupními hodnotami jediného výstupního neuronu (tzv. třívrstvá neuronová síť), dokáže z prostoru $\Re^n$ vydělit průnik m poloprostorů, kde m je počet neuronů v prostřední vrstvě.

Budeme-li mít disjunktní konečné množiny $\bar{A}, \bar{B} \in \Re^n$, pak zřejmě existují vzájemně po dvou disjunktní konvexní množiny $\bar{a}_1, \ldots, \bar{a}_l$, $\bar{b}_1, \ldots, \bar{b}_k$, z nichž každá je tvořena průnikem konečného počtu poloprostorů (je tedy polyedrem) a pro které platí, že $\bar{A} \subset \bigcup_1^l \bar{a}_i$ a $\bar{B} \subset \bigcup_1^k \bar{b}_j$.

Předpokládejme, že máme vícevrstvou neuronovou síť, tvořenou perceptrony, a že počet neuronů ve druhé vrstvě je roven počtu poloprostorů, nezbytných k vytvoření systému množin $\bar{a}_1, \ldots, \bar{a}_l$ a $\bar{b}_1, \ldots, \bar{b}_k$, a že každý z neuronů druhé vrstvy nabývá hodnoty ± 1 podle toho, na jaké straně daného poloprostoru je aktuální vstupní vektor. Chceme-li vytvořit síť, která pro vektory z $\bigcup_i^l \bar{a}_i$ dává hodnoty odlišné od vektorů z $\bigcup_i^k \bar{b}_j$, můžeme se na tento problém dívat jako na problém vytvoření konkrétní booleovské funkce, jejíž proměnné jsou hodnoty neuronů druhé vrstvy. Zřejmě funkce $\overset{\neg}{\text{NOT}}$, $\overset{\wedge}{\text{AND}}$ a $\overset{\vee}{\text{OR}}$ jsou všechny lineárně separabilní a tudíž vyjádřitelné pomocí perceptronu. Současně ale trojice

funkcí $\overset{\neg}{\text{NOT}}$, $\overset{\wedge}{\text{AND}}$ a $\overset{\vee}{\text{OR}}$ tvoří tzv. úplný systém výrokové logiky, což znamená, že jejich skládáním a uzávorkováním lze vyjádřit jakoukoli booleovskou funkci libovolného počtu proměnných. Z toho ale vyplývá, že vícevrstvá síť perceptronů, separující od sebe dvě dané konečné disjunktní množiny $\bar{A}$, $\bar{B}$ (obecněji dva disjunktní systémy disjunktních konvexních polyedrů) vždy existuje.

Otázkou zůstává horní a dolní odhad počtu neuronů ve vrstvách a počet těchto vrstev. V obecném případě aproximace funkce stejnoměrně spojité na kompaktu $\bar{K}$ můžeme vytvořit konečné ϵ-disjunktní pokrytí $\bar{D}$ tohoto kompaktu a aproximovanou funkci nahradit funkcí $\widetilde{\psi}$, po částech konstantní na tomto pokrytí. Zřejmě funkci $\widetilde{\psi}$ můžeme přesně aproximovat neuronovou sítí (disjunktní pokrytí $\bar{D}$ lze volit jako systém polyedrů) a ze stejnoměrné spojitosti plyne, že pro $\epsilon \to 0$ lze stejnoměrně spojitou funkci aproximovat libovolně přesně sítí sestávající se pouze z perceptronů. Otázkou však zůstává odhad na velikost takovéto sítě.

Obecně lze problém analýzy aproximačních vlastností neuronových sítí se třemi vrstvami formulovat jako problém vyjádření dané funkce $\widetilde{f} : \Re^n \to \Re$ ve tvaru součtu

$$\widetilde{f}\left(\vec{\boldsymbol{x}}\right) \overset{\text{def}}{=} \widetilde{g}\left(\sum_{j=1}^{l} \widetilde{\sigma}\left(\vec{\boldsymbol{w}}_j^T \vec{\boldsymbol{x}} + t_j\right)\right),$$

kde $\widetilde{g}$ a $\widetilde{\sigma}$ jsou funkce jedné reálné proměnné.

V následujících podkapitolách se tímto problémem budeme zabývat a ukážeme čtyři přístupy k analýze aproximačních vlastností neuronových sítí. Tyto přístupy jsou primárně založeny na Stone-Weierstrassově větě, vlastnostech duálních prostorů, využití Kolmogorovovy věty a konečně na aplikaci konvoluce funkcí.

3.1 Důsledky Stone-Weierstrasovy věty

Nyní ukážeme přímý přístup k analýze aproximačních vlastností ([Ell94]). Základní přístup je založen na následujících principech. Nejdříve aproximujeme spojitou funkci jedné reálné proměnné funkcí po částech konstantní, a to s libovolnou přesností. Následně se pokusíme zobecnit tento typ aproximace na funkce více proměnných. Následující pojem modulu spojitosti bude užitečný v aproximačních odhadech.

Definice 3.1.1 *Nechť* $\widetilde{f} : \bar{K} \to \Re$ *, kde* $\bar{K} \subset \Re^n$ *je kompaktní množina. Potom definujeme* MODUL SPOJITOSTI *(anglický termín: modulus of continuity) jako číslo*

$$\omega\left(\widetilde{f}, \delta\right) \overset{def}{=} \sup_{\substack{\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \bar{K} \\ \|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}\| < \delta}} \left|\widetilde{f}\left(\vec{\boldsymbol{x}}\right) - \widetilde{f}\left(\vec{\boldsymbol{y}}\right)\right|.$$

Je-li funkce $\widetilde{f}$ spojitá, pak hodnota $\omega\left(\widetilde{f}, \delta\right)$ je konečná a pro $\delta \to 0$ se blíží k nule také. Modul spojitosti v sobě zahrnuje různé způsoby vyjádření hladkosti funkce $\widetilde{f}$. Například je-li $\widetilde{f}$ Lipstichovská, tedy $\left|\widetilde{f}\left(\vec{\boldsymbol{x}}\right) - \widetilde{f}\left(\vec{\boldsymbol{y}}\right)\right| \leq L \cdot \|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}\|$, $L > 0$, $\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \bar{K}$, tak zřejmě platí $\omega\left(\widetilde{f}, \delta\right) \leq L\delta$. Modul spojitosti dává základní odhad, jak přesně lze danou funkci $\widetilde{f}$ aproximovat funkcí, která je po částech konstantní.

Následující tvrzení je založené na zřejmém faktu, že pro každou sigmoidální funkci $\widetilde{\sigma}$ (viz 1.2.1) platí, že pro $a \to +\infty$ funkce $\widetilde{\sigma}\left(ax\right)$ konverguje bodově k 1 pro $x > 0$ a k -1

pro $x < 0$. Tedy $\widetilde{\sigma}(ax)$ pro a rostoucí nade všechny meze konverguje bodově k prahové funkci $\widetilde{sgn}$ až na vyjímku v bodě 0. Z toho plyne, že funkce $\frac{1}{2}(\widetilde{\sigma}(ax) - \widetilde{\sigma}(a(x-1)))$ konverguje k charakteristické funkci intervalu $(0,1)$.

Tvrzení 3.1.1 *Nechť $\widetilde{\sigma}$ je sigmoidální funkce. Pak existuje konstanta c tak, že pro libovolné $\widetilde{f} \in C_{\langle 0,1\rangle}$ platí nerovnost*

$$\left\|\widetilde{f} - \widetilde{g}_n\right\|_\infty \leq c \cdot \omega\left(\widetilde{f}, \frac{1}{n}\right),$$

kde posloupnost funkcí $\{\widetilde{q}_n\}_1^\infty$ je definována předpisem

$$\widetilde{q_n}(x) \stackrel{def}{=} \widetilde{f}(0) + \sum_{v=1}^{n}\left(\widetilde{f}\left(\frac{v}{n}\right) - \widetilde{f}\left(\frac{v-1}{n}\right)\right)\widetilde{\sigma}(A_n(nx-v))$$

a pro číselnou posloupnost $\{A_i\}_1^\infty$ platí, že A_n je nejmenší kladné celé číslo, pro které platí

$$|\widetilde{\sigma}(x)| \leq \frac{1}{n} \quad pro \quad x \leq -A_n \qquad a \qquad 1 - \frac{1}{n} \leq \widetilde{\sigma}(x) \leq 1 + \frac{1}{n} \quad pro \quad x \geq A_n.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Následující větu, jejíž důkaz zde neuvedeme, (viz. např. [Miš89], str. 220), budeme potřebovat pro důkaz úplnosti všech sigmoidálních funkcí v prostoru všech funkcí spojitých na kompaktní množině.

Tvrzení 3.1.2 (Stone-Weierstrass) *Nechť $\mathcal{B}$ je lineární podprostor v prostoru $C_{\Re^n}$ takový, že obsahuje konstantní nenulovou funkci a v prostoru $\Re^n$ odděluje libovolné dva body (t.j. pro libovolné $\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \Re^n$, $\vec{\boldsymbol{x}} \neq \vec{\boldsymbol{y}}$ existuje $\widetilde{f} \in \mathcal{B}$ tak, že $\widetilde{f}(\vec{\boldsymbol{x}}) \neq \widetilde{f}(\vec{\boldsymbol{y}})$). Potom $\mathcal{B}$ je hustá v prostoru $C_{\Re^n}$.*

Před vlastním důkazem hustoty sigmoidálních funkcí v prostoru funkcí spojitých na kompaktní množině $\bar{K}$ dokážeme tvrzení pomocné, ukazující hustotu exponenciálních funkcí v prostoru $C_{\bar{K}}$.

Tvrzení 3.1.3 *Předpokládejme, že $\bar{K}$ je kompaktní podmnožina prostoru $\Re^n$. Potom lineární obal množiny funcí $\bar{\Theta} \stackrel{def}{=} \left\{\widetilde{\rho}(\vec{\boldsymbol{x}}) = e^{\langle \vec{\boldsymbol{a}} | \vec{\boldsymbol{x}} \rangle} \,|\, \vec{\boldsymbol{a}} \in \Re^n\right\}$ je hustý v prostoru $C_{\bar{K}}$.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Následující tvrzení ve své podstatě říká, že neuronová síť typu perceptronu s libovolnou sigmoidální funkcí $\widetilde{\sigma}$ může libovolně přesně aproximovat jakoukoli spojitou funkci na kompaktní množině.

Tvrzení 3.1.4 *Nechť $\bar{K}$ je kompaktní podmnožina prostoru $\Re^n$. Potom množina funkcí tvaru*

$$\widetilde{g}\,(\vec{\boldsymbol{x}}) \stackrel{def}{=} \sum_{j=1}^{k} a_j \widetilde{\sigma}\,(\langle \vec{\boldsymbol{w}}\,|\vec{\boldsymbol{x}}\rangle + c_j),$$

kde $\widetilde{\sigma}$ je spojitá sigmoidální funkce, je hustá v prostoru $C_{\bar{K}}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

3.2 Aplikace duálních prostorů

Jeden z dalších způsobů důkazu hustoty daného systému funkcí v množině funkcí spojitých na kompaktní množině je založen na funkcionálních vlastnostech duálních prostorů. Připomeňme stručně, že duální prostor normovaného vektorového prostoru $\mathcal{X}$ je lineární prostor, tvořený všemi omezenými lineárními funkcionály definovanými na prostoru $\mathcal{X}$. Každý duální prostor s normou definovanou jako

$$\left\|\widetilde{f}\right\| \stackrel{\text{def}}{=} \sup_{\substack{\vec{\boldsymbol{x}}\in\mathcal{X}\\ \|\vec{\boldsymbol{x}}\|=1}} \left|\widetilde{f}\,(\vec{\boldsymbol{x}})\right|,$$

je Banachův prostor, což znamená, že každá Cauchyovská posloupnost (vzhledem ke zmíněné normě) funkcionálů duálního prostoru má v tomto prostoru limitní prvek. Relevance duálních prostorů k hustotě dané množiny funkcí v prostoru všech spojitých funkcí je předmětem následující věty.

Tvrzení 3.2.1 *Nechť $\mathcal{X}$ je normovaný vektorový prostor nad tělesem $\Re$, $\bar{V}$ je jeho podprostor. Potom $\bar{V}$ je hustý v $\mathcal{X}$, právě když jediný lineární funkcionál na $\mathcal{X}$, v jehož nulovém prostoru je $\bar{V}$ obsažen, je nulový funkcionál.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Chceme-li tedy pro daný podprostor $\bar{V}$ prostoru $\mathcal{X}$ dokázat hustotu tohoto podprostoru v celém prostoru, stačí ukázat, že jediný funkcionál, který anihiluje celý podprostor $\bar{V}$ je nulový funkcionál.

Jako příklad aplikace duálních prostorů uvedeme následující analýzu v prostoru $\mathcal{L}_p^{\bar{K}}$ a $C_{\bar{K}}$.

3.2.1 Hustota sigmoidálních funkcí v prostoru $\mathcal{L}_p^{\bar{K}}$

Nyní využijeme předchozí tvrzení pro analýzu hustoty množiny všech sigmoidálních funkcí v prostoru funkcí, jejichž p-tá mocnina, $1 < p$, je Lesbegueovsky integrabilní na kompaktu $\bar{K} \subset \Re^n$.

Je všeobecně známým poznatkem funkcionální analýzy, že každý lineární funkcionál na prostoru $\mathcal{L}_p^{\bar{K}}$ lze vyjádřit jednoznačně ve tvaru integrálu

$$\int_{\bar{K}} \widetilde{f}\,(\vec{\boldsymbol{x}})\widetilde{g}\,(\vec{\boldsymbol{x}})d\,(\vec{\boldsymbol{x}}), \tag{3.1}$$

přičemž $\widetilde{g}$ je prvek z prostoru $\mathcal{L}_q^{\bar{K}}$, kde čísla p a q jsou spolu svázána vztahem $\frac{1}{p}+\frac{1}{q}=1$. Abychom ukázali, že množina funkcí tvaru

$$\widetilde{\sigma}\,(\langle \vec{\boldsymbol{w}}\,|\vec{\boldsymbol{x}}\rangle + c) \tag{3.2}$$

je hustá v $\mathcal{L}_p^{\bar{K}}$, dokážeme, že jediný anihilátor těchto funkcí je nulový funkcionál. Máme tedy dokázat následující implikaci:

$$(\forall\, \widetilde{\sigma}\,(\langle \vec{\boldsymbol{x}}\,|\vec{\boldsymbol{w}}\rangle + c))\left(\int_{\bar{K}} \widetilde{\sigma}\,(\langle \vec{\boldsymbol{x}}\,|\vec{\boldsymbol{w}}\rangle + c)\widetilde{g}\,(\vec{\boldsymbol{x}})d\,(\vec{\boldsymbol{x}}) = 0\right) \Rightarrow \widetilde{g}\,(\vec{\boldsymbol{x}}) = 0 \quad \text{skoro všude na} \quad \bar{K}.$$

Důkaz provedeme sporem, předpokládejme tedy, že $\widetilde{g}$ je anihilátor všech funkcí tvaru 3.2 a $\|\widetilde{g}\| > 0$. Nejdříve z důvodů jednoduššího zápisu dodefinujme funkci $\widetilde{g}$ nulou vně kompaktu $\bar{K}$. Protože funkcionál 3.1 anihiluje všechny funkce tvaru 3.2, musí anihilovat i funkce

$$\widetilde{\sigma_n}\,(\vec{\boldsymbol{x}}) \stackrel{\text{def}}{=} \widetilde{\sigma}\,(n\,(\langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{x}}\rangle - \langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{y}}\rangle)),$$

kde $\vec{\boldsymbol{y}}$ je libovolný vektor z $\bar{K}$ a $\vec{\boldsymbol{a}}$ je libovolný jednotkový vektor, $\|\vec{\boldsymbol{a}}\| = 1$. Zřejmě nadrovina $\langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{x}}\rangle - \langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{y}}\rangle = 0$ protíná množinu $\bar{K}$ v bodě $\vec{\boldsymbol{y}}$. Protože ale

$$\int_{\bar{K}} \widetilde{\sigma_n}\,(\vec{\boldsymbol{x}})\widetilde{g}\,(\vec{\boldsymbol{x}})d\,(\vec{\boldsymbol{x}}) = 0, \tag{3.3}$$

pro libovolné n dostáváme, že ($\widetilde{\sigma}_n$ konverguje bodově k -1 na množině $\langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{x}}\rangle < \langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{y}}\rangle$ a k $+1$ na množině $\langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{x}}\rangle > \langle \vec{\boldsymbol{a}}\,|\vec{\boldsymbol{y}}\rangle$) platí rovnost

$$\int_{\langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{x}}\rangle < \langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{y}}\rangle} \widetilde{g}\,(\vec{\boldsymbol{x}})d\,(\vec{\boldsymbol{x}}) = \int_{\langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{x}}\rangle > \langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{y}}\rangle} \widetilde{g}\,(\vec{\boldsymbol{x}})d\,(\vec{\boldsymbol{x}}),$$

protože jinak bychom na základě spojitosti Lebesgueova integrálu dostali spor s rovnicemi 3.3. Funkce s takovýmito vlastnostmi popisuje následující lemma:

Lemma 3.2.2 *Nechť $\widetilde{g}$ je funkce definovaná na kompaktní množině $\bar{K} \subset \Re^n$ splňující následující podmínku:*

$$(\forall \vec{\boldsymbol{a}} \in \Re^n,\ \|\vec{\boldsymbol{a}}\| = 1)\left(\forall \vec{\boldsymbol{y}} \in \bar{K}\right)\left(\int_{\langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{x}}\rangle < \langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{y}}\rangle} \widetilde{g}\,(\vec{\boldsymbol{x}})d\,(\vec{\boldsymbol{x}}) = \int_{\langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{x}}\rangle > \langle \vec{\boldsymbol{a}}|\vec{\boldsymbol{y}}\rangle} \widetilde{g}\,(\vec{\boldsymbol{x}})d\,(\vec{\boldsymbol{x}}).\right)$$

Potom $\widetilde{g}$ je skoro všude na $\bar{K}$ rovna nule.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Tedy jediný anihilátor sigmoidálních funkcí v prostoru $\mathcal{L}_p^{\bar{K}}$ je nulový funkcionál a proto lineární obal množiny sigmoidálních funkcí 3.2 je hustý v $\mathcal{L}_p^{\bar{K}}$.

3.2.2 Hustota v RBF funkcí v prostoru $C_{\bar{K}}$

Nyní budeme zkoumat hustotu sigmoidální funkce v prostoru $C_{\bar{K}}$, na kterém je definovaná suprémová norma $\left\|\widetilde{f}-\widetilde{g}\right\|_\infty \stackrel{\text{def}}{=} \sup_{\vec{\boldsymbol{x}}\in\bar{K}}\left|\widetilde{f}\left(\vec{\boldsymbol{x}}\right)-\widetilde{g}\left(\vec{\boldsymbol{x}}\right)\right|$.

Hustota v tomto funkcionálním prostoru umožňuje aproximovat spojité funkce na kompaktní množině lokálním způsobem, který zaručuje bodovou konvergenci, narozdíl od prostoru $\mathcal{L}_p^{\bar{K}}$, kde je zaručena pouze konvergence podle středu. Stejně jako v předchozím případě (pro důkaz hustoty sigmoidálních funkcí v $C_{\bar{K}}$) potřebujeme znát obecný tvar lineárního funkcionálu na prostoru $C_{\bar{K}}$. Za tímto účelem definujeme množinu $\bar{\Phi}_{\bar{K}}$ jako množinu všech reálných regulárně spočetně aditivních funkcí $\widetilde{\phi}$, zadaných na σ-algebře B, všech Borelovských množin na $\bar{K}$, majících konečnou totální variaci. Nyní vyslovíme Riesz-Markov-Katukaniho větu ([KA77]), popisující obecný tvar lineárního funkcionálu v $C_{\bar{K}}$.

Tvrzení 3.2.3 *Obecný tvar lineárního spojitého funkcionálu* $\widetilde{g}$ *na prostoru* $C_{\bar{K}}$ *lze zapsat ve tvaru*

$$\widetilde{g}\left(\widetilde{f}\right)=\int_{\bar{K}}\widetilde{f}\left(\vec{\boldsymbol{x}}\right)d\left(\widetilde{\phi}\left(\vec{\boldsymbol{x}}\right)\right),\qquad \widetilde{f}\in C_{\bar{K}},$$

kde $\widetilde{\phi}$ *je libovolný prvek z množiny* $\bar{\Phi}_{\bar{K}}$.

Jedněmi z dosti rozšířených přechodových funkcí používaných v různých modelech neuronových sítí jsou tzv. RBF funkce (anglický termín: Radial Basis Function). Tyto funkce umožňují lokální aproximaci spojitých funkcí. Jejich obecný tvar lze zavést například následující definicí.

Definice 3.2.1 *Nechť* $\widetilde{f}$ *je libovolná spojitá funkce taková, že* $\widetilde{f}:\Re\to\langle 0,1\rangle$, $\lim_{\alpha\to\pm\infty}\widetilde{f}\left(\alpha\right)=0$ *a* $\lim_{\alpha\to 0}\widetilde{f}\left(\alpha\right)=1$. *Potom každou funkci* $\widetilde{g}:\Re^n\to\Re$ *tvaru*

$$\widetilde{g}\left(\vec{\boldsymbol{x}}\right)\stackrel{def}{=}\widetilde{f}\left(\beta\|\vec{\boldsymbol{x}}-\vec{\boldsymbol{c}}\|\right),$$

kde $\vec{\boldsymbol{c}}$ *je libovolný vektor a* $\beta\in\Re$, *nazveme* RBF FUNKCÍ *odvozenou od funkce* $\widetilde{f}$.

Nechť $\widetilde{\phi}$ je nenulová funkce z množiny $\bar{\Phi}_{\bar{K}}$. Jelikož $\widetilde{\phi}$ má nenulovou konečnou variaci, existuje bod $\vec{\boldsymbol{x}}$ takový, že, bez újmy na obecnosti, funkce $\widetilde{\phi}$ je kladná na nějakém kruhovém okolí bodu $\vec{\boldsymbol{x}}$. Pak ale existuje pro dané $\widetilde{f}$ z definice RBF vektor $\vec{\boldsymbol{c}}$ a číslo β tak, že

$$\int_{\Re^n}\widetilde{f}\left(\beta\|\vec{\boldsymbol{x}}-\vec{\boldsymbol{c}}\|\right)\widetilde{\phi}\left(\vec{\boldsymbol{x}}\right)d\left(\vec{\boldsymbol{x}}\right)>0,$$

(vně daného okolí $\widetilde{f}\left(\beta\|\vec{\boldsymbol{x}}-\vec{\boldsymbol{c}}\|\right)$ konverguje k nule, na okolí konverguje k 1 a $\widetilde{\phi}$ má konečnou variaci, tedy i konečný integrál) což ale znamená, že $\widetilde{\phi}$ není anihilátorem RBF funkcí odvozených od funkce $\widetilde{f}$. Proto je lineární obal RBF funkcí odvozených od funkce $\widetilde{f}$ hustý v prostoru $C_{\bar{K}}$.

3.2.3 Hustota sigmoidálních funkcí v prostoru $C_{\langle a,b\rangle}$

Pro prostor $C_{\langle a,b\rangle}$ platí Riesz-Markov-Katukaniho věta 3.2.3, obecný tvar funkcionálu je tedy opět zadán integrálem. Jedním ze způsobů důkazu hustoty sigmoidálních funkcí v prostoru $C_{\langle a,b\rangle}$ je adaptace důkazu hustoty těchže funkcí v prostoru $\mathcal{L}_p^{\bar{K}}$. Zde však ukážeme přímý přístup, který je ve své podstatě totožný s důkazem pro prostor $\mathcal{L}_p^{\bar{K}}$, ale

díky jednorozměrnosti daného definičního oboru nejsme nuceni použít vlastností Fourierovy transformace.

Mějme funci $\widetilde{g}$ z množiny $\bar{\Phi}_{\langle a,b\rangle}$, která je anihilátorem všech sigmoidálních funkcí. Zvolme libovolná přirozená čísla p,q a n tak, aby platilo, že $\frac{p}{q} \in \langle a,b\rangle$ (bez újmy na obecnosti předpokládejme, že $0 < a < b$). Protože $\widetilde{g}$ je anihilátorem sigmoidálních funkcí, platí rovnost

$$\int_a^b \widetilde{\sigma}\left(n\left(x-\frac{p}{q}\right)\right) d\left(\widetilde{g}\left(x\right)\right) = 0.$$

Protože ale funkce $\widetilde{\sigma}\left(n\left(x-\frac{p}{q}\right)\right)$ konverguje pro n rostoucí nade všechny meze k funkci $\widetilde{sgn}\left(x-\frac{p}{q}\right)$, platí, že

$$0 = \lim_{n\to+\infty}\left(\int_a^b \widetilde{\sigma}\left(n\left(x-\frac{p}{q}\right)\right) d\left(\widetilde{g}\left(x\right)\right)\right) = \int_a^b \widetilde{sgn}\left(x-\frac{p}{q}\right) d\left(\widetilde{g}\left(x\right)\right) =$$

$$= -\int_a^{\frac{p}{q}} d\left(\widetilde{g}\left(x\right)\right) + \int_{\frac{p}{q}}^b d\left(\widetilde{g}\left(x\right)\right).$$

Opět tedy dostáváme, že pro libovolný bod c intervalu $\langle a,b\rangle$ platí rovnost

$$\int_a^c d\left(\widetilde{g}\left(x\right)\right) = \int_c^b d\left(\widetilde{g}\left(x\right)\right).$$

Odtud pro libovolná $d, e \in \langle a,b\rangle$ platí

$$\int_d^e d\left(\widetilde{g}\left(x\right)\right) = 0,$$

a proto funkce $\widetilde{g}$ je identicky rovna nule skoro všude na celém intervalu $\langle a,b\rangle$ (stále používáme teorii Lebesgueova integrálu). Z toho ale plyne, že anihilátor sigmoidálních funkcí v prostoru $C_{\langle a,b\rangle}$ je pouze nulový funkcionál, a tedy lineární obal sigmoidálních funkcí je hustý v $C_{\langle a,b\rangle}$.

Cvičení 3.2.3 Aplikace duálních prostorů

1. ***Definice 3.2.2*** *Nechť $\widetilde{f}$ je libovolná spojitá funkce taková, že $\widetilde{f}:\Re\to\langle 0,1\rangle$, $\lim_{\alpha\to\pm\infty}\widetilde{f}\left(\alpha\right)=0$ a $\lim_{\alpha\to 0}\widetilde{f}\left(\alpha\right)=1$. Potom každou funkci $\widetilde{g}:\Re^n\to\Re$ tvaru*

$$\widetilde{g}\left(\vec{\boldsymbol{x}}\right) \stackrel{def}{=} \widetilde{f}\left(\beta\langle\vec{\boldsymbol{x}}\,|\vec{\boldsymbol{c}}\rangle\right),$$

kde $\vec{\boldsymbol{c}}$ je libovolný vektor a $\beta\in\Re$, nazveme SEMILOKÁLNÍ FUNKCÍ *odvozenou od funkce $\widetilde{f}$.*

Rozhodněte, zda lineární obal semilokálních funkcí odvozených od nějaké funkce $\widetilde{f}$ je hustý v prostorech $C_{\bar{K}}$, $C_{\Re^n}$, $\mathcal{L}_p^{\bar{K}}$ a $\mathcal{L}_p^{\Re^n}$.

2. *Rozhodněte, zda lineární obal RBF funkcí odvozených od funkce $\widetilde{f}$ je hustý v $\mathcal{L}_p^{\bar{K}}$ a $\mathcal{L}_p^{\Re^n}$.*

3.3 Analýza na základě Kolmogorovy věty

V této sekci ukážeme přímý přístup k analýze aproximačních možností třívrstvé perceptronové sítě (vstupní a výstupní vrstva plus tzv. skrytá vrstva mezi nimi) s jedním výstupním neuronem, zobrazující $\Re^n$ do $\Re$. K tomu využijeme tzv. Kolmogorovu větu (1957), která vyjadřuje každou reálnou funkci n reálných proměnných, pomocí superpozice konečného počtu monotónních rostoucích funkcí jedné reálné proměnné. Zmíněnou Kolmogorovu větu uvedeme v následující podobě ([Spr93]):

Tvrzení 3.3.1 *Pro každé přirozené $n \geq 2$ existuje $n \times (2n+1)$ spojitých funkcí $\widetilde{\psi}_{pq}$ takových, že pro libovolnou reálnou funkci $\widetilde{f} : \Re^n \to \Re$ existuje $2n$ spojitých funkcí $\widetilde{\Phi}_q$ tak, že pro libovolné $\vec{\boldsymbol{x}} \in \Re^n$ platí rovnice*

$$\widetilde{f}\,(\vec{\boldsymbol{x}}) = \sum_{q=0}^{2n} \widetilde{\Phi}_q \left(\sum_{p=1}^{n} \widetilde{\psi_{pq}}\,(\vec{\boldsymbol{x}}_p) \right).$$

Důkaz této věty neuvedeme pro jeho příliš velkou technickou náročnost. Původní důkaz navíc není relevantní k tematice, které je tato kapitola věnovaná. Kolmogorova věta v podobě právě zmíněné nemá přímý vztah k aproximačním vlastnostem vrstevnatých architektur neuronových sítí.

3.3.1 Konstrukce aproximačního jádra $\widetilde{\psi}$

V původním znění Kolmogorovy věty je spojitá funkce vyjádřena jako součet funkcí, přičemž jejich argumenty jsou hodnoty dané součtem funkcí $\widetilde{\psi}_{pq}$. Funkce $\widetilde{\psi}_{pq}$ mohou být obecně vzájemně různé, což by v korespondenci s funkcemi implementovanými prostřednictvím neuronových sítí znamenalo, že každý neuron má jinou přechodovou funkci $\widetilde{\psi}_{pq}$. Ukážeme ale, že lze zkonstruovat funkci $\widetilde{\phi}$, která umožní vyjádřit spojitou funkci ve tvaru, odpovídajícímu zápisu funkcí, realizovaných třívrstvou neuronovou sítí s předchodovou funkcí $\widetilde{\psi}$.

Definice 3.3.1 *Definujme následující intervaly*

$$\bar{D}_0 \stackrel{def}{=} \left\langle 0, \frac{1}{5!} \right\rangle \quad a \quad \bar{D}_1 \stackrel{def}{=} \left\langle 0, 1 + \frac{1}{5!} \right\rangle.$$

Nejdříve ukážeme konstrukci pomocné funkce $\widetilde{\psi^*} : \bar{D}_0 \to \Re$, indukcí podle k, kde za počáteční hodnotu k vezmeme hodnotu 5. Předpokládejme, že máme dánu posloupnost racionálních čísel $\{\rho_k\}_5^\infty$, která vyhovuje podmínkám

$$\rho_5 < \frac{1}{2 \cdot 5!} \quad \text{a} \quad 0 < \rho_{k+1} < \frac{\rho_k}{k} \quad \text{pro} \quad k \geq 5. \tag{3.4}$$

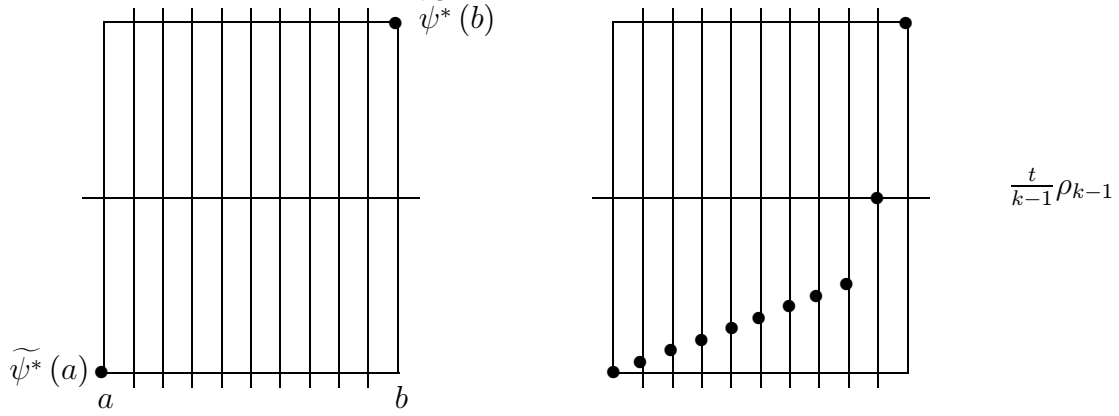
Nejdříve definujme hodnoty funkce $\widetilde{\psi^*}$ v bodech $\frac{j}{6!}$, $j \in \{0, \ldots, 6\}$ jako

$$\begin{gathered} \widetilde{\psi^*}\,(0) \stackrel{\text{def}}{=} 0 \ , \ \widetilde{\psi^*}\left(\tfrac{1}{6!}\right) \stackrel{\text{def}}{=} \tfrac{1}{5}\rho_5 \ , \ \widetilde{\psi^*}\left(\tfrac{2}{6!}\right) \stackrel{\text{def}}{=} \tfrac{2}{5}\rho_5 \ , \ \widetilde{\psi^*}\left(\tfrac{3}{6!}\right) \stackrel{\text{def}}{=} \tfrac{3}{5}\rho_5 \ , \\ \widetilde{\psi^*}\left(\tfrac{4}{6!}\right) \stackrel{\text{def}}{=} \tfrac{4}{5}\rho_5 \ , \ \widetilde{\psi^*}\left(\tfrac{5}{6!}\right) \stackrel{\text{def}}{=} \tfrac{1}{2\cdot 5!} \ , \ \widetilde{\psi^*}\left(\tfrac{1}{5!}\right) \stackrel{\text{def}}{=} \tfrac{1}{5!} \ . \end{gathered} \tag{3.5}$$

Předpokládejme, že hodnoty $\widetilde{\psi^*}\left(\frac{i}{(k-1)!}\right)$ jsou již definovány. Položme $j \stackrel{\text{def}}{=} ki + t$, kde $t \in \{0, \dots, k-1\}$, je-li $i \in \{0, \dots, \frac{k!-1}{5!}\}$ a $t \stackrel{\text{def}}{=} 0$, je-li $i = \frac{k!}{5!}$, a definujme

$$\widetilde{\psi^*}\left(\frac{j}{k!}\right) = \widetilde{\psi^*}\left(\frac{i}{(k-1)!} + \frac{t}{k!}\right) \stackrel{\text{def}}{=} \begin{cases} \widetilde{\psi^*}\left(\frac{i}{(k-1)!}\right) + \frac{t}{k-1}\rho_{k-1} & \text{pro} \quad t \in \{0, \dots, k-2\}, \\ \frac{1}{2}\left\{\widetilde{\psi^*}\left(\frac{i}{(k-1)!}\right) + \widetilde{\psi^*}\left(\frac{i+1}{(k-1)!}\right)\right\} & \text{pro} \quad t = k-1. \end{cases} \tag{3.6}$$

Tedy funkci $\widetilde{\psi^*}$ definujeme rekurzivně na intervalech $\left\langle \frac{i}{(k-1)!}, \frac{i+1}{(k-1)!} \right\rangle$ a to tak, že hodnoty uvnitř tohoto intervalu závisí pouze na funkčních hodnotách $\widetilde{\psi^*}$ v krajních bodech tohoto intervalu. Interval $\left\langle \frac{i}{(k-1)!}, \frac{i+1}{(k-1)!} \right\rangle$ ekvidistantně rozdělíme na k částí a v bodech, které interval $\left\langle \frac{i}{(k-1)!}, \frac{i+1}{(k-1)!} \right\rangle$ rozdělují, definujeme nové funkční hodnoty vždy o $\frac{1}{k-1}\rho_{k-1}$ větší než již definovaná hodnota v levém sousedním bodě, s vyjímkou bodu nejvíce napravo, jemuž přiřadíme funkční hodnotu rovnu aritmetickému průměru hodnot $\widetilde{\psi^*}$ v krajních bodech intervalu $\left\langle \frac{i}{(k-1)!}, \frac{i+1}{(k-1)!} \right\rangle$. Tuto konstrukci ilustruje obrázek 3.1.



Obrázek 3.1: Konstrukce funkce $\widetilde{\psi^*}$.

Z vlastností číselné posloupnosti $\{\rho_k\}_5^\infty$ plyne, že posloupnost takto konstruovaných hodnot funkce $\widetilde{\psi^*}$ je monotonně rostoucí vzhledem ke svým argumentům. Zřejmě množina

$$\left\{ \frac{i}{k!} \,|k \in \{1, \dots, \infty\} \,,\; i \in \{0, \dots, k!\} \right\}$$

je hustá v intervalu $\left\langle 0, \frac{1}{5!} \right\rangle$, máme tedy definovánu funkci $\widetilde{\psi^*}$ na husté podmnožině $\left\langle 0, \frac{1}{5!} \right\rangle$. Na doplňku této podmnožiny do intervalu $\left\langle 0, \frac{1}{5!} \right\rangle$ lze zřejmě funkci $\widetilde{\psi^*}$ spojitě dodefinovat.

S využitím funkce $\widetilde{\psi^*}$ nyní definujme funkci $\widetilde{\psi} : \bar{D}_1 \to \Re$ podle následujícího předpisu:

$$\widetilde{\psi}(x) \stackrel{\text{def}}{=} \begin{cases} \widetilde{\psi^*}(x) & \text{pro} \quad x \in \bar{D}_0, \\ \widetilde{\psi^*}\left(x - \frac{i}{5!}\right) + \frac{i}{5!}, & \text{pro} \quad x \in \left\langle \frac{i}{5!}, \frac{i+1}{5!} \right\rangle, \quad i \in \{0, \dots, 5!\}. \end{cases} \tag{3.7}$$

Evidentně $\widetilde{\psi}$ je monotonně rostoucí, zobrazující interval $\bar{D}_1$ na sebe sama. Základní vlastnost funkce $\widetilde{\psi}$ vystihuje následující lemma.

Lemma 3.3.2 *Nechť platí*

$$\delta_k \stackrel{def}{=} \frac{1}{k!} - \sum_{r=k+1}^{\infty} \frac{1}{r!} \quad a \quad \nu_k \stackrel{def}{=} \sum_{r=k}^{\infty} \frac{r-1}{r} \rho_r.$$

Potom platí

$$\widetilde{\psi}\left(\frac{i}{k!} + \delta_k\right) = \widetilde{\psi}\left(\frac{i}{k!}\right) + \nu_k.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Ještě uveďme lemmu, která bude nutná pro důkaz tvrzení o hustotě.

Definice 3.3.2 *Posloupnost reálných čísel* $\{\lambda_i\}_1^\infty$ *je* CELOČÍSELNĚ NEZÁVISLÁ *, jestliže pro libovolnou konečnou posloupnost celých čísel* $t_1, \ldots, t_n$ *splňujících podmínku*

$$\sum_{p=1}^{n} |t_p| \neq 0$$

platí, že $\sum_{p=1}^{n} t_p \lambda_p \neq 0$. *Dále nechť pro každé* $k \geq 5$ *symbol* $\bar{I}_k$ *označuje množinu celých čísel*

$$\bar{I}_k \stackrel{def}{=} \left\{ i \in \{0, \ldots, +\infty\} \,\middle|\, 0 \leq i \leq \left(1 + \frac{1}{5!}\right) k! \right\}.$$

Lemma 3.3.3 *Nechť* $\{\lambda_i\}_1^\infty$ *,* $\lambda_i > 0$ *je celočíselně nezávislá posloupnost a nechť pro každé* $k \geq 5$ *je*

$$\mu_k \sigma_k \stackrel{def}{=} \min_{i_p, j_p \in \bar{I}_k} \left\{ \left| \sum_{p=1}^{k-3} \lambda_p \left[\widetilde{\psi}\left(\frac{i_p}{k!}\right) - \widetilde{\psi}\left(\frac{j_p}{k!}\right) \right] \right| \,\middle|\, \sum_{p=1}^{k-3} |i_p - j_p| \neq 0 \right\}, \tag{3.8}$$

kde je

$$\sigma_k \stackrel{def}{=} 1 + \sum_{p=1}^{k-3} \lambda_p.$$

Potom je možno volit výchozí racionální hodnoty $\{\rho_k\}_5^\infty$ *ve výrazu 3.4 tak, aby platilo* $\nu_k < \mu_k$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

3.3.2 Konstrukce rekurzivního pokrytí jednotkové krychle

Na základě předchozího výkladu definujme systémy intervalů $\bar{E}_{k,i}$ a $\bar{E}_{k,i}^q$ podle následujících definic. Základní vlastnosti těchto intervalů jsou obsaženy v následujících dvou lemmách.

Definice 3.3.3 *Pro každé přirozené $k \geq 5$ definujme systém uzavřených intervalů*

$$\bar{E}_{k,i} \stackrel{def}{=} \left\langle \frac{i}{k!}, \frac{i}{k!} + \delta_k \right\rangle, \quad i \in \bar{I}_k.$$

Zřejmě tyto intervaly jsou pro každou pevnou hodnotu $k \geq 5$ od sebe vzdáleny mezerou délky

$$e_k \stackrel{\text{def}}{=} \frac{1}{k!} - \delta_k = \sum_{r=k+1}^{\infty} \frac{1}{r!} \tag{3.9}$$

a dá se přímým výpočtem ověřit, že mezi délkou těchto mezer a intervalů platí vztah

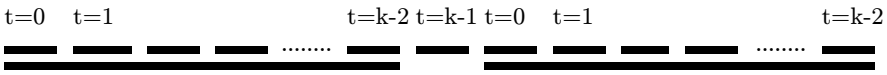
$$\delta_k > (k-1)e_k. \tag{3.10}$$

S vyjímkou těchto mezer intervaly $\bar{E}_{k,i}$ pokrývají interval $\bar{D}_1$. Navíc se snadno ověří, že platnost inkluze

$$\bar{E}_{k,j} \subset \bar{E}_{k-1,i}$$

je ekvivalentní s podmínkou, že $j = ki + t,\ \ t \in \{0, \ldots, k-2\}$. Jestliže tato inkluze platí, pak počáteční body $\bar{E}_{k,j}$ a $\bar{E}_{k-1,i}$ jsou totožné pro $t = 0$ a jejich koncové body jsou totožné pro $t = k-2$. Pro $t = k-1$ interval $\bar{E}_{k,j}$ leží v mezeře, oddělující intervaly $\bar{E}_{k,j}$ a $\bar{E}_{k-1,i}$. Navíc z monotonie funkce $\widetilde{\psi}$ a z lemmy 3.3.2 plyne

$$\widetilde{\psi}\left(\left\langle \frac{i}{k!}, \frac{i}{k!} + \delta_k \right\rangle\right) = \left\langle \widetilde{\psi}\left(\frac{i}{k!}\right), \widetilde{\psi}\left(\frac{i}{k!}\right) + \nu_k \right\rangle.$$



Obrázek 3.2: Vlastnosti intervalů $\bar{E}_{k,j} \subset \bar{E}_{k-1,i}$ ($\bar{E}_{k-1,i}$ dole, $\bar{E}_{k,j}$ nahoře).

Definice 3.3.4 *Pro libovolná přirozená $k \geq 5$, $n \geq 2$ a $q \in \{0, \ldots, 2n\}$ definujme systém uzavřených intervalů*

$$\bar{E}^q_{k,i} \stackrel{def}{=} \left\langle \frac{i}{k!} - qe_{2n+1}, \frac{i}{k!} + \delta_k - qe_{2n+1} \right\rangle, \quad i \in \bar{I}_k,$$

kde $i \in \bar{I}_k$.

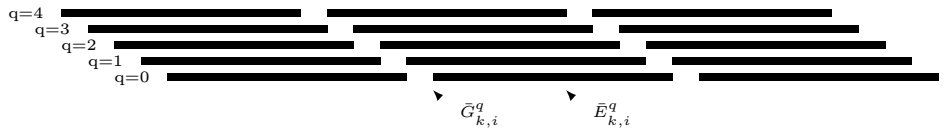
Intervaly $\bar{E}^q_{k,i}$ jsou de facto rovny intervalům $\bar{E}_{k,i}$, posunutým o hodnotu $-qe_{2n-1}$.

Lemma 3.3.4 *Nechť je dáno celé číslo $n \geq 2$. Potom pro každé celé číslo $k \geq 2n+1$ a pro každé $q \in \{0, \ldots, 2n\}$ platí*

1. *délka intervalu $\bar{E}^q_{k,i}$ se blíží k nule pro $k \to \infty$.*
2. *Kdykoli je $i \neq j$, pak*

$$\bar{E}^q_{k,i} \cap \bar{E}^q_{k,j} = \emptyset. \tag{3.11}$$

Dále pro každý bod x z intervalu $\bar{D}_1$ existuje alespoň $2n$ hodnot q takových, že $x \in \bar{E}^q_{k,i}$.



Obrázek 3.3: Intervaly $\bar{E}^q_{k,i}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Lemma 3.3.5 *Nechť jsou dána celá čísla* $n \geq 2$, $q \in \{0, \ldots, 2n\}$, $k \geq 2n+1$ *a pro každé* $p \in \{1, \ldots, n\}$ *nechť je* $i_p \in \bar{I}_k$. *Dále nechť posloupnost čísel* $\{\lambda_i\}_1^n$ *je celočíselně nezávislá. Dále definujme následující intervaly*

$$\bar{T}^q_{k,i_1,\ldots,i_n} \stackrel{def}{=} \left\langle \sum_{p=1}^{n} \lambda_p \widetilde{\psi}\left(\frac{i_p}{k!} + qe_{2n+1}\right), \sum_{p=1}^{n} \lambda_p \left[\widetilde{\psi}\left(\frac{i_p}{k!} + qe_{2n+1}\right) + \nu_k\right]\right\rangle. \tag{3.12}$$

Potom

1. *Velikost intervalů* $\bar{T}^q_{k,i_1,\ldots,i_n}$ *se blíží k nule pro* $k \to \infty$.
2. *Kdykoli je* $(i_1, \ldots, i_n) \neq (j_1, \ldots, j_n)$, *tak*

$$\bar{T}^q_{k,i_1,\ldots,i_n} \cap \bar{T}^q_{k,j_1,\ldots,j_n} = \emptyset. \tag{3.13}$$

Platí-li ještě navíc

$$\xi_q \stackrel{def}{=} \sum_{p=1}^{n} \lambda_p \widetilde{\psi}\left(\vec{\boldsymbol{x}}_p + qe_{2n+1}\right), \tag{3.14}$$

tak pro libovolný bod $\vec{\boldsymbol{x}} \in \langle 0,1\rangle^n$ *existuje alespoň* $n+1$ *hodnot* q, *pro které je* $\xi_q \in \bar{T}^q_{k,i_1,\ldots,i_n}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

3.3.3 Konstrukce přenosových funkcí a neuronové sítě

Vlastnosti intervalů $\bar{E}^q_{k,i}$ uvedené v lemmě 3.3.5 umožňují bezprostřední důkaz modifikované Kolmogorovovy věty ve tvaru vhodně korespondujícím s vrstevnatou architekturou neuronových sítí. Tato konstrukce je odvozena z původního Kolmogorovova důkazu, uvedeného například v [Kol57].

Definice 3.3.5 *Pro každé* $p \in \{1, \ldots, n\}$ *a* $k \geq 2n+1$ *nechť je* $i_p \in \bar{I}_k$. *Potom definujme množinu*

$$\bar{\Omega}^q_{k,i_1,\ldots,i_n} \stackrel{def}{=} \prod_{p=1}^{n} \bar{E}^q_{k,i_p},$$

($\bar{\Omega}^q_{k,i_1,\ldots,i_n}$ *je tedy kartézským součinem intervalů* $\bar{E}^q_{k,i_p}$*).*

Lemma 3.3.6 *Pro libovolnou spojitou funkci* $\widetilde{g} : \langle 0,1\rangle \to \Re$ *a pro libovolnou krychli* $\bar{\Omega}^q_{k,i_1,\ldots,i_n}$ *označme*

$$\Gamma^q_{k,i_1,\ldots,i_n} \stackrel{def}{=} \max\left\{\left|\widetilde{g}\,(\vec{\boldsymbol{x}}) - \alpha^q_{k,i_1,\ldots,i_n}\right| \,\middle|\, \vec{\boldsymbol{x}} \in \bar{\Omega}^q_{k,i_1,\ldots,i_n}\right\},$$

kde

$$\alpha^q_{k,i_1,\ldots,i_n} \stackrel{def}{=} \min\left\{\widetilde{g}\,(\vec{\boldsymbol{x}}) \,\middle|\, \vec{\boldsymbol{x}} \in \bar{\Omega}^q_{k,i_1,\ldots,i_n}\right\}.$$

Nechť $\Gamma_{\widetilde{g},k}$ *je maximum čísel* $\Gamma^q_{k,i_1,\ldots,i_n}$ *přes všechny intervaly* $\bar{\Omega}^q_{k,i_1,\ldots,i_n}$. *Potom pro libovolné číslo* $\epsilon > 0$ *lze volit přirozené číslo* k *tak, že* $\Gamma_{\widetilde{g},k} \leq \epsilon$.

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Definice 3.3.6 *Nechť je dáno přirozené* $n \geq 2$ *a funkce* $\widetilde{f} : \langle 0,1\rangle^n \to \Re$. *Pro každé* $q \in \{0,\ldots,2n\}$ *definujme funkce* $\widetilde{\Phi_{q,0}} : \Re \to \Re$ *jako nulové funkce a* $\widetilde{f_0} : \langle 0,1\rangle^n \to \Re$ *také jako nulovou funkci. Dále definujme*

$$M_0 \stackrel{def}{=} \sup_{\vec{\boldsymbol{x}} \in \langle 0,1\rangle^n} \left|\widetilde{f}\,(\vec{\boldsymbol{x}})\right|$$

a číslo k_0 *tak, aby* $\Gamma_{\widetilde{f},k_0} \leq \frac{1}{2n+2} M_0$.

Předpokládejme, že pro nějaké přirozené $r \geq 0$ *jsou již definovány funkce* $\widetilde{f_r}$ *a* $\widetilde{\Phi_{q,r}}$ *a čísla* M_r, k_r. *Potom definujme*

$$\widetilde{f_{r+1}}\,(\vec{\boldsymbol{x}}) \stackrel{def}{=} \sum_{q=0}^{2n} \widetilde{\Phi_{q,r}}\left(\sum_{p=1}^{n} \lambda_p \widetilde{\psi}\,(\vec{\boldsymbol{x}}_p + q e_{2n+1})\right)$$

a pro všechna $y \in \bar{T}^q_{k,i_1,\ldots,i_n}$

$$\widetilde{\Phi_{q,r+1}}\,(y) \stackrel{def}{=} \widetilde{\Phi_{q,r}}\,(y) + \frac{1}{n+1}\left|\widetilde{f}\,(\vec{\boldsymbol{z}}) - \widetilde{f_r}\,(\vec{\boldsymbol{z}})\right|, \tag{3.15}$$

kde $\vec{\boldsymbol{z}}$ *je libovolný bod ležící v krychli* $\bar{\Omega}^q_{k,i_1,\ldots,i_n}$. *Dále definujme*

$$M_{r+1} \stackrel{def}{=} \sup_{\vec{\boldsymbol{x}} \in \langle 0,1\rangle^n} \left|\widetilde{f}\,(\vec{\boldsymbol{x}}) - \widetilde{f_r}\,(\vec{\boldsymbol{x}})\right|$$

a číslo k_{r+1} *tak, aby platilo* $\Gamma_{\left(\widetilde{f}-\widetilde{f_r}\right),k_{r+1}} \leq \frac{1}{2n+2} M_r$.

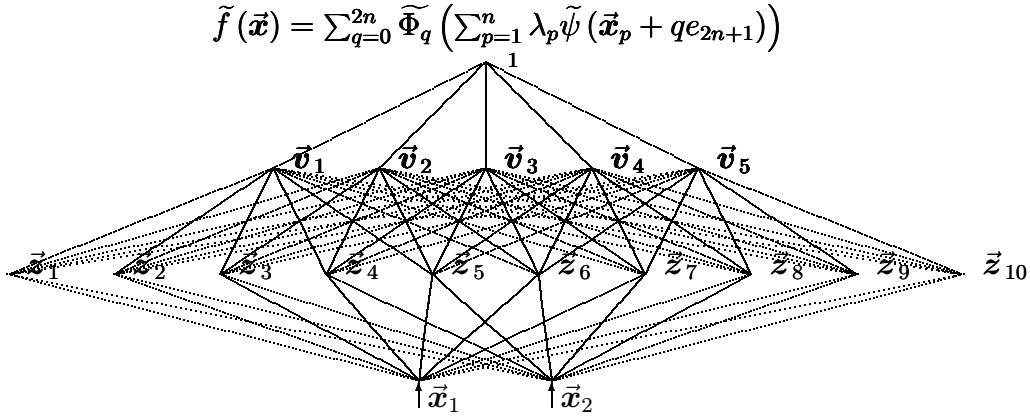
Lemma 3.3.7 *Platí* $\lim_{r\to\infty} M_r = 0$.

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Poznamenejme, $\frac{2n+1}{2n+2}$ se pro velká n blíží k jedné, na základě čehož se dá usuzovat na obecně pomalou rychlost konvergence posloupnosti $\{M_r\}_0^\infty$, což se projeví nutností volit hodnoty k_r velmi vysoké, chceme-li zaručit dostatečnou rychlost konvergence $\{M_r\}_0^\infty$. Na základě konvergence posloupnosti $\{M_r\}_0^\infty$ můžeme vyslovit tvrzení, které je obdobné původní Kolmogorově větě, ale formálně popisuje zobrazení, které lze realizovat prostřednictvím neuronové sítě sestávající ze vstupní vrstvy, dvou následujících vrstev a jednoho výstupního neuronu.



Obrázek 3.4: Schéma architektury sítě odvozené od Kolmogorovy věty pro $n = 2$. ($\vec{z}_{(p-1)n+q} = \widetilde{\psi}(\vec{x}_p + qe_{2n+1})$, $\vec{v}_q = \widetilde{\Phi}_q \left(\sum_{p=1}^{n} \lambda_p \widetilde{\psi}(\vec{x}_p + qe_{2n+1}) \right)$.)

Tvrzení 3.3.8 *Nechť $\lambda_1, \ldots, \lambda_n$ je kladná, celočíselně nezávislá posloupnost. Potom existuje spojitá, monotónně rostoucí funkce $\widetilde{\psi} : \bar{D}_1 \to \bar{D}_1$, taková, že pro každou reálnou spojitou funkci $\widetilde{f} : \langle 0,1 \rangle^n \to \Re$, $n \geq 2$, existují spojité funkce $\widetilde{\Phi}_q$, $q \in \{0, \ldots, 2n\}$, takové, že platí rovnice*

$$\widetilde{f}(\vec{x}) = \sum_{q=0}^{2n} \widetilde{\Phi}_q \left(\sum_{p=1}^{n} \lambda_p \widetilde{\psi}(\vec{x}_p + qe_{2n+1}) \right).$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Zřejmý důsledek předchozího tvrzení udává architekturu a základní vlastnosti neuronů sítě, která přesně počítá spojitou funkci definovanou na jednotkové krychli prostoru $\Re^n$.

Důsledek 3.3.9 *Každá spojitá funkce $\widetilde{f} : \langle 0,1 \rangle^n \to \Re$, $n \geq 2$, může být přesně spočítána neuronovou sítí s n neurony ve vstupní vrstvě, $(2n+1)n$ neurony ve druhé vrstvě, $2n+1$ neurony ve třetí vrstvě a s jedním výstupním neuronem (=čtvrtá vrstva). Navíc platí, že váhy mezi vstupní vrstvou a druhou vrstvou jsou buď 1 nebo 0, váhy mezi druhou a třetí vrstvou jsou buď 0 nebo λ_i, $i \in \{1, \ldots, n\}$, váhy mezi neurony třetí vrstvy a výstupním neuronem jsou rovny jedničce a žádná váha není závislá na $\widetilde{f}$. Dále neurony ve druhé vrstvě a výstupní neuron nezávisí na implementované funkci $\widetilde{f}$.*

3.4 Konvoluční přístup

Základní myšlenka konvoluční metody spočívá v konstrukci jádra, které je odvozené ze sigmoidálních funkcí a které může aproximovat Diracovu delta funkci (t.j. distribuci, jejíž nosič je pouze jednobodová množina). Samotná konvoluce a současně jádro jsou aproximovány kvadraturou, na základě které dostáváme požadovanou aproximaci dané spojité

funkce. Začneme tvrzením o základních vlastnostech uniformní konvergence konvolucí. (Připomeňme, že $\|\vec{\boldsymbol{x}}\|_\infty \stackrel{\text{def}}{=} \max_{i\in\{1,\dots,n\}} |\vec{\boldsymbol{x}}_i|$ a $\|\vec{\boldsymbol{x}}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^n |\vec{\boldsymbol{x}}_i|$).

Tvrzení 3.4.1 *Předpokládejme, že* $\widetilde{f}$ *je omezená, stejnoměrně spojitá funkce na* $\Re^n$ *a že pro* $\widetilde{g} \in \mathcal{L}_1^{\Re^n}$ *platí:*

$$\int_{\Re^n} \widetilde{g}\,(\vec{\boldsymbol{x}}) d\,(\vec{\boldsymbol{x}}) = 1.$$

Dále definujme funkce $\widetilde{g_m}\,(\vec{\boldsymbol{x}}) \stackrel{def}{=} m^n \widetilde{g}\,(m\vec{\boldsymbol{x}})$. *Potom platí:*

1. $\left(\widetilde{f} * \widetilde{g_m}\right)$ *konverguje stejnoměrně k funkci* $\widetilde{f}$ *pro* $m \to \infty$.
2. *pro libovolné číslo* $R > 0$ *platí*

$$\left\|\left(\widetilde{f} * \widetilde{g_m}\right) - \widetilde{f}\right\|_\infty \le \omega\left(\widetilde{f}, \frac{2R}{m}\right) \|\widetilde{g}\|_1 + 2\left\|\widetilde{f}\right\|_\infty \int_{\|\vec{\boldsymbol{s}}\|>R} |\widetilde{g}\,(\vec{\boldsymbol{s}})| d\vec{\boldsymbol{s}},$$

   kde norma $\|.\|_\infty$ *je brána vzhledem k celému* $\Re^n$.

3. *Nechť* $\widetilde{f}$ *Lipstichovská s konstantou* λ *a*

$$M \stackrel{def}{=} \int_{\Re^n} \|\vec{\boldsymbol{x}}\| |\widetilde{g}\,(\vec{\boldsymbol{x}})| d\vec{\boldsymbol{x}} < \infty.$$

   Potom platí odhad $\left\|\left(\widetilde{f} * \widetilde{g_m}\right) - \widetilde{f}\right\|_\infty \le \frac{M\lambda}{m}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Definice 3.4.1 *Pro libovolnou spojitou funkci* $\widetilde{\phi} \in C_{\Re^n}$ *definujme* KONVOLUČNÍ JÁDRO *jako funkci*

$$\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}}) \stackrel{def}{=} \frac{1}{\alpha_n} \int_{\|\vec{\boldsymbol{u}}\|=1} \widetilde{\phi}\,(\langle \vec{\boldsymbol{x}}\,|\vec{\boldsymbol{u}}\rangle) d\,(\vec{\boldsymbol{u}}), \quad \textit{kde} \quad \alpha_n \stackrel{def}{=} \int_{\|\vec{\boldsymbol{u}}\|=1} d\,(\vec{\boldsymbol{u}}), \tag{3.16}$$

(α_n *je povrch jednotkové koule v prostoru* $\Re^n$ *).*

Toto jádro použijeme k důkazu následující stěžejní věty, která je založena na obsahu tvrzení 3.4.1.

Tvrzení 3.4.2 *Nechť pro nějaké* $a > 0$ *je* $\bar{K} \stackrel{def}{=} \langle a, a\rangle^n$ *a* $\widetilde{\phi} \in C_\Re$ *je stejnoměrně spojitá. Předpokládejme, že* $\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}})$ *je definována podle 3.16. Potom jestliže* $\widetilde{G}_{\widetilde{\phi}}(\cdot) \in \mathcal{L}_1^{\Re^n}$ *a současně*

$$\int_{\Re^n} \widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}}) d\,(\vec{\boldsymbol{x}}) \neq 0,$$

tak množina funkcí ve tvaru $\widetilde{\phi}\,(\langle \vec{\boldsymbol{x}}\,|\vec{\boldsymbol{a}}\rangle - c)$, $\vec{\boldsymbol{a}} \in \Re^n$, $c \in \Re$, *je hustá v prostoru* $C_{\bar{K}}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

3.4.1 Volba konvolučního jádra

Abychom ukázali, že množina všech funkcí tvaru $\widetilde{\phi}\left(\langle \vec{\boldsymbol{x}} \,|\vec{\boldsymbol{a}}\rangle - c\right)$, $\vec{\boldsymbol{a}} \in \Re^n$, $c \in \Re$ je hustá v $C_{\bar{K}}$, musíme jednak dokázat, že konvoluční jádro $\widetilde{G}_{\widetilde{\phi}}(\cdot)$ definované v 3.16 patří do prostoru $\mathcal{L}_1^{\Re^n}$, a dále, že integrál tohoto jádra přes celý definiční obor je od nuly různý. Následující lemma nám pomůže stanovit takovou mocninu proměnné $r \stackrel{\text{def}}{=} \|\vec{\boldsymbol{x}}\|$, aby jádro $\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}})$, omezené pro velká r touto mocninou, bylo integrabilní a tedy v $\mathcal{L}_1^{\Re^n}$.

Lemma 3.4.3 *Nechť $p, R \in \Re$, $R > 0$. Potom platí, že*

$$\int_{\|\vec{\boldsymbol{x}}\|=1} \frac{1}{\|\vec{\boldsymbol{x}}\|^p} d\left(\vec{\boldsymbol{x}}\right) < \infty$$

právě když $p > n$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Abychom dokázali, že $\widetilde{G}_{\widetilde{\phi}}(\cdot) \in \mathcal{L}_1^{\Re^n}$, stačí ukázat, že $\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}}) = O\left(\frac{1}{\|\vec{\boldsymbol{x}}\|^n}\right)$ pro $\|\vec{\boldsymbol{x}}\| \to \infty$. V tomto důkazu nám pomůže vyjádření jádra $\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}})$ v následujícím tvaru:

Lemma 3.4.4 *Nechť $\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}})$ je definováno dle 3.16.Potom $\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}}) = \widetilde{g_0}\left(\widetilde{\phi}, r\right)$, kde $r \stackrel{def}{=} \|\vec{\boldsymbol{x}}\|$ a*

$$\begin{aligned} \widetilde{g_0}\left(\widetilde{\phi}, r\right) &\stackrel{def}{=} \frac{\alpha_{n-2}}{\alpha_{n-1}} \int_{-1}^{1} \widetilde{\phi}\left(rs\right)\left(1-s^2\right)^{\frac{n-3}{2}} d\left(s\right) \\ &= \frac{\alpha_{n-2}}{\alpha_{n-1}} \int_{-r}^{r} r^{2-n} \widetilde{\phi}\left(t\right)\left(r^2-t^2\right)^{\frac{n-3}{2}} d\left(t\right), \quad r \neq 0. \end{aligned} \tag{3.17}$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Ukážeme, že pro důkaz hustoty sigmoidálních funkcí v množině funkcí spojitých na kompaktní množině je vhodné zkoumat ne přímo vlastnosti dané sigmoidální funkce, ale jisté funkce od ní odvozené. Zkoumejme otázku, jakou podmínku musí splňovat funkce $\widetilde{\phi}$, aby jádro $\widetilde{G}_{\widetilde{\phi}}(\cdot)$ bylo v $\mathcal{L}_1^{\Re^n}$. Základním požadavkem je, aby $\widetilde{\phi}$ šlo dostatečně rychle k nule pro velké absolutní hodnoty argumentů. Zřejmě hodnota jádra $\widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}})$ je nulová, je-li $\widetilde{\phi}$ lichá funkce. Zvolme tedy $\widetilde{\phi}$ tak, aby byla sudá (není to ale nutné). Pro sigmoidální funkci $\widetilde{\sigma}$ definujme funkci

$$\widetilde{\psi}\left(t\right) \stackrel{\text{def}}{=} \widetilde{\sigma}\left(1+t\right) + \widetilde{\sigma}\left(1-t\right).$$

Zřejmě $\widetilde{\psi}$ je sudá funkce, která jde v $\pm\infty$ k nule. Předpokládejme, že $\widetilde{\sigma}$ je spojitá funkce a že

$$\left|\widetilde{\psi}\left(t\right)\right| < \frac{\eta}{|t|^p}, \quad p > n-2,$$

pro nějakou kladnou hodnotu η. Rozvineme-li výraz 3.17 do mocninné řady, dostaneme

$$\widetilde{g_0}\left(\widetilde{\phi}, r\right) = \frac{1}{r^{n-2}} \sum_{j=0}^{\lambda} \widetilde{\beta_j}\left(r\right) r^{2\lambda-2j}, \tag{3.18}$$

kde

$$\widetilde{\beta_j}\left(r\right) \stackrel{\text{def}}{=} \frac{\alpha_{n-2}}{\alpha_{n-1}} \binom{\lambda}{j} \int_{-r}^{+r} \widetilde{\phi}\left(t\right) t^{2j} d\left(t\right).$$

Tvaru tohoto rozkladu využijeme pro důkaz následující lemmy.

Lemma 3.4.5 *Nechť n je liché a pro $\widetilde{\psi}\left(t\right) \stackrel{def}{=} \widetilde{\phi}\left(1+t\right) + \widetilde{\phi}\left(1-t\right)$ a pro nějaké $\eta > 0$ platí $\left|\widetilde{\psi}\left(t\right)\right| \leq \frac{\eta}{|t|^p}$, $p > n-2$. Potom $\widetilde{G}_{\widetilde{\phi}}(\cdot) \in \mathcal{L}_1^{\Re^n}$ tehdy a jen tehdy, je-li*

$$\int_0^\infty \widetilde{\psi}\left(t\right) t^{2j} d\left(t\right) = 0, \quad j \in \{0, \ldots, \frac{n-3}{2}\}. \tag{3.19}$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Lemma 3.4.6 *Předpokládejme, že n je liché a že funkce $\widetilde{\psi}$ splňuje předpoklady a tvrzení lemmy 3.4.5. Dále předpokládejme, že $\widetilde{\psi}$ je sudá funkce. Potom platí*

$$\int_{\Re^n} \widetilde{G}_{\widetilde{\phi}}(\vec{\boldsymbol{x}}) d\left(\vec{\boldsymbol{x}}\right) = -2\alpha_{n-2}\tau_n \int_0^\infty \widetilde{\psi}\left(t\right) t^{n-1} d\left(t\right), \tag{3.20}$$

kde je

$$\tau_n \stackrel{def}{=} \int_0^1 r \left(1-r^2\right)^{\frac{n-3}{2}} d\left(r\right) > 0.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Spojíme-li všechny předchozí lemmy dohromady, můžeme dokázat následující tvrzení, kterým uzavřeme kapitolu, týkající se aproximačních vlastností neuronových sítí.

Tvrzení 3.4.7 *Nechť $\widetilde{\sigma} \in C_{\Re}$ je stejnoměrně spojitá na $\Re$ a $\widetilde{\psi}\left(t\right) \stackrel{def}{=} \widetilde{\sigma}\left(1+t\right) + \widetilde{\sigma}\left(1-t\right)$. Předpokládejme, že n je liché a $\bar{K} = \langle -a, a\rangle^n$, $a > 0$ a pro nějaké $\eta > 0$ platí $\left|\widetilde{\psi}\left(t\right)\right| \leq \frac{\eta}{|t|^p}$, $p > n-2$. Dále nechť*

$$\int_0^\infty \widetilde{\psi}\left(t\right) t^{n-1} d\left(t\right) \neq 0.$$

Potom množina funkcí tvaru $\widetilde{\sigma}\left(\langle \vec{\boldsymbol{x}} \,|\vec{\boldsymbol{w}}\rangle + c\right)$, kde $\vec{\boldsymbol{w}} \in \Re^n$ a $c \in \Re$, je hustá v prostoru $C_{\bar{K}}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Kapitola 4

Vapnik-Chervonenkova dimenze

Einsteinův axiom

Pokud vycházejí matematické poučky ze skutečnosti,
nejsou spolehlivé.
Pokud jsou spolehlivé,
nevycházejí ze skutečnosti.

Nedílnou součástí při výkladu teorie umělých neuronových sítí je i problematika spojená se schopnostmi neuronových sítí efektivně přiřazovat svým vnitřním parametrům (váhy jednotlivých spojů, prahy uzlů atd.) takové hodnoty, aby odezva neuronové sítě na dané vstupní hodnoty byla rovna hodnotám požadovaným, nebo je alespoň uspokojivě aproximovala. Procesu nastavování těchto parametrů na základě extrakce vlastností vstupních hodnot se říká učení. Teorie učení byla rozpracována obecně, nejenom pro neuronové sítě. K rigorózním výkladu té části teorie učení, která bezprostředně souvisí s neuronovými sítěmi je zapotřebí nejdříve zavést a detailně analyzovat jistou míru složitosti množin, která se označuje jako VC-dimenze. Této analýze věnujeme tuto část.

4.1 Pojem konceptu a třídy konceptů

Základním předmětem, ke kterému s váže pojem VC-dimenze, je jakákoli podmnožina potenční množiny libovolné, pevně dané množiny $\bar{X}$. V literatuře relevantní k našemu dalšímu výkladu jsou pro takovéto systémy podmnožin používány termíny koncept, hypotéza, třída konceptů a třída hypotéz. Proto této terminologie budeme v následujících definicích a výkladu používat také.

Definice 4.1.1 *Nechť $\bar{X}$ je libovolná množina. Potom $\bar{c} \subset \bar{X}$ nazveme* KONCEPTEM *(nad množinou $\bar{X}$). Neprázdnou množinu $\mathsf{C} \subset 2^{\bar{X}}$ nazveme* TŘÍDOU KONCEPTŮ *(nad množinou X). Prvky této množiny $\bar{c} \in \mathsf{C}$, budeme nazývat koncepty třídy C .*

Jedním ze způsobů popisu konceptu je zadání konceptu jako vzor dané množiny při nějakém zobrazení. Formálně tedy definujeme pro booleovskou funkci $\widetilde{f}$ koncept $\mathsf{C}_{\widetilde{f}}$.

Definice 4.1.2 *Nechť $\widetilde{f}$ je zobrazení z množiny $\bar{X}$ do dvouprvkové množiny $\{-1, +1\}$. Potom definujme koncept $\bar{c}_{\widetilde{f}}$ jako množinu*

$$\bar{c}_{\widetilde{f}} \stackrel{def}{=} \left\{ x \in \bar{X} \,\middle|\, \widetilde{f}\,(x) = 1 \right\}.$$

Předchozí definici lze zřejmě rozšířit z definice konceptu na definici celé třídy konceptů, odpovídající dané množině zobrazení.

Definice 4.1.3 *Předpokládejme, že* $\bar{F}$ *je libovolná množina funkcí definovaných na množině* $\bar{X}$ *s oborem hodnot* $\{-1,+1\}$ *. Pak říkáme, že* $\bar{F}$ REPREZENTUJE TŘÍDU KONCEPTŮ

$$\mathsf{C}_{\bar{F}} \stackrel{def}{=} \left\{ \bar{A} \subset \bar{X} \,\middle|\, (\exists \widetilde{f} \in \bar{F})(\bar{A} = \bar{c}_{\widetilde{f}}) \right\}.$$

V souladu s touto definicí lze definovat třídu konceptů, které jsou generovány neuronovou sítí typu perceptron.

Definice 4.1.4 *Nechť* n *je přirozené číslo a*

$$\bar{F} \stackrel{def}{=} \left\{ \widetilde{f} : \Re^n \to \{-1,+1\} \,\middle|\, \widetilde{f} = \widetilde{sgn}\,(\langle \vec{\boldsymbol{x}} \,|\vec{\boldsymbol{w}}\rangle - t), \quad t \in \Re, \;\; \vec{\boldsymbol{w}}, \vec{\boldsymbol{x}} \in \Re^n \right\}.$$

Potom definujme třídu konceptů

$$\mathbf{HALFSPACE}_n \stackrel{def}{=} \mathsf{C}_{\bar{F}}.$$

Zřejmě třída konceptů $\mathbf{HALFSPACE}_n$ obsahuje veškeré poloprostory v Eukleidovském prostoru dimenze n.

4.2 VC-dimenze třídy konceptů

Nyní uvedeme stěžejní pojem vyjadřující schopnost pevně daného systému množin vymezit v jiné množině libovolnou podmnožinu. Zřejmě vlastnosti každé neuronové sítě velmi silně závisí na vlastnostech třídy konceptů, která je reprezentována všemi možnými zobrazeními, která je daná architektura neuronové sítě schopna realizovat s ohledem na možné hodnoty vnitřních parametrů sítě. Zřejmě čím složitější geometrické vlastnosti konceptů, které lze pomocí neuronové sítě generovat, tím potenciálně vyšší možnost existence parametrů (vah atd.), které jsou vhodné pro řešení požadované konkrétní úlohy (buď aproximace funkce či separace množin). Intuitivně se dá očekávat, že taková třída konceptů, jejíž pomocí lze v dané množině vymezit co největší počet všech možných podmnožin dané velikosti je vhodnější, pro úlohy založené na separaci množin. Mírou této schopnosti je z jistého úhlu pohledu právě Vapnik-Chervonenkova dimenze.

Definice 4.2.1 *Nechť* $\bar{X}$ *je daná množina,* $\bar{Y} \subset \bar{X}$, $\mathsf{C} \subset 2^{\bar{X}}$. *Potom říkáme, že* C ROZDĚLUJE $\bar{Y}$ *, právě když platí:*

$$(\forall \bar{z} \subset \bar{Y})(\exists \bar{c} \in \mathsf{C})(\bar{c} \cap \bar{Y} = \bar{z}).$$

Dále definujme pojem VAPNIK-CHERVONENKOVA DIMENZE *systému množin* C *jako*

$$\mathtt{VC}_{dim}(\mathsf{C}) \stackrel{def}{=} \max \left\{ \left|\bar{Y}\right| \,\middle|\, \mathsf{C} \textit{ rozděluje } \bar{Y} \right\},$$

nebo jako $+\infty$ *v případě, že toto maximum neexistuje.*

Poznamenejme, že Vapnik-Chervonenkova dimenze je rovna mohutnosti maximální množiny $\bar{A} \in \mathsf{C}$, pro kterou platí

$$\left\{ \bar{B} \,\middle|\, (\exists \bar{Z} \in \mathsf{C})(\bar{B} = \bar{A} \cap \bar{Z}) \right\} = 2^{\bar{A}}.$$

Uveďme několik příkladů na výpočet VC-dimenze ([AB92]).

Příklad 4.2.1

1. *Nechť* $\bar{X} = \Re$ *a* C *je tvořeno všemi intervaly tvaru* $(a, +\infty)$ *, kde* $a \in \Re$*. Potom zřejmě pro libovolnou* $\{b, c \in \Re \,|b < c\}$ *neexistuje žádný interval* $(a, +\infty)$ *, který obsahuje bod* b *a neobsahuje bod* c*. Zřejmě tedy* $\mathtt{VC}_{dim}(\mathsf{C}) = 1$*.*

2. *Nechť* $\bar{X} = \Re$*,* s *je přirozené číslo,* $s > 1$ *a* C *je množina obsahující sjednocení právě* s *intervalů. Nechť* $\bar{S} = \{x_1, \ldots, x_{2s} \,|x_i < x_{i+1}\}$*. Zřejmě* $\bar{S}$ *je rozdělena třídou konceptů* C *. Je-li ale* $\bar{S} = \{x_1, \ldots, x_{2s+1} \,|x_i < x_{i+1}\}$*, pak každý koncept obsahující body* $x_1, x_3, \ldots, x_{2s+1}$ *nutně obsahuje i body* $x_2, x_4, \ldots, x_{2s}$*. Tedy takováto* $\bar{S}$ *není rozdělena třídou konceptů* C *, tedy* $\mathtt{VC}_{dim}(\mathsf{C}) = 2s$*.*

3. *Nechť* $\bar{X} = \Re^n$ *a* C *je tvořena všemi intervaly v prostoru* $\Re^n$ *. Evidentně množina* $2n$ *bodů ležících ve středu stěn jednotkové krychle je třídou* C *rozdělena. Vezmeme-li ale libovolnou množinu* $\bar{A}$ $2n + 1$ *bodů, lze k ní přiřadit minimální interval obsahující celou* $\bar{A}$ *. Pak ale musí existovat prvek* $\vec{\boldsymbol{x}} \in \bar{A}$*, který je buď uvnitř minimálního intervalu, nebo na nějaké jeho stěně, a zřejmě množinu* $\bar{A} - \{\vec{\boldsymbol{x}}\}$ *nelze intervalem oddělit od množiny* $\{\vec{\boldsymbol{x}}\}$*. Proto* $\mathtt{VC}_{dim}(\mathsf{C}) = 2n$*.*

Příklad 4.2.2 *Nechť* C *je konečná třída konceptů. Jelikož k rozdělení množiny mohutnosti* d *je třeba minimálně* 2^d *různých konceptů, množina o mohutnosti větší než* $\log_2(|\mathsf{C}|)$ *nemůže být třídou* C *rozdělena. Proto* $\mathtt{VC}_{dim}(\mathsf{C}) \leq \log_2(|\mathsf{C}|)$*.*

Příklad 4.2.3 *Pro úplnost ukážeme příklad třídy konceptů s nekonečnou VC-dimenzí. Položme*

$$\mathsf{C} \stackrel{def}{=} \left\{ \bar{A}_\alpha \,\middle|\, (\exists \alpha \in \Re^n) \left(\bar{A}_\alpha = \left\{ x \in \Re \,\middle|\, \widetilde{\sin}(\alpha x) \geq 0 \right\} \right) \right\}.$$

Potom pro libovolné přirozené l *vezměme posloupnost čísel* $\bar{Z}_l \stackrel{def}{=} \{z_i\}_1^l$*,* $z_i = \frac{1}{10^i}$*. Dále každou podmnožinu množiny* $\bar{Z}_l$*, lze charakterizovat posoupností* $\delta_1, \ldots, \delta_l$ *,* $\delta_i \in \{0, 1\}$*. Zvolíme-li*

$$\alpha \stackrel{def}{=} \pi \left(\sum_{i=1}^{l} (1 - \delta_i)\, 10^i + 1 \right),$$

pak platí rovnost

$$\alpha \frac{1}{10^j} = \pi \left(\sum_{i=1}^{l} (1 - \delta_i)\, 10^{i-j} + \frac{1}{10^j} \right).$$

Odtud

$$\alpha \frac{1}{10^j} = \pi \left(\sum_{i=1}^{j-1} \frac{1 - \delta_i}{10^{j-i}} + (1 - \delta_j) + \sum_{i=j+1}^{l} (1 - \delta_i)\, 10^{i-j} + \frac{1}{10^j} \right).$$

Proto je $\widetilde{\sin}(\alpha z_j) < 0$ *pro* $\delta_j = 0$ *a* $\widetilde{\sin}(\alpha z_j) > 0$ *pro* $\delta_j = 1$*. Zřejmě tedy množina* $\bar{A}_\alpha$ *separuje každou podmnožinu* $\bar{Z}_l$ *definovanou prostřednictvím* $\delta_1, \ldots, \delta_l$ *. Tedy pro libovolné* l *existuje množina mohutnosti* l*, která je rozdělena třídou konceptů* C *, tedy* $\mathtt{VC}_{dim}(\mathsf{C}) = +\infty$*.*

Velmi užitečným nástrojem pro horní odhad VC-dimenze systému poloprostorů Eukleidovského prostoru $\Re^n$ je následující Radonovo lemma (viz. [Lej85]). Předtím, než toto lemma vyslovíme, připomeňme, že lineární kombinaci $\sum_{i=1}^n \alpha_i \vec{\boldsymbol{x}}_i$ libovolného konečného počtu vektorů z daného vektorového prostoru nazveme afinní kombinací vektorů, právě když $\sum_{i=1}^n \alpha_i = 1$. Dále připomeňme, že v n dimenzionálním prostoru je každá $n+2$ členná posloupnost vektorů afinně závislá, což znamená, že platnost rovnic

$$\sum_{i=1}^k \alpha_i \vec{\boldsymbol{x}}_i = \vec{\mathbf{0}} \quad \text{a} \quad \sum_{i=1}^k \alpha_i = 0$$

implikuje, že všechna α_i jsou rovna 0, $i \in \{1, \ldots, k\}$.

Lemma 4.2.1 (Radonova) *Nechť* $\bar{S} \stackrel{def}{=} \{\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_k\} \subset \Re^n$, $k \geq n+2$, $\vec{\boldsymbol{x}}_i$ *jsou po dvou od sebe různé. Potom existují množiny* $\bar{S}_1$ *a* $\bar{S}_2$ *tak, že* $\bar{S}_1 \cup \bar{S}_2 = \bar{S}$, $\bar{S}_1 \cap \bar{S}_2 = \emptyset$ *a*

$$[\bar{S}_1]_\kappa \cap [\bar{S}_2]_\kappa \neq \emptyset,$$

($[\bar{S}]_\kappa$ *značí konvexní obal množiny* $\bar{S}$ *).*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Aplikací Radonova lemmatu dokážeme následující tvrzení pro VC-dimenzi poloprostorů.

Tvrzení 4.2.2 $\mathtt{VC}_{dim}(\mathbf{HALFSPACE}_n) = n+1$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Vezměme v úvahu pouze takové třídy konceptů, která jsou definovány nad konečnými množinami $\bar{X}$. Potom následující Sauerovo lemma dává horní odhad mohutnosti každé třídy konceptů se zadanou VC-dimenzí. Zdůrazněme, jak plyne z důkazu lemmatu ([Sau72]), že tento odhad je čistě kombinatorickou vlastností konečných množin.

Lemma 4.2.3 (Sauer) *Nechť* $\bar{X}$ *je konečná množina,* $\mathsf{C} \subset 2^{\bar{X}}$. *Potom*

$$|\mathsf{C}| \leq \sum_{i=0}^{\mathtt{VC}_{dim}(\mathsf{C})} \binom{\left|\bar{X}\right|}{i}.$$

Navíc existuje $\mathsf{C} \subset 2^{\bar{X}}$ *tak, že platí rovnost.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Pojem zavedený v následující definici má velmi úzký vztah k VC-dimenzi a také k tvrzení Sauerovy lemmy.

Definice 4.2.2 *Nechť* C *je neprázdná třída konceptů nad* $\bar{X}$ *a* $\bar{S} \subset \bar{X}$*. Potom definujme*

$$\Pi_{\mathsf{C}}\left(\bar{S}\right) \stackrel{def}{=} \left\{\bar{S} \cap \bar{c} \,|\bar{c} \in \mathsf{C}\right\}.$$

Dále pro pevné $m \geq 0$ *definujme*

$$\Pi_{\mathsf{C}}(m) \stackrel{def}{=} \max\left\{\left|\Pi_{\mathsf{C}}\left(\bar{S}\right)\right| \left|\bar{S} \subset \bar{X}\ ,\ \left|\bar{S}\right| = m\right\},$$

(maximum se bere přes všechny množiny $\bar{S} \subset \bar{X}$ *mohutnosti* m*).*

Zřejmě $\Pi_{\mathsf{C}}\left(\bar{S}\right)$ je systém všech podmnožin množiny $\bar{S}$, které lze oddělit od jejich doplňku v $\bar{S}$ pomocí konceptu ležícího ve třídě konceptů C. Číslo $\Pi_{\mathsf{C}}(m)$ vyjadřuje mohutnost maximálního takovéhoto systému $\Pi_{\mathsf{C}}\left(\bar{S}\right)$, kde množina $\bar{S}$ je konečná m prvková podmnožina $\bar{X}$. Mezi těmito čísly a VC-dimenzí platí evidentní vztah

$$\mathtt{VC}_{dim}(\mathsf{C}) = \max\{m \,|\Pi_{\mathsf{C}}(m) = 2^m\}.$$

Definice 4.2.3 *Pro všechna* $d \geq 0$ *a* $m \geq 0$ *položme* $\Phi_{d,m} \stackrel{def}{=} \sum_{i=0}^{d} \binom{m}{i}$*, je-li* $m \geq d$ *a* $\Phi_{d,m} \stackrel{def}{=} 2^m$ *je-li* $m < d$*.*

Tvrzení obsažená v následující lemmě budou nezbytná k důkazu horních odhadů délky vzorků, nutných pro učení algoritmů s danou přesností (ve smyslu nalezení hypotézy mající "malou" symetrickou diferenci s daným konceptem při požadované pravděpodobnosti nalezení takovéto hypotézy), které je analyzované v následujících oddílech.

Lemma 4.2.4 *Je-li* C *libovolná třída konceptů nad konečnou množinou* $\bar{X}$*, tak platí*

1. *Je-li* $\mathtt{VC}_{dim}(\mathsf{C}) = d$*, tak* $\Pi_{\mathsf{C}}(m) \leq \Phi_{d,m}$ *pro všechna* $m \geq 0$*.*
2. $\Phi_{d,m} \leq m^d + 1$ *pro* $d \geq 0$ *a* $\Phi_{d,m} \leq m^d$ *pro* $d \geq 0$ *a* $m \geq 2$*.*
3. $\Phi_{d,m} \leq 2\frac{m^d}{d!}m + 1 \leq \left(\frac{em}{d}\right)^d$ *pro* $m \geq d \geq 1$*.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Důsledek 4.2.5 *Nechť* $\bar{X}$ *je konečná množina,* $\mathsf{C} \subset 2^{\bar{X}}$ *a* $\mathtt{VC}_{dim}(\mathsf{C}) > 0$*. Potom*

$$\mathtt{VC}_{dim}(\mathsf{C}) > \frac{\log_e\left(\left|\bar{C}\right|\right)}{1 + \log_e\left(\left|\bar{X}\right|\right)}.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Cvičení 4.2.0 Vapnik-Chervonenkova dimenze

1. *Dokažte druhou část lemmy 4.2.4.*

2. *Nechť* $\vec{\boldsymbol{x}} \in \Re^n$ *a* $\mathsf{G}_{n,\vec{\boldsymbol{x}}}$ *je systém všech poloprostorů v* $\Re^n$ *, které neobsahují bod* $\vec{\boldsymbol{x}}$ *,*

$$\mathsf{G}_{n,\vec{\boldsymbol{x}}} \stackrel{def}{=} \{\bar{c} \subset \Re^n \,|(\exists \vec{\boldsymbol{w}} \in \Re^n,\ \alpha \in \Re)\,(\bar{c} = \{\vec{\boldsymbol{z}} \in \Re^n \,|\langle \vec{\boldsymbol{w}} \,|\vec{\boldsymbol{z}}\rangle \leq \alpha \ a\ \vec{\boldsymbol{x}} \notin \bar{c}\})\} \tag{4.1}$$

Dokažte, že $\mathtt{VC}_{dim}\left(\mathsf{G}_{n,\vec{\boldsymbol{x}}}\right) = n$.

3. *Ukažte, že* $\Pi_{\mathsf{G}_{n,\vec{\boldsymbol{x}}}}(m) = \Phi_{n,m}$.

4. *Nechť* A *je třída konceptů definovaná nad* $\bar{X}$ *a* $\mathtt{VC}_{dim}(\mathsf{A})$ *je konečná. Nechť* B *je třída konceptů obsahující doplňky konceptů z* A *do množiny* $\bar{X}$ *. Dokažte, že potom platí* $\mathtt{VC}_{dim}(\mathsf{A}) = \mathtt{VC}_{dim}(\mathsf{B})$.

5. *Nechť* C *je systém všech uzavřených intervalů na* $\Re$ *. Ukažte, že*

$$\Pi_{\mathsf{C}}(m) = 1 + m + \frac{1}{2}m(m-1).$$

6. *Řekneme, že třída konceptů* C *je* LINEÁRNĚ USPOŘÁDANÁ *, právě když obsahuje alespoň dva koncepty a pro libovolné* $\bar{a}, \bar{b} \in \mathsf{C}$ *je buď* $\bar{a} \subset \bar{b}$ *nebo* $\bar{b} \subset \bar{a}$. *Potom platí, že* $\mathtt{VC}_{dim}(\mathsf{C}) = 1$. *Dokažte.*

7. *Ukažte, že systém všech přímek v* $\Re^n$ *má VC-dimenzi rovnu* 1.

4.3 Vapnik-Chervonenkova dimenze neuronových sítí

Pro libovolný systém funkcí můžeme definovat VC-dimenzi těchto funkcí jako VC-dimenzi všech množin, které jsou vzorem pro intervaly $(-\infty, \alpha\rangle$ při zobrazení nějakou funkcí z uvažovaného systému funkcí. Obdobně lze definovat VC-dimenzi obecných neuronových sítí či booleovských obvodů, budeme-li na ně pohlížet jako na funkce zobrazující hodnoty na vstupních uzlech do množiny reálných čísel. Pro účely další analýzy VC-dimenze neuronových sítí zadefinujme obecnější model neuronových sítí, než je booleovský obvod.

Definice 4.3.1 DOPŘEDNÝ OBVOD *je acyklický orientovaný graf s hranovým a vrcholovým ohodnocením a s pouze jediným vrcholem, ze kterého nevede hrana. Vrchol tohoto grafu nazýváme* VSTUPNÍ VRCHOL *, nevede-li do něj žádná hrana,* VÝSTUPNÍ VRCHOL *, nevede-li z něj žádná hrana, a* VNITŘNÍ VRCHOL *, není-li ani vstupní ani výstupní. Hodnoty hranového ohodnocení jsou obecně reálná čísla, ohodnocení hrany z vrcholu* i *do vrcholu* j *nazveme* VÁHOU HRANY $i \to j$ *a budeme značit* $w_{i\to j}$. *Pro daný vrchol* j *nechť* n_j *je počet hran vedoucích do vrcholu* j *a* $\vec{\boldsymbol{w}}_j \in \Re^{n_j}$ *je vektor, jehož složky jsou rovny ohodnocení těchto hran. Ohodnocení vrcholů jsou reálná čísla, která pro vrchol* j *budeme značit jako* v_j. *Hodnoty* v_j *u vstupních vrcholů nazýváme vstupními hodnotami, ohodnocení* v_j *u výstupního vrcholu výstupní hodnotou. Nechť* $\vec{\boldsymbol{s}}_j \in \Re^{n_j}$ *je vektor tvořený hodnotami ohodnocení vrcholů, z nichž vede hrana do vrcholu* j. *Dále, ohodnocení vnitřního a výstupního vrcholu* j *splňuje rovnost*

$$v_j \stackrel{def}{=} \widetilde{f_j}\,(\vec{\boldsymbol{w}}_j, \vec{\boldsymbol{s}}_j),$$

kde $\widetilde{f_j}$ *jsou libovolné pevně dané funkce,* $\widetilde{f_j} : \Re^{n_j} \times \Re^{n_j} \to \Re$ *.*

V následujícím textu odvodíme základní horní odhad VC-dimenze dopředného obvodu využívající vlastností funkcí $\widetilde{f_j}$ v tomto obvodu. K tomu bude zapotřebí pracovat s částmi dopředného obvodu, které budeme definovat prostřednictvím tzv. vlastního očíslování vrcholů obvodu.

Definice 4.3.2 *Předpokládejme, že G je acyklický orientovaný graf a že uzly grafu G, které mají alespoň jednu vstupní hranu, jsou očíslovány přirozenými čísly $i \in \{1, \ldots, z\}$. Potom toto očíslování nazveme* VLASTNÍ OČÍSLOVÁNÍ *, právě když pro všechny očíslované uzly platí, že hrany z daného uzlu vedou pouze do uzlů, jejichž očíslování je vyšší.*

Vlastní očíslování acyklického orientovaného grafu dostaneme například takovým způsobem, že po odstranění všech uzlů nemajících vstupní hranu, v takto vzniklém grafu nejdříve očíslujeme postupně všechny uzly nemající vstupní hranu (to lze, neboť graf je acyklický), tyto uzly z grafu odstraníme, a tento postup zopakujeme na takto vzniklém grafu, přičemž číslujeme postupně dalšími čísly.

Definice 4.3.3 *Nechť D je dopředný obvod, který má $(z-1)$ vnitřních vrcholů. Potom pro každé $l \in \{1, \ldots, z\}$ definujme třídu konceptů*

$$\mathsf{C}_l \stackrel{def}{=} \left\{ \bar{c} \subset \Re^{n_l} \,\middle|\, (\exists \vec{\boldsymbol{w}} \in \Re^{n_l}) \left(\bar{c} = \left\{ \vec{\boldsymbol{s}} \in \Re^{n_l} \,\middle|\, \widetilde{f}_l\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{s}}\right) \leq 0 \right\} \right) \right\}.$$

Každou třídu konceptů C_l nazveme LOKÁLNÍ TŘÍDOU KONCEPTŮ *.*

Nechť dále $\bar{\Delta}$ je množina všech možných hranových ohodnocení obvodu D a hodnotu výstupního vrcholu z označme v závislosti na $\vec{\boldsymbol{h}} \in \bar{\Delta}$ a vstupních hodnotách $\vec{\boldsymbol{x}} \in \Re^n$ obvodu D jako $v_{z,\vec{\boldsymbol{h}},\vec{\boldsymbol{x}}}$. Potom definujme TŘÍDU KONCEPTŮ DOPŘEDNÉHO OBVODU *jako systém množin*

$$\mathsf{C}_D \stackrel{def}{=} \left\{ \bar{c} \subset \Re^n \,\middle|\, \left(\exists \vec{\boldsymbol{h}} \in \bar{\Delta}\right) \left(\bar{c} = \left\{ \vec{\boldsymbol{x}} \in \Re^n \,\middle|\, v_{z,\vec{\boldsymbol{h}},\vec{\boldsymbol{x}}} \leq 0 \right\} \right) \right\}.$$

VC-DIMENZÍ DOPŘEDNÉHO OBVODU *D budeme rozumět číslo $\mathtt{VC}_{dim}\left(\mathsf{C}_D\right)$.*

Jinými slovy, pro dané l je C_l je systém všech podmnožin $\bar{c} \in \Re^{n_l}$ takových, že existuje váhový vektor v $\Re^{n_l}$ tak, že množina $\bar{c}$ je vzorem intervalu $(-\infty, 0\rangle$ při zobrazení $\widetilde{f}_l\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{s}}\right)$. Obdobným způsobem je definována třída konceptů C_D.

Definice 4.3.4 *Nechť $\bar{V} \subset \Re^n$, $\bar{V} \stackrel{def}{=} \{\vec{\boldsymbol{x}}_i, \ldots, \vec{\boldsymbol{x}}_m\}$ a D je dopředný obvod dimenze n s $n+z$ vrcholy a s vlastním očíslováním vrcholů. Potom řekneme, že dvě hranová ohodnocení ω_1 a ω_2 jsou l-*ROZLIŠITELNÁ POMOCÍ *$\bar{V}$, $l \in \{1, \ldots, z\}$, právě když existuje $\vec{\boldsymbol{x}}_i \in \bar{V}$ a existuje $j \in \{1, \ldots, l\}$ taková, že hodnota uzlu j se liší, je-li hranové ohodnocení obvodu rovno hodnotám ω_1 nebo hodnotám ω_2. Dále označme symbolem $\mathsf{S}_{l,\bar{V}}$ maximální počet všech hranových ohodnocení, která jsou vzájemně po dvou l-rozlišitelná pomocí $\bar{V}$.*

Tvrzení 4.3.1 *Nechť m je libovolné kladné celé číslo, $\bar{V} \subset \Re^n$, $\bar{V} \stackrel{def}{=} \{\vec{\boldsymbol{x}}_i, \ldots, \vec{\boldsymbol{x}}_m\}$. Potom pro všechna $l \in \{1, \ldots, z\}$ platí odhad*

$$\mathsf{S}_{l,\bar{V}} \leq \Pi_{\mathsf{C}_1}(m) \cdot \Pi_{\mathsf{C}_2}(m) \cdot \cdots \cdot \Pi_{\mathsf{C}_l}(m).$$

Navíc platí

$$\Pi_{\mathsf{C}_D}(m) \leq \mathsf{S}_{z,\bar{V}} \leq \Pi_{\mathsf{C}_1}(m) \cdot \Pi_{\mathsf{C}_2}(m) \cdot \cdots \cdot \Pi_{\mathsf{C}_z}(m).$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Až do tohoto místa výkladu jsme nikde nevyužili vlastností přechodových funkcí $\widetilde{f}_i$ dopředného obvodu, podstatná byla pouze skutečnost, že graf dopředného obvodu je orientovaný a acyklický. Tvrzení 4.3.1 tedy zřejmě platí pro dosti širokou třídu neuronových sítí a dává nám návod, jak postupovat při odvozování horního odhadu VC-dimenze takovýchto obecnějších neuronových sítí, neboť umožňuje rozložit nalezení takovéhoto odhadu na odhady VC-dimenze sítí obsahujících pouze jednotlivé uzly sítě původní. Budeme ilustrovat tento postup na jednoduchém příkladě tzv. lineárního sigmoidálního obvodu, který v sobě zahrnuje i booleovské obvody.

Definice 4.3.5 *Dopředný obvod L nazveme* LINEÁRNÍM SIGMOIDÁLNÍM OBVODEM *jestliže pro každý nevstupní uzel obvodu L jemu příslušná funkce $\widetilde{f}_j$ je tvaru*

$$\widetilde{f}_j(\vec{\boldsymbol{s}}_j) \stackrel{def}{=} \widetilde{\sigma}_j(\langle \vec{\boldsymbol{s}}_j | \vec{\boldsymbol{w}}_j \rangle - t_j), \tag{4.2}$$

kde $\widetilde{\sigma}_j$ jsou libovolné, pevně dané sigmoidální funkce, $t_j \in \Re$, a $\vec{\boldsymbol{s}}_j$, $\vec{\boldsymbol{w}}_j$ mají stejný význam jako v definici 4.3.1.

Určíme nyní horní odhad VC-dimenze lineárního sigmoidálního obvodu. Jelikož každý takovýto obvod je konkrétní realizací dopředného obvodu zmíněného v předchozím, všechny potřebné definice a tvrzení ohledně dopředného obvodu jsou platné i pro tyto obvody. Přistupme proto přímo k tvrzení, jež se týká lineárních sigmoidálních obvodů.

Tvrzení 4.3.2 *Nechť L je lineární sigmoidální obvod, w je počet jeho hran, z je počet jeho vrcholů, které nejsou vstupní, $q \stackrel{def}{=} w + z$. Potom pro každé $m \geq \max\{n_i | i \in \{1, \ldots, z\}\}$ platí*

$$\Pi_{\mathsf{C_L}}(m) \leq \left(\frac{ezm}{q}\right)^q \tag{4.3}$$

a dále platí odhad

$$\mathtt{VC}_{dim}(\mathsf{C}_L) \leq 2q\log_2(ez). \tag{4.4}$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Cvičení 4.3.0 VC-dimenze neuronových sítí

1. *Předpokládejme, že v definičním vztahu 4.2 jsou všechna t_j rovna nule. Pro takto modifikovaný lineární sigmoidální obvod odvoďte odhad 4.4 na VC-dimenzi takovéhoto obvodu. Vysvětlete souvislost tohoto odhadu s vlastnostmi třídy konceptů $\mathsf{G}_{n,\vec{\boldsymbol{x}}}$ definované v 4.1. Objasněte souvislost s existencí či neexistencí prahových hodnot t_j v lineárním sigmoidálním obvodu.*

2. *Mějme dopředný obvod, jehož přechodové funkce $\widetilde{f}_j$ jsou tvaru*

$$\widetilde{f}_j(\vec{\boldsymbol{s}}_j) \stackrel{def}{=} \max\left\{\left|(\vec{\boldsymbol{s}}_j)_k - (\vec{\boldsymbol{w}}_j)_k\right| | k \in \{1, \ldots, n_j\}\right\} - t_j.$$

Spočtěte VC-dimenzi tohoto obvodu.

3. Mějme dopředný obvod, jehož přechodové funkce $\widetilde{f_j}$ jsou tvaru

$$\widetilde{f_j}\left(\vec{\boldsymbol{s}}_j\right) \stackrel{def}{=} \min\left\{\widetilde{sgn}\left(\left(\vec{\boldsymbol{s}}_j\right)_k - \left(\vec{\boldsymbol{w}}_j\right)_k\right) | k \in \{1,\ldots,n_j\}\right\}.$$

Spočtěte VC-dimenzi tohoto obvodu.

Kapitola 5

Teorie učení a neuronové sítě

Saganovská variace

Tvrdit, že matematické vzorce jsou jen čáry na papíře,
je stejné, jako tvrdit,
že Shakespearovy hry jsou jen slova.

Tuto kapitolu věnujeme výkladu základů tzv. teorie učení, která analyzuje možnosti extrakce vlastností množin na základě znalosti konečného počtu jejich reperezentantů. Tato teorie poskytuje základní pohled na možnost generalizace vjemů a vstupů, přicházejících do jakéhokoliv systému, který je schopen na základě dřívějších znalostí generovat hypotézy týkající se vlastností vjemů a vstupů následujících. Tato kapitola se tedy netýká pouze neuronových sítí, ale obecně analyzuje problém učení. Náš výklad bude založen zejména na teorii rozpracované v citaci ([BEHW89]). Ukážeme, že VC-dimenze třídy konceptů a VC-dimenze neuronových sítí jsou klíčové pojmy pro studium obecného problému učení. Dále zavedeme různá kriteria efektivity učících algoritmů, co se týče počtu reprezentantů nutných k realizaci daného algoritmu s požadovanou kvalitou výstupu. Ukážeme základní odhady počtu reprezentantů vzhledem ke složitosti konceptu a vzhledem k dimenzi reprezentantů konceptu pro polynomiálně složité algoritmy učení. Vyložené pojmy plynule navazují na problematiku neuronových sítí, prezentovanou v předchozích kapitolách.

5.1 Deskripce učicích algoritmů

Uvažujme problém, kdy máme zadánu jakousi množinu $\bar{A} \in \Re^n$, o které apriori víme, že $\bar{A}$ je z geometrického hlediska koule. Předpokládejme, že máme dánu posloupnost bodů $\vec{\boldsymbol{x}}_i \in \Re^n$ a o každém jednom z nich je známo, zda leží v množině $\bar{A}$ nebo ne. Naším úkolem je rozhodnout, jak vypadá množina $\bar{A}$, v tomto konkrétním případě najít střed a poloměr příslušné koule. Je zřejmé, že naše rozhodnutí je velmi závislé na vlastnostech posloupnosti $\vec{\boldsymbol{x}}_i$. Předpokládejme, že na základě nějakého algoritmu (deterministického či nedeterministického) jsme se rozhodli pro nějakou kouli v $\Re^n$, o které budeme tvrdit, že je rovna množině $\bar{A}$. Je přirozené očekávat, že v této množině by měly být obsaženy všechny body $\vec{\boldsymbol{x}}_j \in \bar{A}$ a naopak by v ní nemělo být žádné $\vec{\boldsymbol{x}}_j$, které v $\bar{A}$ neleží. Tomuto požadavku budeme říkat konzistentnost a o množině, o které se domníváme, že je rovna kouli $\bar{A}$, budeme hovořit jako o hypotéze (rovnost námi zvolené koule a $\bar{A}$ je pouze naše hypotéza, založená na vlastnostech použitého algoritmu a na posloupnosti bodů $\vec{\boldsymbol{x}}_i$).

Označme námi zvolenou kouli jako $\bar{K}$. Pak je zřejmé, že podmínka konzistentnosti musí být splněna v případě, kdy symetrická diference množin $\bar{K}$ a $\bar{A}$ bude prázdná množina, je tedy přirozené požadovat, aby toto platilo pro generovanou hypotézu, nebo aby to platilo alespoň s vysokou mírou pravděpodobnosti. Výše diskutované pojmy jsou obsahem následujících definic.

Definice 5.1.1 *Nechť* C *je třída konceptů a pro systém množin* H *platí* $\mathsf{C} \subset \mathsf{H} \subset 2^{\bar{X}}$. *Potom* H *nazveme* TŘÍDOU HYPOTÉZ *pro třídu konceptů* C.

Definice 5.1.2 *Nechť* $\breve{x} \stackrel{def}{=} \{x_1, \ldots, x_m\}$, $x_i \in \bar{X}$, $i \in \{1, \ldots, m\}$, $\vec{\boldsymbol{z}} \in \{-1, +1\}^m$ *a nechť* $\bar{c} \subset \bar{X}$. *Potom uspořádanou dvojici*

$$\left(\breve{x}, \vec{\boldsymbol{z}}\right)$$

nazveme VZOREK KONCEPTU $\bar{c}$ DÉLKY m *právě když*

$$(\forall i \in \{1, \ldots, m\}) \left((x_i \in \bar{c}) \Leftrightarrow (\vec{\boldsymbol{z}}_i = 1)\right).$$

Pro třídu konceptů C *definujme* PROSTOR VZORŮ TŘÍDY KONCEPTŮ *jako*

$$\bar{S}_{\mathsf{C}} \stackrel{def}{=} \bigcup_{m \geq 1} \left\{ \bigcup_{\bar{c} \in \mathsf{C}} \left\{ \left(\breve{x}, \vec{\boldsymbol{z}}\right) \middle| \left(\breve{x}, \vec{\boldsymbol{z}}\right) \text{ je vzorek délky } m \text{ konceptu } \bar{c} \right\} \right\}.$$

Množina $\bar{b} \subset \bar{X}$ *je* KONZISTENTNÍ MNOŽINA *se vzorkem* $\left(\breve{x}, \vec{\boldsymbol{z}}\right)$ *jestliže pro všechna* $i \in \{1, \ldots, m\}$, *platí* $\left(x_i \in \bar{b} \Leftrightarrow \vec{\boldsymbol{z}}_i = 1\right)$.

Pro další výklad musíme mít k dispozici nějakou míru odlišnosti dvou množin. Tuto míru odlišnosti definujeme takto:

Definice 5.1.3 *Nechť* $\bar{X}$ *je pravděpodobnostní prostor s hustotou pravděpodobnosti* $\widetilde{P}$. *Nechť* C *je třída konceptů definovaná nad prostorem* $\bar{X}$ *a nechť je* H *třída hypotéz pro* C . *Potom pro každé* $\bar{c} \in \mathsf{C}$ *a* $\bar{h} \in \mathsf{H}$ *číslo*

$$\mathsf{e}_{\widetilde{P}}\left(\bar{c}, \bar{h}\right) \stackrel{def}{=} Prob_{\widetilde{P}}\left(\bar{h} \triangle \bar{c}\right)$$

nazveme CHYBA HYPOTÉZY $\bar{h}$ VZHLEDEM KE KONCEPTU $\bar{c}$ *a hustotě pravděpodobnosti* $\widetilde{P}$ *(symbol* $\bar{h} \triangle \bar{c}$ *označuje symetrickou diferenci množin* $\bar{h}$ *a* $\bar{c}$ *).*

Protože množina $\bar{A}$ byla popsána pouze posloupností svých prvků (plus apriorní informací o tvaru) a pouze na základě těchto údajů jsme se měli rozhodnout, která koule je rovna množině $\bar{A}$, je náš uvažovaný učící algoritmus ve své podstatě zobrazení mezi množinou všech vzorků množiny $\bar{A}$ do množiny všech koulí v $\Re^n$. Od smysluplného učícího algoritmu nebudeme požadovat přesnou hypotézu pro všechny možné úlohy tohoto typu, nýbrž budeme pouze požadovat, aby učicí algoritmus ve většině případů produkoval "uspokojivou" hypotézu (např. hypotézu konzistentní se zadaným vzorkem). Kvantifikace tohoto požadavku je obsahem následující definice.

Definice 5.1.4 *Mějme definovánu funkci* $\widetilde{m}(\epsilon,\delta)$ *zobrazující množinu* $(0,1)\times(0,1)$ *do množiny přirozených čísel a* H *nechť je třída hypotéz pro třídu konceptů* C *definovanou nad množinou* $\bar{X}$ *. Pak* $\widetilde{P}$ -UČICÍ ALGORITMUS SLOŽITOSTI $\widetilde{m}(\epsilon,\delta)$ *třídy konceptů* C *je každé zobrazení* $\widetilde{\mathrm{A}^*}:\bar{S}_{\mathsf{C}}\to\mathsf{H}$ *takové, že pro všechna* $\bar{c}\in\mathsf{C}$*, pro všechna* $0<\epsilon<1$ *a* $0<\delta<1$ *a pro libovolnou pravděpodobnost* $\widetilde{P}$ *definovanou na* $\bar{X}$ *je pravděpodobnost množiny*

$$\left\{\breve{x}\in\bar{X}^m\,\middle|\,\left(\breve{x},\vec{\boldsymbol{z}}\right)\text{ je vzorek }\bar{c}\quad a\quad \mathsf{e}_{\widetilde{P}}\left(\bar{c},\widetilde{\mathrm{A}^*}\left(\left(\breve{x},\vec{\boldsymbol{z}}\right)\right)\right)\geq\epsilon\right\}$$

menší než číslo δ*. Jestliže takovýto učící algoritmus existuje, říkáme, že* C JE UNIFORMĚ NAUČITELNÁ *podle třídy hypotéz* H *. Každý takovýto učicí algoritmus nazveme* (ϵ,δ)-UČICÍM ALGORITMEM *.*

POZNÁMKA:
Tento model výpočetní složitosti se v literatuře označuje jako PAC učení (anglický termín: Probably Approximately Correct) .

Je zřejmé, že každou množinu lze velmi přesně popsat velkým počtem jejích prvků. Algoritmy využívající vzorků velké délky mohou zřejmě produkovat mnohem věrohodnější hypotézy. Spokojíme-li se s předem danou přesností generování hypotéz, popsanou čísly ϵ a δ, je přirozené se ptát, jak dlouhý vzorek pro danou třídu konceptů musíme použít, abychom měli tuto přesnost zaručenu. Proto zavedeme pojem vzorové složitosti učících algoritmů následovně:

Definice 5.1.5 *Minimální hodnotu* $\widetilde{m}(\epsilon,\delta)$ *, pro kterou je* A^* *učicím algoritmem, nazveme* VZOROVOU SLOŽITOSTÍ *učicího algoritmu* A^* *.*

Tato složitost je definována pouze vzhledem k délce vzorku použitého učícím algoritmem, otázka vnitřní výpočetní složitosti vlastního algoritmu zde není brána v úvahu! Nejhrubší odhad vzorové složitosti dává následující tvrzení, založené pouze na vlastnostech pravděpodobnosti (poskytuje základní odhad počtu vzorů, které musíme pro učicí algoritmus produkující konzistentní hypotézu použít).

Tvrzení 5.1.1 *Nechť* C *je třída konceptů nad konečnou množinou* $\bar{X}$ *a* $\mathsf{H}=\mathsf{C}$*. Potom pro každý učící algoritmus vyžadující*

$$\frac{1}{\epsilon}\log_e\left(\frac{|\mathsf{C}|}{\delta}\right) \tag{5.1}$$

dotazů a produkující pro daný koncept $\bar{c}\in\mathsf{C}$ *konzistentní hypotézu a pro každou hustotu pravděpodobnosti* $\widetilde{P}$ *definovanou na* $\bar{X}$ *platí*

$$Prob_{\widetilde{P}}\left(\mathsf{e}_{\widetilde{P}}\left(\bar{c},\widetilde{\mathrm{A}^*}\left(\left(\breve{x},\vec{\boldsymbol{z}}\right)\right)\right)\geq\epsilon\right)<\delta.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Poznamenejme, že odhad 5.1 není závislý na konkrétní míře pravděpodobnosti $\widetilde{P}$. To lze vysvětlit tím, že vliv různé pravděpodobnosti prvků v $\bar{X}$ je kompenzován pravděpodobností výběru tohoto prvku podle hustoty pravděpodobnosti $\widetilde{P}$. Jinak řečeno, prvky z $\bar{X}$ s velkou pravděpodobností, které v případě příslušnosti k symetrické differenci $\bar{c}\triangle\bar{h}$

přispívají významě k velikosti chyby, jsou v průměrném vzorku častější, což vzhledem ke skutečnosti, že algoritmus produkuje vždy konzistentní hypotézu, eliminuje jejich vliv. Proto je odhad 5.1 nezávislý na pravděpodobnosti $\widetilde{P}$.

V následujícím výkladu ukážeme, že tento odhad lze velmi výrazně zlepšit, právě na základě aplikace pojmu VC-dimenze.

5.2 Odhady složitosti učení a VC-dimenze

V této podkapitole se budeme zabývat dolními a horními odhady počtu vzorů, potřebných pro učící algoritmy složitosti $\widetilde{m}(\epsilon, \delta)$.

Jeden takovýto odhad na postačující množství dotazů je dán v tvrzení 5.1.1. V této kapitole ukážeme, že dominantní roli při těchto odhadech hraje právě dříve studovaný pojem VC-dimenze. Na základě tohoto pojmu lze vyslovit lepší odhady na nezbytnou délku vzorku použitého pro generování hypotézy, aproximující daný koncept.

Nejdříve se budeme zabývat zmíněnými odhady pro pevně daný systém konceptů a hypotéz. V dalším budeme uvažovat případy, kdy dané koncepty jsou rozlišeny podle dimenze prostoru, jehož podmnožinami jsou, a budeme analyzovat vzorovou složitost učících algoritmů s ohledem na dimenzi vektorových prostorů, obsahujících jako své podmnožiny dané koncepty. Přirozeným požadavkem je postulovat kritérium, které zaručí, že délka vzorku $\widetilde{m}(\epsilon, \delta)$ učícího algoritmu bude zhora omezena polynomem v dimenzi prostoru.

Závěrem provedeme obdobnou analýzu vzhledem k délce slov, která popisují koncepty z daného systému konceptů a která ve své podstatě vyjadřují svou délkou popisnou složitost konceptů.

5.2.1 Odhad počtu vzorů

V této části dokážeme sérii lemm a tvrzení, které budou nutné k důkazu tvrzení 5.2.6, udávajícího dolní a horní odhad vzorové složitosti pro (ϵ, δ)-algoritmy.

V průběhu celé této kapitoly a kapitol následujících budeme předpokládat, že každá uvažovaná třída konceptů, stejně jako každá uvažovaná třída hypotéz, bude tvořena pouze Borelovskými množinami. Pro úplnost připomeňme pojem Borelovských množin ([Jar55]). Předpokládejme, že Ω je neprázdný systém množin, který je σ-aditivní, to jest pro libovolnou posloupnost množin $\bar{A}_n \in \Omega$ platí, že $\bigcup_{n=1}^{\infty} \bar{A}_n \in \Omega$. Platí-li navíc, že pro libovolné $\bar{A}, \bar{B} \in \Omega$ je též i $\bar{A} \dot{-} \bar{B} \in \Omega$, nazveme systém množin Ω σ-aditivním okruhem. Dá se ukázat, že pro každý neprázdný systém množin Ω existuje právě jeden nejmenší σ-aditivní okruh obsahující Ω (je to právě průnik všech σ-aditivních okruhů, obsahujících systém Ω). Tento minimální σ-aditivní okruh nazveme Borelovským okruhem nad systémem Ω . Budiž $\mathcal{M}$ libovolný metrický prostor a Θ nechť je Borelovský okruh nad systémem všech otevřených množin v $\mathcal{M}$. Potom prvky tohoto okruhu nazveme BORELOVSKÝMI MNOŽINAMI prostoru $\mathcal{M}$.

Kromě požadavku ohledně Borelovskosti množin, které tvoří jednotlivé koncepty, bude nezbytné pro další důkazy detailněji charakterizovat uvažované koncepty a hypotézy. Tuto nutnou charakterizaci lze vhodně postihnout pojmem ϵ-transversála.

Definice 5.2.1 *Pro libovolné* $\mathsf{R} \subset 2^{\bar{X}}$ *a pro libovolnou hustotu pravděpodobnosti* $\widetilde{P}$ *definovanou na* $\bar{X}$ *a pro libovolné* $\epsilon > 0$ *definujme* $\mathsf{R}_{\widetilde{P},\epsilon} \stackrel{def}{=} \left\{ \bar{r} \in \mathsf{R} \,\middle|\, Prob_{\widetilde{P}}(\bar{r}) > \epsilon \right\}$. $\mathsf{R}_{\widetilde{P},\epsilon}$ *Potom* $\mathsf{N} \subset 2^{\bar{X}}$ *nazveme* ϵ-TRANSVERSÁLA *, jestliže* N *obsahuje bod z každé množiny patřící do systému* $\mathsf{R}_{\widetilde{P},\epsilon}$.

Příklad 5.2.1 *Nechť* $\bar{a} \stackrel{def}{=} \langle 0,1\rangle$ *a* C *je systém všech uzavřených intervalů na* $\bar{a}$ *. Potom množina všech bodů tvaru* ϵk, $1 \leq k \leq \frac{1}{\epsilon}$ *tvoří* ϵ*-transversálu* $\bar{a}$ *pro libovolnou hodnotu* $\epsilon > 0$ *a pro libovolnou hustotu pravděpodobnosti definovanou na* $\bar{a}$ *. Naopak je-li* C *tvořeno všemi otevřenými podmnožinami intervalu* $\bar{a}$ *, pak pro rovnoměrnou hustotu pravděpodobnosti neexistuje pro žádné* ϵ *konečná* ϵ*-transversála.*

Budeme se zajímat o pravděpodobnost výběru ϵ-transversály systému R na základě náhodného výběru bodů z $\bar{X}$. Konkrétně nás bude zajímat pravděpodobnost události popsané v následující definici.

Definice 5.2.2 *Pro každé* $m \geq 1$ *a* $\epsilon > 0$ *nechť* $\bar{Q}_{m,\epsilon}$ *označuje množinu všech* $\breve{x} \in \bar{X}^m$ *takovou, že navzájem různé prvky* $\breve{x}$ *netvoří* ϵ*-transversálu pro* R *s ohledem na hustotu* $\widetilde{P}$ *. Nechť potom* $\bar{J}_\epsilon^{2m}$ *označuje množinu všech* $\breve{x}\breve{y} \in \bar{X}^{2m}$*, kde* $\breve{x}, \breve{y} \in \bar{X}^m$ *tak, že existuje* $\bar{r} \in \mathsf{R}_{\widetilde{P},\epsilon}$*, pro kterou je* $\breve{x} \cap \bar{r} = \emptyset$ *a* $\left|\left\{i \,\middle|\, \breve{y}_i \in \bar{r}, i \in \{1,\ldots,m\}\right\}\right| \geq \frac{\epsilon m}{2}$.

V důkazech pravděpodobnostních vlastností učících algoritmů nás budou zajímat zejména vlastnosti systému množin $\mathsf{R} \stackrel{\text{def}}{=} \left\{\bar{h} \triangle \bar{c} \,\middle|\, \bar{h} \in \mathsf{H}\right\}$, kde $\bar{c}$ je nějaký pevný koncept $\bar{c} \subset \bar{X}$ a H je třída hypotéz pro tento koncept. Význam pojmu ϵ-transversála je ten, že jestliže vzorek konceptu $\bar{c}$ je současně ϵ-transversálou pro R , pak obsahuje protipříklad pro každou hypotézu, jejíž chyba vzhledem k cílovému konceptu $\bar{c}$ je větší než ϵ. Abychom mohli zkoumat vlastnosti systému R , vezmeme v úvahu třídy hypotéz splňující následující kritérium uvedené v definici.

Definice 5.2.3 *Třída hypotéz* H *je* DOBŘE UTVOŘENÁ *(anglický termín: well-behaved) jestliže množiny* $\bar{Q}_{m,\epsilon}$ *a* $\bar{J}_\epsilon^{2m}$ *jsou měřitelné pro libovolnou hustotu pravděpodobnosti* $\widetilde{P}$ *, libovolné* $m \geq 1$*,* $\epsilon > 0$ *a libovolný systém množin* $\mathsf{R} \stackrel{def}{=} \left\{\bar{h} \triangle \bar{c} \,\middle|\, \bar{h} \in \mathsf{H}\right\}$*, kde* $\bar{c}$ *je libovolná Borelovská množina.*

Výrazná většina tříd konceptů a tříd hypotéz běžně používaných je dobře utvořená. Jednou z možností verifikace této vlastnosti je ověření univerzální separibility.

Definice 5.2.4 *Třída hypotéz* $\mathsf{H} \subset 2^{\bar{X}}$ *je* UNIVERZÁLNĚ SEPARABILNÍ *, existuje-li spočetná podmnožina* T *třídy* H *tak, že pro všechny* $\bar{h} \in \mathsf{H}$ *existuje posloupnost* $\{\bar{h}_i\}_1^\infty$ *množin z* T *taková, že*

$$\left(\forall x \in \bar{X}\right)(\exists n \geq 1)\left((\forall i \geq n)\left(x \in \bar{h}_i \;\textit{ právě když }\; x \in \bar{h}\right)\right).$$

Platí implikace vyjádřená v následujícím tvrzení.

Tvrzení 5.2.1 *Je-li* H *univerzálně separabilní, pak* H *je dobře utvořená.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Dokážeme nyní následující technickou lemmu poskytující základní nerovnosti nezbytné při odhadech vzorové složitosti učících algoritmů (zaveďme konvenci, že $\widetilde{P}^m$ je hustota pravděpodobnosti definovaná na kartézském součinu $\bar{X}^m$, odvozená od hustoty pravděpodobnosti $\widetilde{P}$ definované na $\bar{X}$).

Lemma 5.2.2 *Nechť* R *je neprázdná třída konceptů na* $\bar{X}$ *a* $\widetilde{P}$ *je hustota pravděpodobnosti na* $\bar{X}$ *, pro kterou* $\bar{Q}_{m,\epsilon}$ *a* $\bar{J}_{\epsilon}^{2m}$ *jsou měřitelné pro libovolné* $m \geq 1$ *a* $\epsilon > 0$*. Potom*

1. *pro každé* $\epsilon > 0$ *a* $m \geq \frac{m}{\epsilon}$ *platí*

$$Prob_{\widetilde{P}^m}\left(\bar{Q}_{m,\epsilon}\right) < 2Prob_{\widetilde{P}^{2m}}\left(\bar{J}_{\epsilon}^{2m}\right),$$

2. *pro každé* $\epsilon > 0$ *a* $m \geq 1$ *platí*

$$Prob_{\widetilde{P}^{2m}}\left(\bar{J}_{\epsilon}^{2m}\right) \leq \Pi_{\mathsf{R}}\left(2m\right) 2^{-\frac{\epsilon m}{2}}.$$

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Předešlá lemma umožňuje získat horní odhad pravděpodobnosti události, kdy na základě náhodného vzorku nezískáme ϵ-transversálu pro třídu hypotéz R . K získání odhadu budeme potřebovat lemmu následující.

Lemma 5.2.3 *Pro každé* $m \geq 1$*,* $\bar{c} \subset \bar{X}$ *a* $\mathsf{H} \subset 2^{\bar{X}}$ *je* $\Pi_{\mathsf{H}}\left(m\right) = \Pi_{\mathsf{R}}\left(m\right)$*, kde je* $\mathsf{R} \stackrel{def}{=} \left\{\bar{h} \triangle \bar{c} \middle| \bar{h} \in \mathsf{H}\right\}$.

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Nyní vyslovme tvrzení, zhora odhadující pravděpodobnost existence hypotézy konzistentní se vzorkem konceptu $\bar{c}$ a s chybou menší než ϵ.

Tvrzení 5.2.4 *Nechť* H *je neprázdná dobře utvořená třída hypotéz nad* $\bar{X}$ *,* $\widetilde{P}$ *je hustota pravděpodobnosti na* $\bar{X}$ *a* $\bar{c} \subset \bar{X}$ *je libovolná Borelovská množina. Potom pro libovolná pevně zvolená* $\epsilon > 0$*,* $m \geq 1$ *a pro libovolný vzorek* $\left(\breve{x}, \vec{\boldsymbol{z}}\right)$ *konceptu* $\bar{c}$ *délky* m *nechť* Γ *označuje pravděpodobnost existence hypotézy* $\bar{h} \in \mathsf{H}$*, která je konzistentní s* $\left(\breve{x}, \vec{\boldsymbol{z}}\right)$ *a pro kterou je* $\mathsf{e}_{\widetilde{P}}\left(\bar{h}, \bar{c}\right) > \epsilon$*. Potom platí odhad*

$$\Gamma \leq 2\Pi_{\mathsf{H}}\left(2m\right) 2^{-\frac{\epsilon m}{2}}.$$

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Na základě lemmy 4.2.4 je zřejmé, že je-li VC-dimenze třídy hypotéz H konečná, potom číslo $\Pi_{\mathsf{H}}\left(m\right)$ je polynomiálně omezené vzhledem k m. Proto asymptotické chování výrazu $\Pi_{\mathsf{H}}\left(2m\right) 2^{-\frac{\epsilon m}{2}}$ v okolí nekonečna je dáno členem exponenciálním a zřejmě se tato hodnota velmi rychle blíží nule pro velké hodnoty m. Nyní odhadněme velikost m (tedy délku vzorku) tak, aby výraz $2\Pi_{\mathsf{H}}\left(2m\right) 2^{-\frac{\epsilon m}{2}}$ byl menší než číslo δ.

Lemma 5.2.5 *Je-li*

$$m \geq \max\left(\frac{4}{\epsilon}\log_2\left(\frac{2}{\delta}\right), \frac{8d}{\epsilon}\log_2\left(\frac{13}{\epsilon}\right)\right),$$

tak platí $2\Phi_{d,2m}2^{-\frac{\epsilon m}{2}} \leq \delta$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Lemmy a tvrzení vyslovená v předchozím umožňují již přímo dokázat základní tvrzení této podkapitoly. Vynechejme ale z našich úvah triviální případy, popsané v následující definici.

Definice 5.2.5 *Třídu konceptů* C *definovanou nad* $\bar{X}$ *nazveme* TRIVIÁLNÍ TŘÍDA KONCEPTŮ *, obsahuje-li pouze jeden koncept, nebo obsahuje-li koncepty dva, které spolu tvoří disjunktní pokrytí množiny* $\bar{X}$.

Nyní již vyslovme stěžejní tvrzení této části.

Tvrzení 5.2.6 *Nechť* C *je netriviální, dobře utvořená třída konceptů. Potom*

1. *Je-li* $\mathtt{VC}_{dim}(\mathsf{C}) = d$ *a* $d < \infty$*, tak*

   (a) *pro každé* $0 < \epsilon < \frac{1}{2}$ *žádný učící algoritmus vyžadující méně než*

   $$\max\left(\frac{1-\epsilon}{\epsilon}\log_e\left(\frac{1}{\delta}\right), d\left(1-2\left(\epsilon\left(1-\delta\right)+\delta\right)\right)\right) \tag{5.2}$$

   dotazů není pro žádný prostor hypotéz H (ϵ,δ)*-učícím algoritmem.*

   (b) *pro každé* $0 < \epsilon < 1$ *je každý učící algoritmus vyžadující alespoň*

   $$\max\left(\frac{4}{\epsilon}\log_2\left(\frac{2}{\delta}\right), \frac{8d}{\epsilon}\log_2\left(\frac{13}{\epsilon}\right)\right) \tag{5.3}$$

   dotazů a produkující konzistentní hypotézu (ϵ,δ)*-učící algoritmus.*

2. C *je uniformě naučitelná, právě když* $\mathtt{VC}_{dim}(\mathsf{C}) < \infty$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Zřejmě toto tvrzení je mnohem silnější než tvrzení 5.1.1, protože hodnota $|\mathsf{C}|$ v horním odhadu je nyní nahrazena hodnotou $\mathtt{VC}_{dim}(\mathsf{C})$, která může být podstatně menší.

5.2.2 Polynomiální učení vztažené k dimenzi

Vezměme v úvahu výše zmíněný problém nalézt v n dimenzionálním prostoru koncept (konkrétně nějakou kouli), který bude konzistentní s nějakým zadaným vzorkem. Předpokládejme, že máme k dispozici nějaký algoritmus, generující na základě vzorku odpovídající hypotézu a že tento algoritmus pracuje s přesností ϵ, δ ve smyslu uvedeném v předešlých kapitolách. Tvrzení 5.2.6 nám dává odhad na nutnou délku vzorku, který zajistí, že námi každý použitý algoritmus produkující konzistentní hypotézy bude mít přesnost popsanou čísly ϵ a δ. Nás tato délka zajímá především proto, abychom si pro danou třídu konceptů byli jisti, že snaha hledat efektivní učicí algoritmus je smysluplná v tom smyslu, že počet příkladů nutných k naučení algoritmu bude polynomiální v $\frac{1}{\epsilon}$ a $\frac{1}{\delta}$. Jsou ale obvyklé případy, kdy koncepty v dané třídě konceptů jsou navíc parametrizovány dalšími parametry, např. dimenzí konceptu (koule v prostoru dimenze n), a nás budou pochopitelně zajímat takové podmínky, které zaručí, že pro koncepty s vyšší dimenzí nevzroste neúměrně délka vzorku nezbytně nutná ke garanci (ϵ, δ) přesnosti učení. Touto problematikou se budeme zabývat nyní.

Definice 5.2.6 *Nechť $\bar{U}$ je nějaká univerzální množina, $n \geq 1$ a $\bar{X}_n \subset \bar{U}^n$. Nechť $\mathsf{C}_n \subset 2^{\bar{X}_n}$ je třída konceptů definovaná nad $\bar{X}_n$ a $\mathsf{H}_n \subset 2^{\bar{X}_n}$ je prostor hypotéz pro C_n. Předpokládejme dále, že každá H_n je dobře utvořená a obsahuje prázdnou množinu a celý prostor $\bar{X}_n$. Množinu $\mathcal{C} \stackrel{def}{=} \bigcup_{i=1}^{\infty} \mathsf{C}_n$ nazveme* SYSTÉM KONCEPTŮ *nad posloupností $\{\bar{X}_i\}_1^{\infty}$. Stejně tak množinu $\mathcal{H} \stackrel{def}{=} \bigcup_{i=1}^{\infty} \mathsf{H}_n$ nazveme* SYSTÉMEM HYPOTÉZ .

Ještě definujme prostor vzorů pro systém konceptů analogicky, jako pro prostor vzorů pro třídu konceptů.

Definice 5.2.7 *Množinu*

$$\bar{S}_{\mathcal{C}} \stackrel{def}{=} \bigcup_{i=1}^{\infty} \bar{S}_{\mathsf{C}_n}$$

nazveme PROSTOR VZORŮ SYSTÉMU KONCEPTŮ $\mathcal{C}$.

Z formálního hlediska zavedeme reprezentaci konceptů pomocí slov nějakého jazyka a budeme předpokládat, že tato reprezentace umožňuje pracovat s koncepty a jejich prvky v čase, který je polynomiální vzhledem k dimenzi jednotlivých prvků v daném konceptu, tedy vzhledem k n. To nám v dalším umožní odhlédnout od problému, jakou výpočetní složitost má samotný přístup ke konceptům a jejich prvkům, a zaměříme se pouze na problém složitosti vzhledem k počtu vzorů pro danou dimenzi konceptů. V následující definici symbol $\{0,1\}^*$ označuje množinu všech konečných posloupností tvořených nulami a jedničkami.

Definice 5.2.8 *Řekneme, že systém konceptů $\mathcal{C}$ (resp. třída konceptů C) je* POLYNOMIÁLNĚ DEFINOVANÝ SYSTÉM KONCEPTŮ *, (resp.* POLYNOMIÁLNĚ DEFINOVANÁ TŘÍDA KONCEPTŮ *) existuje-li jazyk $R \subset \{0,1\}^*$ takový, že:*

1. *Mezi jazykem R a množinou $\mathcal{C}$ (resp. C) existuje vzájemně jedoznačné zobrazení.*
2. *Existuje polynomiální algoritmus (polynomiální vzhledem k n), který rozhodne, zda slovo z $\{0,1\}^*$ délky n je v jazyce R.*

3. *Existuje polynomiální algoritmus (polynomiální vzhledem k n), který pro dané slovo $r \in R$, n a $x \in \bar{X}_n$ (resp. $x \in \bar{X}$) rozhodne, zda x leží v konceptu, který je reprezentován slovem r.*

Stejně tak definujeme POLYNOMIÁLNĚ DEFINOVANÝ SYSTÉM HYPOTÉZ *(resp.* POLYNOMIÁLNĚ DEFINOVANOU TŘÍDU HYPOTÉZ *)*.

Zřejmě například $\mathbf{HALFSPACE}_n$ je polynomiálně definovaná třída konceptů. Většina tříd konceptů přirozeně vzniklých je polynomiálně definovaná třída konceptů.

V dalším textu se budeme zabývat již pouze takovými systémy konceptů a hypotéz, které budou polynomiálně definované. Dále rozšíříme pojem naučitelnosti tříd konceptů na naučitelnost systému konceptů podle následující definice.

Definice 5.2.9 *Předpokládejme, že $\mathcal{C}$ je polynomiálně definovaný systém konceptů a $\mathcal{H}$ je polynomiálně definovaný systém hypotéz, $\mathcal{C} \subset \mathcal{H}$ a $\bar{N}$ je množina přirozených čísel. Mějme definovánu funkci $\widetilde{m}(\epsilon, \delta, n)$ zobrazující množinu $(0,1) \times (0,1) \times \bar{N}$ do množiny $\bar{N}$. Potom řekneme, že $\mathcal{C}$* JE D-POLYNOMIÁLNĚ NAUČITELNÝ SYSTÉM KONCEPTŮ PODLE *$\mathcal{H}$, existuje-li polynomiální algoritmus (vzhledem k času) A^* zobrazující $\bar{S}_{\mathcal{C}}$ do $\mathcal{H}$ tak, že pro všechna $0 < \epsilon < 1$, $0 < \delta < 1$ a $n \geq 1$ existuje přirozené číslo $\widetilde{m}(\epsilon, \delta, n)$ polynomiálně omezené vzhledem k $\frac{1}{\epsilon}$, $\frac{1}{\delta}$ a n tak, že pro všechna $n \geq 1$, pro všechny koncepty $\bar{c} \in \mathsf{C}_n$ a všechny hustoty pravděpodobnosti $\widetilde{P}$ definované na $\bar{X}_n$ je pravděpodobnost množiny*

$$\left\{ \breve{x} \in \bar{X}_n^{\widetilde{m}(\epsilon,\delta,n)} \,\middle|\, \left(\breve{x}, \vec{\boldsymbol{z}}\right) \text{ je vzorek } \bar{c} \quad a \quad \mathsf{e}_{\widetilde{P}}\left(\bar{c}, \widetilde{\mathrm{A}^*}\left(\left(\breve{x}, \vec{\boldsymbol{z}}\right)\right)\right) \geq \epsilon \right\}$$

menší než δ. Je-li $\mathcal{C}$ d-polynomiálně naučitelný systém konceptů podle $\mathcal{C}$ (tedy podle sebe sama), říkáme, že $\mathcal{C}$ je TOTÁLNĚ NAUČITELNÝ SYSTÉM KONCEPTŮ *. Minimální hodnotu $\widetilde{m}(\epsilon, \delta, n)$ s touto vlastností nazveme* VZOROVOU SLOŽITOSTÍ *učícího algoritmu A^* .*

Následující tvrzení postuluje ekvivalentní podmínku pro to, aby systém konceptů byl totálně naučitelný. Ukazuje se, že nutně musí být $\mathtt{VC}_{dim}(\mathsf{C}_n)$ polynomiálně omezena v n a musí existovat možnost, jak lze na základě nějakého nedeterministického algoritmu generovat hypotézu konzistentní s libovolným vzorkem z $\bar{S}_{\mathcal{C}}$ s pravděpodobností alespoň γ, kde $\gamma > 0$ je nezávislé na n.

Definice 5.2.10 *Nechť $\mathcal{C}$ je polynomiálně definovaný systém konceptů. Nechť pro dané $\gamma > 0$ existuje nedeterministický polynomiální algoritmus A^* zobrazující $\bar{S}_{\mathcal{C}}$ do $\mathcal{C}$ tak, že pro všechna $\left(\breve{x}, \vec{\boldsymbol{z}}\right) \in \bar{S}_{\mathcal{C}}$ je koncept $\widetilde{\mathrm{A}^*}\left(\left(\breve{x}, \vec{\boldsymbol{z}}\right)\right)$ konzistentní s $\left(\breve{x}, \vec{\boldsymbol{z}}\right)$ s pravděpodobností alespoň γ. Potom takovýto algoritmus nazveme* NEDETERMINISTICKÝ POLYNOMIÁLNÍ GENERÁTOR HYPOTÉZ *pro systém konceptů $\mathcal{C}$.*

Tvrzení 5.2.7 *Libovolný systém konceptů $\mathcal{C}$ je totálně naučitelný, právě když existuje nedeterministický polynomiální generátor hypotéz pro $\mathcal{C}$ a $\mathtt{VC}_{dim}(\mathsf{C}_n)$ je omezena polynomem v n.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

V případě, kdy systém konceptů je konstruován nad binární množinou $\{-1,+1\}$, existuje velmi užitečný vztah mezi čísly $\mathtt{VC}_{dim}(\mathsf{C}_n)$ a čísly $\log_2(|\mathsf{C}_n|)$, který je popsán v následující lemmě.

Lemma 5.2.8 *Předpokládejme, že* $\bar{X}_n \stackrel{def}{=} \{-1,+1\}^n$ *a* $\mathsf{C}_n \subset 2^{\bar{X}_n}$, $n \geq 1$. *Potom* $\mathtt{VC}_{dim}(\mathsf{C}_n)$ *roste polynomiálně vzhledem k* n, *právě když* $\log_2(|\mathsf{C_n}|)$ *roste polynomiálně v* n.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Cvičení 5.2.2 Učení systému konceptů

1. *Předpokládejme, že třídy* **P** *a* **NP** *nejsou totožné (což se obecně předpokládá, že platí). Potom na základě nějakého vybraného NP-úplného problému zkonstruujte nepolynomiálně definovaný sytém konceptů.*
2. *Předpokládejme, že* $\bar{X}_n \stackrel{def}{=} \{-1,+1\}^n$. *Zkonstruujte systém konceptů* $\mathcal{C}$, *pro který posloupnost* $\log_2(|\mathsf{C_n}|)$, $\mathsf{C}_n \subset 2^{\bar{X}_n}$, $n \geq 1$ *není omezena polynomem v* n.

5.2.3 Přibližné řešení problému pokrytí množin

Při dokazování některých vlastností učicích algoritmů vyvstává problém vybrat pro daný konečný systém konečných množin $\bar{S}_i$, $i \in \{1,\ldots,n\}$, a danou množinu $\bar{A} \subset \bigcup_{i=1}^n \bar{S}_i$, pokud možno co nejmenší počet množin $\bar{S}_i$, které mají tu vlastnost, že jejich sjednocení obsahuje množinu $\bar{A}$. Obecně se dá tento problém formulovat jako problém pokrytí konečného sjednocení konečných množin, kdy hledáme minimální počet množin tvaru $\bar{S}_i \cap \bar{A}$, jejichž sjednocení je právě rovno množině $\bar{A}$. Problém nalezení takovéhoto optimálního pokrytí (pokrytí sestávající z minimálního počtu množin) leží ve třídě NP-úplných problémů. Existuje však hladový algoritmus, který sice nenalezne optimální pokrytí, ale v čase úměrném $m\log_e(m)$ nalezne takové pokrytí množiny $\bar{A}$ mohutnosti m, které obsahuje nejvýše $(\log_e(m)+1)$-krát více množin nežli optimální pokrytí. V dalším výkladu uvidíme, že tento odhad je dostatečný pro analýzu některých vlastností učicích algoritmů. Věnujme se nyní výkladu zmíněného hladového algoritmu (viz. [Joh74]).

Definice 5.2.11 *Nechť* $\bar{S}_1,\ldots,\bar{S}_n$ *je libovolná konečná posloupnost konečných množin. Pro každé* $i \in \{1,\ldots,n\}$ *položme* $\bar{T}_{i,0} \stackrel{def}{=} \bar{S}_i$, *číslo* α_0 *nechť splňuje podmínku*

$$\left|\bar{T}_{\alpha_0,0}\right| \stackrel{def}{=} \max_{i \in \{1,\ldots,n\}} \left|\bar{T}_{i,0}\right|$$

a definujme

$$\bar{\Omega}_0 \stackrel{def}{=} \bar{S}_{\alpha_0} .$$

Potom pro každé $j \geq 1$ *a* $i \in \{1,\ldots,n\}$ *rekurzivně definujme množiny*

$$\bar{T}_{i,j} \stackrel{def}{=} \bar{T}_{i,j-1} \dot{-} \bar{T}_{\alpha_{j-1},j-1} , \tag{5.4}$$

číslo α_j *tak, aby platilo*

$$\left|\bar{T}_{\alpha_j,j}\right| \stackrel{def}{=} \max_{i\in\{1,\dots,n\}} \left|\bar{T}_{i,j}\right|$$

a množinu

$$\bar{\Omega}_j \stackrel{def}{=} \bar{\Omega}_{j-1} \cup \bar{S}_{\alpha_j} \ .$$

Základní vlastnosti právě zavedených množin popisuje následující lemma.

Lemma 5.2.9 *Nechť* $\bar{S}_1, \dots, \bar{S}_n$ *je konečná posloupnost konečných množin a* $\bar{\Omega}_j$, $j \geq 0$, *je posloupnost množin definovaná v definici 5.2.11. Potom platí následující výroky:*

1. *Existuje* l *takové, že* $\bar{\Omega}_l = \bigcup_{i=1}^n \bar{S}_i$.
2. $(\forall j \geq 0)\,(\forall k > j)\,(\forall i \in \{1,\dots,n\})$ *platí, že* $\bar{S}_{\alpha_j} \cap \bar{T}_{i,k} = \emptyset$.
3. $(\forall r \geq 0)$ *je*
$$\bigcup_{j>r} \bar{T}_{\alpha_j,r} = \left(\bigcup_{i=1}^n \bar{S}_i\right) \dot{-} \bar{\Omega}_r. \tag{5.5}$$
4. *Nechť* $r \geq 0$. *Potom platí*
$$(\forall j > r)\,(\forall i \in \{1,\dots,n\})\left(\left|\bar{T}_{i,j}\right| \leq \left|\bar{T}_{\alpha_r,r}\right|\right).$$
5. *Nechť* $q, r \geq 0$ *jsou přirozená čísla a* $\bar{M} \subset \{1,\dots,n\}$. *Navíc nechť platí*
$$\bigcup_{i\in\bar{M}} \bar{T}_{i,q+1} = \left(\bigcup_{i=1}^n \bar{S}_i\right) \dot{-} \bar{\Omega}_q. \tag{5.6}$$
Potom platí
$$\bigcup_{i\in\bar{M}} \left(\bar{T}_{i,q+1} \dot{-} \bar{T}_{i,r+1}\right) = \bar{\Omega}_r \dot{-} \bar{\Omega}_q.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Bod 1 předchozí lemmy nás opravňuje k zavedení následujícího pojmu.

Definice 5.2.12 *Nejmenší takové číslo* l, *pro které platí rovnost*

$$\bar{\Omega}_l = \bigcup_{i=1}^n \bar{S}_i \ ,$$

nazveme DÉLKOU APROXIMACE OPTIMÁLNÍHO POKRYTÍ *množiny* $\bigcup_{i=1}^n \bar{S}_i$.

Pro porovnání velikosti pokrytí, nalezeného algoritmem popsaným v definici 5.2.11, a velikosti optimálního pokrytí pro konkrétní zadanou úlohu slouží odhad vyslovený v lemmě 5.2.10.

Lemma 5.2.10 *Nechť l je délka aproximace optimálního pokrytí množiny $\bigcup_{i=1}^{n} \bar{S}_i$. Nechť dále pro nějaké $q \geq 0$ a $\bar{M} \subset \{1, \ldots, n\}$ platí rovnost*

$$\bigcup_{i \in \bar{M}} \bar{T}_{i,q+1} = \left(\bigcup_{i=1}^{n} \bar{S}_i \right) \dot{-} \bar{\Omega}_q. \tag{5.7}$$

Potom platí nerovnost

$$l - q \leq \sum_{i \in \bar{M}} \left(\sum_{z=1}^{\left|\bar{T}_{i,q+1}\right|} \frac{1}{z} \right).$$

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Přímým důsledkem vyslovené lemmy je odhad, který říká, že pokrytí nalezené analyzovaným algoritmem je v porovnání s pokrytím optimálním v nejhorším případě

$$(\log_e(m) + 1)$$

kráte větší (co do počtu množin realizujících pokrytí), kde m je maximální mohutnost množin tvořících sjednocení.

Tvrzení 5.2.11 *Nechť $\bar{S}_1, \ldots, \bar{S}_n$ je konečná posloupnost konečných množin splňujících pro nějaké přirozené m podmínku $\left|\bar{S}_i\right| \leq m$, $i \in \{1, \ldots, n\}$. Dále nechť $\bar{M} \subset \{1, \ldots, n\}$ je taková množina, že systém množin $\left\{ \bar{S}_i \middle| i \in \bar{M} \right\}$ tvoří minimální pokrytí množiny $\bigcup_{i=1}^{n} \bar{S}_i$, to jest*

$$\left|\bar{M}\right| \stackrel{def}{=} \min \left\{ \left|\bar{Q}\right| \middle| \bar{Q} \subset \{1, \ldots, n\} \text{ a } \bigcup_{i \in \bar{Q}} \bar{S}_i = \bigcup_{i=1}^{n} \bar{S}_i \right\}.$$

Nechť l je jakákoliv délka aproximace optimálního pokrytí množiny $\bigcup_{i=1}^{n} \bar{S}_i$. Potom platí odhad

$$\frac{l}{\left|\bar{M}\right|} \leq \sum_{z=1}^{m} \frac{1}{z} \leq \log_e(m) + 1.$$

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Závěrem této malé exkurze do oblasti algoritmů přibližně řešících NP-úplné problémy vyslovme tvrzení o aproximaci optimálního pokrytí množin ve tvaru, který je vhodný z hlediska analýzy vlastností učicích algoritmů pro třídy konceptů, kde jednotlivé koncepty jsou dány jako sjednocení (či průnik) konceptů s omezenou popisnou složitostí.

Tvrzení 5.2.12 *Nechť $\bar{S}_1, \ldots, \bar{S}_n$ je konečná posloupnost konečných množin a $\bar{A}$ je libovolná podmnožina $\bigcup_{i=1}^{n} \bar{S}_i$ mohutnosti m, nechť V označuje počet množin $\bar{S}_i$, tvořících minimální pokrytí množiny $\bar{A}$. Potom lze nalézt pokrytí množiny $\bar{A}$ obsahující méně než $V(\log_e(m) + 1)$ množin $\bar{S}_i$ v čase úměrném hodnotě $nmV\log_e(m)$.*

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

5.2.4 Polynomiální učení vztažené ke složitosti konceptu.

Obraťme nyní pozornost na další aspekt problematiky učení. Předpokládejme, že máme dánu třídu konceptů C a že koncepty z této třídy mají různou složitost. Přesněji, za tuto složitost budeme považovat hodnotu míry složitosti, což bude funkce definovaná pro každý koncept a vyjadující míru jeho složitosti. Budeme řešit otázku, jak odhadnout vzorovou složitost učicích algoritmů vzhledem k míře složitosti cílového konceptu. Pro jednoduchost ztotožníme v této kapitole třídu hypotéz H s třídou konceptů C, což však není újma na obecnosti, protože všechny závěry této podkapitoly se dají elementárně zobecnit i pro obecný případ $\mathsf{C} \subset \mathsf{H}$. Nejdříve definujme zmíněnou míru složitosti.

Definice 5.2.13 *Nechť C je třída konceptů nad $\bar{X}$ a nechť je definována funkce $\widetilde{ccm}$ zobrazující třídu konceptů C do množiny přirozených čísel. Pak tuto funkci nazveme* KONCEPTUÁLNÍ MÍROU SLOŽITOSTI *třídy konceptů C.*

Analogicky jako v případě dimenzionální složitosti definujme konceptuální složitost algoritmu A^* a s-polynomiální naučitelnost třídy konceptů C.

Definice 5.2.14 *Předpokládejme, že C je polynomiálně definovaná třída konceptů, $\bar{N}$ je množina přirozených čísel, a $\mathsf{H} = \mathsf{C}$ je polynomiálně definovaná třída hypotéz. Mějme definovánu funkci $\widetilde{m}(\epsilon, \delta, s)$ zobrazující množinu $(0,1) \times (0,1) \times \bar{N}$ do množiny přirozených čísel. Potom řekneme, že C* JE S-POLYNOMIÁLNĚ NAUČITELNÁ TŘÍDA KONCEPTŮ *, existuje-li polynomiální algoritmus (vzhledem k času) A^* zobrazující $\bar{S}_{\mathsf{C}}$ do H tak, že pro všechna $0 < \epsilon < 1$, $0 < \delta < 1$ a $s \geq 1$ existuje přirozené číslo $\widetilde{m}(\epsilon, \delta, s)$ polynomiálně omezené vzhledem k $\frac{1}{\epsilon}$, $\frac{1}{\delta}$ a s tak, že pro všechny koncepty $\bar{c} \in \mathsf{C}$ pro které je $\widetilde{ccm}(\bar{c}) \leq s$ a všechny hustoty pravděpodobnosti $\widetilde{P}$ definované na $\bar{X}$, je pravděpodobnost množiny*

$$\left\{ \breve{x} \in \bar{X}^{\widetilde{m}(\epsilon,\delta,s)} \,\middle|\, \left(\breve{x}, \vec{\boldsymbol{z}}\right) \text{ je vzorek } \bar{c} \quad a \quad \mathsf{e}_{\widetilde{P}}\left(\bar{c}, \widetilde{\mathrm{A}^*}\left(\left(\breve{x}, \vec{\boldsymbol{z}}\right)\right)\right) \geq \epsilon \right\}$$

menší než δ. Minimální hodnotu $\widetilde{m}(\epsilon, \delta, s)$ s touto vlastností nazveme KONCEPTUÁLNÍ SLOŽITOSTÍ *učícího algoritmu A^*.*

Naše další úvahy velmi výrazně zjednoduší zavedení pojmu efektivní prostor hypotéz, neboť nám umožní pro daný algoritmus pracovat pouze s takovou podmnožinou třídy konceptů C, pro jejíž každý koncept $\bar{a}$ existuje při zobrazení daným algoritmem vzor splňující předem zadaná omezení na míru složitosti konceptu $\bar{a}$ a na délku vzorku, nezbytného ke generování konceptu $\bar{a}$ uvažovaným algoritmem.

Definice 5.2.15 *Předpokládejme, že C je polynomiálně definovaná třída konceptů a že existuje polynomiální algoritmus (vzhledem k času) A^* zobrazující $\bar{S}_{\mathsf{C}}$ do C tak, že pro všechna $\left(\breve{x}, \vec{\boldsymbol{z}}\right) \in \bar{S}_{\mathsf{C}}$ je $\widetilde{\mathrm{A}^*}\left(\left(\breve{x}, \vec{\boldsymbol{z}}\right)\right)$ konzistentní s $\left(\breve{x}, \vec{\boldsymbol{z}}\right)$. Pro pevně zvolená přirozená čísla m, s definujme následující množiny:*

$$\mathsf{Z}_s \stackrel{def}{=} \{\bar{c} \in \mathsf{C} \,|\, \widetilde{ssc}(\bar{c}) \leq s\}$$

a

$$\mathsf{Y}_{m,s} \stackrel{def}{=} \left\{ \left(\breve{x}, \vec{\boldsymbol{z}}\right) \,\middle|\, \breve{x} \in \bar{X}^m \;\; a \;\; (\exists \bar{c} \in \mathsf{Z}_s)\left(\left(\breve{x}, \vec{\boldsymbol{z}}\right) \text{ je konzistentní s } \bar{c}\right) \right\}.$$

Potom množinu

$$\mathsf{EF}_{\mathrm{A}^*,m,s} \stackrel{def}{=} \left\{ \widetilde{\mathrm{A}^*}\left(\left(\breve{x}, \vec{\boldsymbol{z}}\right)\right) \,\middle|\, \left(\breve{x}, \vec{\boldsymbol{z}}\right) \in \mathsf{Y}_{m,s} \right\}$$

nazveme EFEKTIVNÍ PROSTOR HYPOTÉZ ALGORITMU A^* PRO KONCEPTY SLOŽITOSTI s A VZORKY DÉLKY m *(stručněji s,m-*EFEKTIVNÍ PROSTOR *).*

Definice 5.2.16 *Předpokládejme, že pro nějaký algoritmus* A^* *je definován efektivní prostor hypotéz* $\mathsf{EF}_{\mathrm{A}^*,m,s}$. *Potom algoritmus* A^* *je* OCCAMŮV ALGORITMUS , *právě když existuje číslo* α, $0 \leq \alpha < 1$ *a polynom* $\widetilde{p}(s)$ *tak, že pro všechna* $m, s \geq 1$ *platí, že*

$$\mathtt{VC}_{dim}\left(\mathsf{EF}_{\mathrm{A}^*,m,s}\right) \leq \widetilde{p}(s)m^{\alpha}.$$

Nyní dokážeme jednu technickou lemmu sloužící k důkazu tvrzení následujícího.

Lemma 5.2.13

1. *Je-li* $0 < \epsilon$, $\delta < 1$, $k \geq 1$, $0 \leq \alpha < 1$ *a*

$$m \stackrel{def}{=} \max\left(\frac{4}{\epsilon}\log_2\left(\frac{2}{\delta}\right), \left(\frac{8k}{\epsilon}\log_2\left(\frac{13}{\epsilon}\right)\right)^{\frac{1}{1-\alpha}}\right), \tag{5.8}$$

potom platí

$$\frac{2\left(\frac{2em}{km^{\alpha}}\right)^{km^{\alpha}}}{2^{\frac{m\epsilon}{2}}} \leq \delta.$$

2. *Je-li* $l \geq 1$ *a druhá část omezující* m *ve výrazu 5.8 je nahrazena výrazem*

$$\frac{2^{l+4}k}{\epsilon}\left(\log_2\left(\frac{8(2l+2)^{l+1}k}{\epsilon}\right)\right)^{l+1}, \tag{5.9}$$

pak

$$\frac{2(2m)^{k(\log_2(m))^l}}{2^{\frac{m\epsilon}{2}}} \leq \delta.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Následující tvrzení dává do vzájemné souvislosti Vapnik-Chervonenkovu dimenzi efektivního prostoru hypotéz a délku vzorku, nezbytnou pro příslušný učící algoritmus A^* .

Tvrzení 5.2.14 *Nechť* C *je třída konceptů s definovanou konceptuální mírou složitosti. Potom platí následující tvrzení.*

1. (a) *Je-li* $\mathtt{VC}_{dim}\left(\mathsf{EF}_{\mathrm{A}^*,m,s}\right) \leq \widetilde{p}(s)m^{\alpha}$ *pro nějaký polynom* $\widetilde{p}(s) \geq 1$ *a* $0 \leq \alpha < 1$, *potom* A^* *je* (ϵ, δ)*-učící algoritmus vyžaduje-li alespoň*

$$m \stackrel{def}{=} \max\left(\frac{4}{\epsilon}\log_2\left(\frac{2}{\delta}\right), \left(\frac{8\widetilde{p}(s)}{\epsilon}\log_2\left(\frac{13}{\epsilon}\right)\right)^{\frac{1}{1-\alpha}}\right) \tag{5.10}$$

dotazů.

(b) *Je-li* $\mathtt{VC}_{dim}\left(\mathsf{EF}_{\mathrm{A}^*,m,s}\right) \leq \widetilde{p}(s)\left(\log_2(m)\right)^l$ *pro nějaký polynom* $\widetilde{p}(s) \geq 2$ *a* $l \geq 1$, *potom předchozí odhad platí s tím, že druhá část je rovna výrazu*

$$\frac{2^{l+4}\widetilde{p}(s)}{\epsilon}\left(\log_2\left(\frac{8(2l+2)^{l+1}\widetilde{p}(s)}{\epsilon}\right)\right)^{l+1}. \tag{5.11}$$

2. *Existuje-li Occamův algoritmus pro* C *, pak* C *je s-polynomiálně naučitelná třída konceptů.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Definice 5.2.17 *Nechť* C *je polynomiálně definovaná třída konceptů a nechť existuje polynomiální algoritmus* A^* *zobrazující* $\bar{S}_{\mathsf{C}}$ *do* C *tak, že pro všechna* $\left(\breve{x}, \vec{\boldsymbol{z}}\right) \in \bar{S}_{\mathsf{C}}$ *je koncept* $\widetilde{\mathrm{A}^*}\left(\left(\breve{x}, \vec{\boldsymbol{z}}\right)\right)$ *konzistentní s* $\left(\breve{x}, \vec{\boldsymbol{z}}\right)$*. Potom takovýto algoritmus nazveme* POLYNOMIÁLNÍ GENERÁTOR HYPOTÉZ *pro třídu konceptů* C *. Dále definujme* PROBLÉM KONZISTENCE PRO C *jako rozhodovací problém, zda pro daný vzorek z* $\bar{S}_{\mathsf{C}}$ *existuje koncept* $\bar{c} \in \mathsf{C}$*, konzistentní s tímto vzorkem.*

Důležitou vlastností polynomiálního generátoru hypotéz je to, že existence takovéhoto generátoru zaručuje, že problém konzistence pro třídu konceptů C leží ve třídě $\mathbf{P}$ polynomiálně řešitelných problémů. Poznamenejme, že výše zavedený polynomiální generátor hypotéz je deterministický algoritmus.

Obdobně jako v analýze vlastností učicích algoritmů vzhledem k dimenzi platí i zde obdobné tvrzení, které říká, že za předpokladu existence polynomiálního generátoru hypotéz má třída konceptů C z jistého úhlu pohledu "rozumné vlastnosti" ve smyslu tvrzení následující lemmy.

Lemma 5.2.15 *Nechť třída konceptů má konečnou VC-dimenzi a problém konzistence pro třídu konceptů* C *je ve třídě* $\mathbf{P}$ *. Potom pro každou konečnou množinu* $\bar{Z} \subset \bar{X}$ *množina* $\Pi_{\mathsf{C}}\left(\bar{Z}\right)$ *všech podmnožin množiny* $\bar{Z}$ *, které jsou rozděleny třídou konceptů* C *, může být proběhnuta v čase polynomiálním vzhledem k mohutnosti množiny* $\bar{Z}$ *.*

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Vlastnost třídy konceptů popsaná v předešlé lemmě nám umožní zabývat se otázkou, zda pro danou třídu konceptů C (pro kterou například může existovat efektivní učicí algoritmus) nelze nalézt třídu konceptů, která je složitější (např. s vyšší VC-dimenzí), ale která současně od třídy C převzala vlastnosti, zaručující efektivitu učicích algoritmů pro tuto novou odvozenou třídu konceptů. Uvidíme dále, že dva způsoby odvození takovýchto složitějších tříd lze realizovat prostřednictvím konečných průniků a sjednocení konceptů z původní třídy konceptů C .

Definice 5.2.18 *Nechť* $\mathsf{C} \subset 2^{\bar{X}}$ *je třída konceptů. Označme* U_{C} *množinu všech konečných sjednocení množin z* C *a* I_{C} *množinu všech konečných průniků množin z* C *. Dále definujme pro každé přirozené číslo* k *množiny*

$$\mathsf{U}_{k,\mathsf{C}} \stackrel{def}{=} \left\{ \bigcup_{i=1}^{k} \bar{c}_i \,|\, (\forall i \in \{1,\ldots,k\})\,(\bar{c}_i \in \mathsf{C}) \right\}$$

a

$$\mathsf{I}_{k,\mathsf{C}} \stackrel{def}{=} \left\{ \bigcap_{i=1}^{k} \bar{c}_i \,|\, (\forall i \in \{1,\ldots,k\})\,(\bar{c}_i \in \mathsf{C}) \right\}.$$

Lemma 5.2.16 *Nechť* $\mathsf{C} \subset 2^{\bar{X}}$ *je třída konceptů,* $\mathtt{VC}_{dim}(\mathsf{C}) = d \geq 1$ *je konečná. Potom pro všechna* $k \geq 1$ *je* $\mathtt{VC}_{dim}(\mathsf{U}_{k,\mathsf{C}}) \leq 2dk\log_2(3k)$ *a* $\mathtt{VC}_{dim}(\mathsf{I}_{k,\mathsf{C}}) \leq 2dk\log_2(3k)$.

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Vyslovená lemma umožňuje důkaz následujícího tvrzení o s-polynomiální naučitelnosti tříd konceptů U_C a I_C.

Tvrzení 5.2.17 *Předpokládejme, že* C *je třída konceptů pro kterou existuje polynomiální generátor hypotéz, a nechť navíc* $\mathtt{VC}_{dim}(\mathsf{C})$ *je konečná. Potom třídy konceptů* U_C *a* I_C *jsou s-polynomiálně naučitelné.*

■ *Důkaz:*

Důkaz je uveden v originální tištěné verzi.

■

Kapitola 6

Numerická analýza učicích algoritmů

Finaglův numerický postulát

Poslední iterace řeší zadanou úlohu.
Nenechte se zmást fakty.

6.1 Analýza gradientních metod

Nyní se zabývejme numerickou analýzou vybraných učicích algoritmů, které využívají při nalezní cílových vah gradientních metod pro minimalizaci chybové (účelové, energetické) funkce. Tento přístup je široce používán v rozličných aplikacích neuronových sítí. Zmíníme se o tzv. delta pravidlu a různých vlastnostech jeho modifikací, dále pak stručně zmíníme nelineární zobecnění delta pravidla, jehož speciálním případem je i často používaná metoda zpětného šíření (anglický termín: back-propagation) .

6.1.1 Delta pravidlo

Nejdříve vezměme v úvahu gradientní algoritmus pro učení jednoho perceptronu, jehož váhy budeme chápat jako vektor a označovat symbolem $\vec{\boldsymbol{w}}$. Učicí algoritmus po předložení vstupního vektoru $\vec{\boldsymbol{x}}$ a příslušné výstupní hodnoty y adaptuje hodnoty vah podle formule ($\eta > 0$)

$$\delta\vec{\boldsymbol{w}} \stackrel{\text{def}}{=} \eta \left(y - \langle \vec{\boldsymbol{w}} \,|\vec{\boldsymbol{x}}\rangle\right) \vec{\boldsymbol{x}} \,, \tag{6.1}$$

což vede k iterační posloupnosti definované jako

$$\vec{\boldsymbol{w}}_{k+1} \stackrel{\text{def}}{=} \vec{\boldsymbol{w}}_k + \delta\vec{\boldsymbol{w}}_k = \vec{\boldsymbol{w}}_k + \eta \left(y - \vec{\boldsymbol{w}}_k^T \vec{\boldsymbol{x}}\right) \vec{\boldsymbol{x}} = \left(\boldsymbol{I} - \eta\vec{\boldsymbol{x}}\vec{\boldsymbol{x}}^T\right) \vec{\boldsymbol{w}}_k + \eta y \vec{\boldsymbol{x}}$$

(v dalším textu bude symbol $\vec{\boldsymbol{w}}_k$ označovat k-tý člen iterační posloupnosti, $\vec{\boldsymbol{x}}_i$ i-tý vstupní vektor). Adaptaci vah podle 6.1 budeme říkat učení podle delta pravidla. Význam tohoto učení spočívá v tom, že přírůstek váhy mezi i-tým neuronem vstupní vrstvy a neuronem výstupním je přímo úměrný jak chybě aproximace hodnoty y, tak i i-té souřadnici vektoru $\vec{\boldsymbol{x}}$. Je-li chyba aproximace y kladná, pak zvyšujeme velikost těch vah, pro které je $\vec{\boldsymbol{x}}_i$ kladné (a tedy chyba se snižuje), a snižujeme hodnoty vah, pro které je $\vec{\boldsymbol{x}}_i$ záporné (a

tedy chyba se opět snižuje). Obdobně tato argumentace platí pro případ, kdy chyba aproximace y je záporná (všechny tyto úvahy platí samozřejmě pro dostatečně malou hodnotu parametru η). Z geometrického pohledu, vzorec 6.1 vyjadřuje tu skutečnost, že při adaptaci váhy $\vec{\boldsymbol{w}}_i$ se nadrovina definovaná váhovým vektorem $\vec{\boldsymbol{w}}$ natočí do takového směru, že hodnota lokální chyby $\|y - \langle \vec{\boldsymbol{w}} \,|\vec{\boldsymbol{x}}\rangle\|$ se sníží. Problémem je, aby toto snížení nebylo na straně druhé zaplaceno příliš vysokým přírůstkem lokální chyby pro jiné vektory, na které chceme daný perceptron naučit (kritériem celkové kvality naučení se nejčastěji bere hodnota "energetické" funkce, kterou zavedeme níže).

V následujících lemmách a tvrzeních ukážeme, že pro malé hodnoty parametru η je učicí algoritmus založený na delta pravidle v jistém kontextu konvergentní a také numericky stabilní.

Lemma 6.1.1 *Nechť* $\boldsymbol{B} \stackrel{def}{=} \left(\boldsymbol{I} - \eta\vec{\boldsymbol{x}}\vec{\boldsymbol{x}}^T\right)$. *Potom* $\boldsymbol{B}$ *má pouze vlastní číslo* $1 - \eta\|\vec{\boldsymbol{x}}\|^2$, *které odpovídá vlastnímu vektoru* $\vec{\boldsymbol{x}}$, *a vlastní číslo* 1 *odpovídající vlastním vektorům v ortogonálním doplňku vektoru* $\vec{\boldsymbol{x}}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Lemma 6.1.2 *Nechť* $\boldsymbol{B} \stackrel{def}{=} \left(\boldsymbol{I} - \eta\vec{\boldsymbol{x}}\vec{\boldsymbol{x}}^T\right)$ *a* $0 \leq \eta \leq \frac{2}{\|\vec{\boldsymbol{x}}\|^2}$. *Potom je* $\|\boldsymbol{B}\| = \rho(\boldsymbol{B}) = 1$, *kde* $\rho(\boldsymbol{B})$ *označuje spektrální poloměr matice* $\boldsymbol{B}$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Matice $\boldsymbol{\Lambda}$ definovaná v následující definici bude hrát klíčivou roli při analýze zmíněné iterační posloupnosti pro váhový vektor $\vec{\boldsymbol{w}}$.

Definice 6.1.1 *Předpokládejme, že* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t \in \Re^n$. *Pro každý vektor* $\vec{\boldsymbol{x}}_p$ *definujme matici* $\boldsymbol{B}_p = \left(\boldsymbol{I} - \eta\vec{\boldsymbol{x}}_p\vec{\boldsymbol{x}}_p^T\right)$. *Potom pro každou permutaci* π *množiny* $(1, \ldots, t)$ *definujme matici*

$$\boldsymbol{\Lambda}_\pi \stackrel{def}{=} \boldsymbol{B}_{\pi(t)}\boldsymbol{B}_{\pi(t-1)}\cdots\boldsymbol{B}_{\pi(1)}$$

a vektor

$$\vec{\boldsymbol{h}} \stackrel{def}{=} y_{\pi(1)}\left(\boldsymbol{B}_{\pi(t)}\boldsymbol{B}_{\pi(t-1)}\cdots\boldsymbol{B}_{\pi(2)}\right)\vec{\boldsymbol{x}}_{\pi(1)} + \cdots + y_{\pi(t-1)}\boldsymbol{B}_{\pi(t)}\vec{\boldsymbol{x}}_{\pi(t-1)} + y_{\pi(t)}\vec{\boldsymbol{x}}_{\pi(t)}.$$

Lemma 6.1.3 *Nechť pro všechna* $p \in \{1, \ldots, t\}$ *platí* $0 \leq \eta \leq \frac{2}{\|\vec{\boldsymbol{x}}_p\|^2}$ *a nechť soubor vektorů* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t$ *generuje celý prostor* $\Re^n$. *Potom pro libovolnou permutaci* π *je* $\|\boldsymbol{\Lambda}_\pi\| < 1$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Nyní rigorózním způsobem zadefinujeme posloupnost váhových parametrů vzniklou aplikací delta pravidla, se kterou budeme pracovat dále.

Definice 6.1.2 *Předpokládejme, že* $(\vec{\boldsymbol{x}}_1, y_1), \ldots, (\vec{\boldsymbol{x}}_t, y_t)$ *je zadaná posloupnost uspořádaných dvojic z* $\Re^n \times \Re$, $t \geq 1$. *Nechť pro dané* $\eta > 0$ *je posloupnost vektorů* $\{\vec{\boldsymbol{w}}_i\}_1^\infty$ *definovaná rekurzivní předpisem*

$$\begin{array}{lll} \vec{\boldsymbol{w}}_1 \stackrel{def}{=} \vec{\boldsymbol{w}}_0 & & \vec{\boldsymbol{w}}_0 \in \Re^n, \text{ libovolné}, \\ \vec{\boldsymbol{w}}_{k+1} \stackrel{def}{=} \left(\boldsymbol{I} - \eta \vec{\boldsymbol{x}}_k \vec{\boldsymbol{x}}_k^T\right) \vec{\boldsymbol{w}}_k + \eta y_k \vec{\boldsymbol{x}}_k, & pro & k \in \{1, \ldots, t\} \\ \vec{\boldsymbol{w}}_{k+t} \stackrel{def}{=} \boldsymbol{\Lambda}_e \vec{\boldsymbol{w}}_k + \eta \vec{\boldsymbol{h}}_e & & k \geq 1 \end{array},$$

kde e *je identická permutace. Potom řekneme, že tato posloupnost vznikla iterací*

$$(\vec{\boldsymbol{x}}_1, y_1), \ldots, (\vec{\boldsymbol{x}}_t, y_t)$$

podle DELTA PRAVIDLA .

Předchozí definice popisuje proces učení perceptronu na základě předložení posloupnosti $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t$ vstupních vektorů a odpovídající posloupnosti $y_1, \ldots, y_t$ výstupních hodnot. Učení probíhá tak, že postupně cyklicky nastavujeme váhové parametry na základě vektoru $\vec{\boldsymbol{x}}_j$ a hodnoty y_j. Dále budeme předpokládat, že pořadí ve kterém předkládáme jednotlivé vektory $\vec{\boldsymbol{x}}_j$ je pevné, a v průběhu učení se nemění.

Tvrzení 6.1.4 *Nechť posloupnost* $\{\vec{\boldsymbol{w}}_i\}_1^\infty$ *vznikla iterací* $(\vec{\boldsymbol{x}}_1, y_1), \ldots, (\vec{\boldsymbol{x}}_t, y_t)$ *podle delta pravidla. Předpokládejme, že* $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t$ *generuje prostor* $\Re^n$ *a že vektor* $\vec{\boldsymbol{w}}^*$ *minimalizuje hodnotu funkce*

$$\widetilde{E}(\vec{\boldsymbol{w}}) \stackrel{def}{=} \sum_{p=1}^{t} \left(y_p - \langle \vec{\boldsymbol{w}} \,|\vec{\boldsymbol{x}}_p \rangle\right)^2. \tag{6.2}$$

Potom jestliže pro všechna $p \in \{1, \ldots, t\}$ *platí* $0 \leq \eta \leq \frac{2}{\|\vec{\boldsymbol{x}}_p\|^2}$, *konverguje posloupnost* $\{\vec{\boldsymbol{w}}_i\}_1^\infty$ *ke konečnému cyklu* $\{\vec{\tilde{\boldsymbol{w}}}_i\}_1^t$ *a platí, že* $\vec{\tilde{\boldsymbol{w}}}_i$ *je jediný pevný bod kontrahujícího zobrazení*

$$\widetilde{F_i}(\vec{\boldsymbol{w}}) \stackrel{def}{=} \boldsymbol{\Lambda}_{\pi_i} \vec{\boldsymbol{w}} + \eta \vec{\boldsymbol{h}}_{\pi_i},$$

kde permutace π_i *je definována jako* $(i+1, \cdots, t, 1, \cdots, i)$ *(tedy* π_0 *je identická permutace).*

Nechť $\vec{\boldsymbol{w}}(\eta)$ *je libovolný člen tohoto cyklu pro pevně zvolenou hodnotu parametru* $\eta > 0$. *Potom platí:*

1. $\|\vec{\boldsymbol{w}}(\eta) - \vec{\boldsymbol{w}}^*\| = o(\eta)$
2. $\left|\widetilde{E}(\vec{\boldsymbol{w}}(\eta)) - \widetilde{E}(\vec{\boldsymbol{w}}^*)\right| = o(\eta^2)$.

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Poznamenejme, že důkaz předešlých lemm a tvrzení lze modifikovat pro případ, kdy pořadí předkládání vektorů $\vec{\boldsymbol{x}}_i$ se v průběhu adaptace vektoru $\vec{\boldsymbol{w}}$ mění. Chování takovéto posloupnosti $\{\vec{\boldsymbol{w}}_i\}_1^\infty$ se oproti původní posloupnosti změní v tom, že není zajištěna konvergence k nějakému cyklu. Obecně může tato posloupnost oscilovat. Nezmění se však charakter konvergence vzhledem k parametru η, jelikož i nadále bude platit

$$\limsup_{k\to\infty} \|\vec{\boldsymbol{w}}_k(\eta) - \vec{\boldsymbol{w}}^*\| = o(\eta).$$

Stejně tak bude i nadále platit

$$\limsup_{k\to\infty} \left|\widetilde{E}(\vec{\boldsymbol{w}}_k(\eta)) - \widetilde{E}(\vec{\boldsymbol{w}}^*)\right| = o\left(\eta^2\right).$$

Zřejmě i v případě, že vektory $\vec{\boldsymbol{x}}_i$ vybíráme pro nastavení vektoru $\vec{\boldsymbol{w}}$ podle rovnoměrného rozložení pravděpodobnosti, kdy každý z vektorů $\vec{\boldsymbol{x}}_i$ má stejnou pravděpodobnost výběru, je s vysokou mírou pravděpodobnosti zachována podmínka, že vektory použité pro učení jsou předkládány v po sobě následujících permutacích, a tedy výše popsané chování iterační posloupnosti bude s vysokou mírou pravděpodobnosti platné i v tomto případě.

Závěry odvozené v této kapitole lze přenést bezprostředne i na případ, kdy síť je tvořena více perceptrony, které jsou zařazeny vedle sebe, protože v průběhu učení se změny vah u jednoho perceptronu nijak neprojeví na váhách jiných perceptronů. Tvrzení 6.1.4 tedy popisuje i konvergenci vah pro dvouvrstvou architekturu sítě.

6.1.2 Diskrétní delta pravidlo

Nyní zkoumejme diskrétní případ předešlého algoritmu. I když se v tomto případě již nejedná o gradientní algoritmus, uvádíme tento algoritmus právě zde z důvodů podobnosti s původním delta algoritmem.

Definice 6.1.3 *Předpokládejme, že* $(\vec{\boldsymbol{x}}_1, y_1), \ldots, (\vec{\boldsymbol{x}}_t, y_t)$ *je zadaná posloupnost uspořádaných dvojic z* $\Re^n \times \{-1, +1\}$, $t \geq 1$. *Nechť pro dané* $\eta > 0$ *je posloupnost vektorů* $\{\vec{\boldsymbol{w}}_i\}_1^\infty$ *definovaná následujícím rekurzivním předpisem*

1. *položme* $\vec{\boldsymbol{w}}_1 \stackrel{def}{=} \vec{\boldsymbol{0}}$
2. *zvolme index* $j_k \in \{1, \ldots, t\}$ *tak, aby platilo* $\widetilde{sgn}(\langle \vec{\boldsymbol{w}}_k | \vec{\boldsymbol{x}}_{j_k} \rangle) \neq y_{j_k}$ *a dále:*
   (a) *neexistuje-li takovéto* j_k, *položme* $\vec{\boldsymbol{w}}_{k+1} = \vec{\boldsymbol{w}}_k$,
   (b) *existuje-li takovéto* j_k, *položme*

$$\vec{\boldsymbol{w}}_{k+1} \stackrel{def}{=} \vec{\boldsymbol{w}}_k + \eta \vec{\boldsymbol{x}}_{j_k} \left(\widetilde{sgn}(\langle \vec{\boldsymbol{w}}_k | \vec{\boldsymbol{x}}_{j_k} \rangle) - y_{j_k}\right).$$

Potom řekneme, že tato posloupnost vznikla iterací $(\vec{\boldsymbol{x}}_1, y_1), \ldots, (\vec{\boldsymbol{x}}_t, y_t)$ *podle* DISKRÉTNÍHO DELTA PRAVIDLA .

Základní vlastnost diskrétního delta pravidla popisuje následující tvrzení ([ŠN96]):

Tvrzení 6.1.5 *Nechť posloupnost* $\{\vec{\boldsymbol{w}}_i\}_1^\infty$ *vznikla iterací podle diskrétního delta pravidla a nechť dále existuje* $\widehat{\vec{\boldsymbol{w}}}$ *takové, že pro všechna* $i \in \{1, \ldots, t\}$ *platí* $\widetilde{sgn}\left(\left\langle \widehat{\vec{\boldsymbol{w}}} | \vec{\boldsymbol{x}}_i \right\rangle\right) = y_i$. *Potom existuje přirozené číslo* $z > 0$, *pro které* $\vec{\boldsymbol{w}}_{z+1} = \vec{\boldsymbol{w}}_z$, *a navíc platí odhad*

$$z \leq \frac{\left\|\widehat{\vec{\boldsymbol{w}}}\right\|^2 \alpha}{4\eta^2\beta^2} + 1 \ ,$$

kde je

$$\alpha \stackrel{def}{=} \max_{i\in\{1,\dots,t\}} \left\{ \|\vec{\boldsymbol{x}}_i\|^2 \right\}$$

a

$$\beta \stackrel{def}{=} \min_{i\in\{1,\dots,t\}} \left\{ \left| \left\langle \widehat{\vec{\boldsymbol{w}}} \,|\vec{\boldsymbol{x}}_i \right\rangle \right| \right\}.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

6.1.3 Rozšířené delta pravidlo a SVD rozklad

V obou případech diskrétního i obyčejného delta pravidla se hodnota váhového vektoru měnila vždy v závislosti na chybě aproximace pro jeden vstupní vektor. Existují ale i jiné přístupy ke způsobu nastavování váhového vektoru v průběhu učení. Jedním z nich, který se často zmiňuje, je algoritmus, při kterém se hodnoty vah nastavují v závislosti na všech chybách přes všechny vstupní vektory. Tento algoritmus adaptuje váhový vektor podle následující iterační formule:

$$\vec{\boldsymbol{w}}_{k+1} \stackrel{\text{def}}{=} \vec{\boldsymbol{w}}_k - \eta \sum_{p=1}^{t} \left(\vec{\boldsymbol{x}}_p \vec{\boldsymbol{x}}_p^T \right) \vec{\boldsymbol{w}}_k + \eta \sum_{p=1}^{t} y_p \vec{\boldsymbol{x}}_p \stackrel{\text{def}}{=} (\boldsymbol{I} - \eta \boldsymbol{L})\, \vec{\boldsymbol{w}}_k - \eta \sum_{p=1}^{t} y_p \vec{\boldsymbol{x}}_p. \tag{6.3}$$

Zavedeme-li označení $\boldsymbol{\Omega} \stackrel{\text{def}}{=} \boldsymbol{I} - \eta \boldsymbol{L}$, lze předchozí rovnici psát ve tvaru

$$\vec{\boldsymbol{w}}_{k+1} = \Omega \vec{\boldsymbol{w}}_k - \eta \sum_{p=1}^{t} (y_p \vec{\boldsymbol{x}}_p). \tag{6.4}$$

Proveďme si nyní analýzu takovéto iterační posloupnosti. Dosadíme-li do rovnice 6.3 $\vec{\boldsymbol{w}}_{k+1} = \vec{\boldsymbol{w}}_k = \vec{\boldsymbol{w}}$, vidíme (viz. **??**), že řešení této rovnice je rovno vektoru, který minimalizuje hodnotu chybové funkce 6.2. Rovnice 6.3 vlastně představuje iterační posloupnost pro metodu nejmenších čtverců. Matice $\boldsymbol{L}$ je symetrická a pozitivně semidefinitní. Budeme-li navíc předpokládat, že $\{\vec{\boldsymbol{x}}_p\}_1^t$ generuje prostor $\Re^n$, je matice $\boldsymbol{L}$ pozitivně definitní, viz. lemma 6.1.6. Všechna vlastní čísla matice $\boldsymbol{\Omega}$ jsou ve tvaru $1 - \eta\lambda$, kde λ je vlastní číslo matice $\boldsymbol{L}$. Tedy pro dostatečně malé η je spektrální poloměr a odtud i indukovaná norma ($\boldsymbol{\Omega}$ je symetrická) menší než 1 a iterační posloupnost 6.4 bude tedy konvergentní na základě věty o kontrahujícím zobrazení.

Ukážeme ale, že tato konvergence obecně neprobíhá dostatečně rychle, a to i v případech separace konvexních množin, které jsou od sebe velmi vzdáleny v porovnání s jejich rozměry. Abychom demonstrovali tento fakt, uveďme následující odhady velikosti maximálního vlastního čísla matice $\boldsymbol{L}$.

Lemma 6.1.6 *Budiž $\vec{\boldsymbol{x}}_1, \dots, \vec{\boldsymbol{x}}_t \in \Re^n$ a nechť λ je maximální vlastní číslo matice*

$$\boldsymbol{L} \stackrel{def}{=} \sum_{p=1}^{t} \vec{\boldsymbol{x}}_p \vec{\boldsymbol{x}}_p^T.$$

Potom platí odhad

$$\max \left\{ \|\vec{\boldsymbol{x}}_p\|^2 \,|p \in \{1,\dots,t\} \right\} \le \lambda \;\le\; \sum_{p=1}^{t} \|\vec{\boldsymbol{x}}_p\|^2.$$

■ *Důkaz:*

DŮKAZ JE UVEDEN V ORIGINÁLNÍ TIŠTĚNÉ VERZI.

■

Předpokládejme, že $\vec{u}$ a $\vec{v}$ jsou dva ortonormální vektory v $\Re^n$ a že pro $\epsilon > 0$ a $z \geq 1$ máme dány libovolné množiny $\bar{A}$ a $\bar{B}$,

$$\bar{A} \stackrel{\text{def}}{=} \left\{ \vec{x}_p \in \Re^n, p \in \{1, \ldots, z\} \,\middle|\, \|\vec{x}_p - \vec{u}\| \leq \epsilon \right\}$$

a

$$\bar{B} \stackrel{\text{def}}{=} \left\{ \vec{y}_p \in \Re^n, p \in \{1, \ldots, z\} \,\middle|\, \left\|\vec{y}_p - \vec{v}\right\| \leq \epsilon \right\}.$$

Zřejmě pro malé hodnoty ϵ není problém množiny $\bar{A}$ a $\bar{B}$ od sebe lineárně odseparovat. Označme vlastní čísla matice $\boldsymbol{L}$ jako $0 \leq \lambda_n \leq \lambda_{n-1} \leq \cdots \leq \lambda_1$. Podle odhadů z lemmy 6.1.6 je ale zřejmé, že $\lambda_1 \geq (1-\epsilon)^2$ a současně matice $\boldsymbol{L}$ se pro $\epsilon \to 0$ blíží matici, jejíž hodnost je rovna jedné. Tedy pro $j \geq 2$ je $\lim_{\epsilon \to 0} \lambda_j = 0$. Jinými slovy, matice $\boldsymbol{L}$ je špatně podmíněná a to tak, že všechna vlastní čísla s vyjímkou maximálního (které se blíží k 1, $\epsilon \to 0$) se stále více přibližují k nule. Protože $\boldsymbol{\Omega} = \boldsymbol{I} - \eta \boldsymbol{L}$, vlastní čísla matice $\boldsymbol{\Omega}$ jsou tvaru $1 - \eta\lambda_n \geq 1 - \eta\lambda_{n-1} \geq \cdots \geq 1 - \eta\lambda_1$. To je ale pro konvergenci ten nejhorší možný případ, protože pro libovolnou počáteční iteraci se každá iterační posloupnost velmi rychle přiblíží k lineárnímu obalu vlastního vektoru příslušejícímu maximálnímu vlastnímu číslu (rychlost tohoto přiblížení je daná velikostí ostatních vlastních čísel, které jsou velmi malé, lineární operátor $\boldsymbol{\Omega}$ zúžený na ortogonální doplněk vlastního vektoru pro $1 - \eta\lambda_n$ je kontrahující s velmi malým kontrahujícím faktorem) a výsledná rychlost konvergence de fakto závisí pouze na maximálním vlastním čísle, které je poblíž hodnoty $1 - \eta$.

Zřejmě tedy individuální korekce vah podle jednotlivých vektorů použitých k učení, jako je tomu v případě původního delta pravidla, je dominantní pro efektivitu učicího algoritmu.

SVD rozklad a jev přeučení

Přestože rozšířené delta pravidlo je z numerického hlediska mnohem méně vhodné pro učení nežli původní delta pravidlo, použijeme nyní SVD (anglický termín: singular value decomposition) rozkladu k analýze jevu přeučení, který při aplikaci tohoto pravidla nastává. Důvod, proč tuto analýzu neprovedeme pro původní delta pravidlo, je ten, že tato analýza by byla technicky mnohem obtížnější. Jev přeučení spočívá v tom, že schopnost sítě generalizovat data se mnohdy s pokračující minimalizací chybové funkce snižuje, přestože vektor vah konverguje k nějaké limitní množině.

Vezměme v úvahu iterační posloupnost 6.4. Aplikujme na matici $\boldsymbol{X}$ SVD rozklad, tedy vyjádřeme matici $\boldsymbol{X}$ ve tvaru

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{S}\boldsymbol{Q}^T,$$

kde $\boldsymbol{P}$ a $\boldsymbol{Q}$ jsou ortogonální matice (mohou být různého řádu) a matice $\boldsymbol{S}$ je obecně obdélníková diagonální matice, na jejíž diagonále jsou singulární čísla matice $\boldsymbol{X}$ (což jsou vlastní čísla matice $\boldsymbol{X}^T\boldsymbol{X}$). Dosadíme-li tuto dekompozici do vztahu 6.4, obdržíme

$$\begin{aligned} \vec{\boldsymbol{w}}_{k+1} &= \left(\boldsymbol{I} - \eta \boldsymbol{P}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{P}^T\right)\vec{\boldsymbol{w}}_k - \eta\boldsymbol{X}\vec{\boldsymbol{y}} \\ &= \boldsymbol{P}\left(\boldsymbol{I} - \eta\boldsymbol{S}\boldsymbol{S}^T\right)\boldsymbol{P}^T\vec{\boldsymbol{w}}_k - \eta\boldsymbol{P}\boldsymbol{S}\boldsymbol{Q}^T\vec{\boldsymbol{y}}. \end{aligned}$$

Po zavedení substituce $\vec{\boldsymbol{z}}_k \stackrel{\text{def}}{=} \boldsymbol{P}^T \vec{\boldsymbol{w}}_k$ a $\vec{\boldsymbol{u}} \stackrel{\text{def}}{=} \boldsymbol{P}^T \vec{\boldsymbol{y}}$ lze tento výraz přepsat do tvaru

$$\vec{\boldsymbol{z}}_{k+1} = \left(\boldsymbol{I} - \eta \boldsymbol{S}\boldsymbol{S}^T\right) \vec{\boldsymbol{z}}_k - \eta \boldsymbol{S}\vec{\boldsymbol{u}}.$$

Předpokládejme, že $\boldsymbol{X}$ má r nenulových singulárních čísel, pro která platí $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_r$. Potom lze složky vektoru $\vec{\boldsymbol{z}}_{k+1}$ zapsat ve tvaru

$$\left(\vec{\boldsymbol{z}}_{k+1}\right)_i = \left(1 - \eta\nu_i^2\right) \left(\vec{\boldsymbol{z}}_k\right)_i - \eta\nu_i \vec{\boldsymbol{u}}_i$$

pro všechny indexy $i \in \{1, \ldots, r\}$ a

$$\left(\vec{\boldsymbol{z}}_{k+1}\right)_i = \left(\vec{\boldsymbol{z}}_k\right)_i$$

pro indexy $r + 1 \leq i \leq n$.

Je-li η dostatečně malé na to, aby zaručovalo konvergenci (t.j. aby $0 < 1 - \eta\nu^2 < 1$), je zřejmé, že konvergence bude mnohem rychlejší pro taková $\vec{\boldsymbol{z}}_i$, která odpovídají velkým singulárním hodnotám. Takovéto složky vektorů $\vec{\boldsymbol{z}}$ jsou pro charakterizaci dat významné a v průběhu iterování zkonvergují nejdříve. Složky vektoru $\vec{\boldsymbol{z}}$, které korespondují s malými singulárními čísly, naopak konvergují pomalu, ale mohou významně přispívat k velikosti chybové funkce. Současně nejsou pro dané vektory $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t$ signifikantní. V průběhu iterování se váhy nastaví nejdříve tak, aby daná síť separovala dominantní vlastnosti učicích vzorů, při dalším průběhu se síť snaží separovat nevýznamné vlastnosti, eventuelně šumy.

6.1.4 Zobecnění na nelineární systémy

Efektivní aplikovatelnost různých obměn delta pravidla je omezena pouze na případy, kdy separované množiny jsou lineárně separovatelné. To vyplývá z toho, že výchozím bodem pro návrhy učicích gradientních algoritmů založených na delta pravidle, je vztah 6.1. Tento výraz ale vyjadřuje změnu vah takovým způsobem, aby byla minimalizována chyba parametrizovaná polohou nějaké nadroviny, oddělující body patřící do separovaných množin. Budeme-li brát v úvahu takové neuronové sítě, jejichž dělicí plochy nejsou nadroviny ale plochy vyšších řádů, je potřeba adaptovat pravidlo 6.1 tak, aby změna vah byla úměrná chybě daného vzoru vzhledem k dělicím plochám, které odpovídají uvažovanému typu neuronových sítí. Proto nyní popíšeme způsob zobecnění delta pravidla na nelineární případ a ukážeme, že lze obdobným způsobem jako v lineárním případě analyzovat konvergentní vlastnosti iterační posloupnosti v okolí stabilních bodů, kterých ale v nelineárním případě může být (a zpravidla i je) více. Nechť M označuje celkový počet vah dané sítě, n je dimenze vstupního vektoru. Obecně síť realizuje zobrazení $\widetilde{g} : \Re^M \times \Re^n \to \Re^m$, kde m je dimenze výstupních vektorů. Pro daný vstupní vektor $\vec{\boldsymbol{x}}_p$ označme výstupní vektor $\vec{\boldsymbol{v}}_p = \widetilde{g}\left(\vec{\boldsymbol{x}}_p, \vec{\boldsymbol{w}}\right)$. Budeme předpokládat, že $\widetilde{g}$ je diferencovatelné a nechť jeho derivace je označena jako $\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{x}}, \vec{\boldsymbol{w}}\right)$. Ze spojitosti $\widetilde{g}$ plyne po aplikaci Tylerova rozvoje, že

$$\widetilde{g}\left(\vec{\boldsymbol{w}} + \delta\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}\right) = \widetilde{g}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}\right) + \widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}\right)\delta\vec{\boldsymbol{w}} + o\left(\|\delta\vec{\boldsymbol{w}}\|\right).$$

Pro pevný vstupní vektor $\vec{\boldsymbol{x}}_p$ nechť $\vec{\boldsymbol{y}}_p$ označuje požadovaný výstupní vektor. Potom označme chybu aproximace tohoto výstupního vektoru jako

$$\epsilon_p^2 \stackrel{\text{def}}{=} \left(\vec{\boldsymbol{y}}_p - \vec{\boldsymbol{v}}_p\right)^T \left(\vec{\boldsymbol{y}}_p - \vec{\boldsymbol{v}}_p\right) \stackrel{\text{def}}{=} \vec{\boldsymbol{q}}_p^T \vec{\boldsymbol{q}}_p$$

a celkovou chybovou funkci naučení vzorů $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t$ jako

$$\widetilde{E}\left(\vec{\boldsymbol{w}}\right) \stackrel{\text{def}}{=} \sum_{i=1}^{t} \epsilon_i^2.$$

Zřejmě při změně váhového vektoru o hodnotu $\delta\vec{\boldsymbol{w}}$ se změní chyba pro $\vec{\boldsymbol{x}}_p$ o hodnotu

$$\delta\epsilon_p^2 = \left(\vec{\boldsymbol{q}}_p + \delta\vec{\boldsymbol{q}}_p\right)^T \left(\vec{\boldsymbol{q}}_p + \delta\vec{\boldsymbol{q}}_p\right) - \vec{\boldsymbol{q}}_p^T\vec{\boldsymbol{q}}_p = 2\left(\delta\vec{\boldsymbol{q}}_p\right)^T \vec{\boldsymbol{q}}_p + \left(\delta\vec{\boldsymbol{q}}_p\right)^T \delta\vec{\boldsymbol{q}}_p.$$

Protože $\vec{\boldsymbol{y}}_p$ je konstantní, platí

$$\delta\vec{\boldsymbol{q}}_p = -\delta\vec{\boldsymbol{v}}_p = -\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}_p\right)\delta\vec{\boldsymbol{w}} + o\left(\|\delta\vec{\boldsymbol{w}}\|\right).$$

Tedy

$$\begin{aligned} \delta\epsilon_p^2 &= -2\left(\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}_p\right)\delta\vec{\boldsymbol{w}}\right)^T \left(\vec{\boldsymbol{y}}_p - \widetilde{g}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}_p\right)\right) + o\left(\|\delta\vec{\boldsymbol{w}}\|\right) \\ &= -2\left(\delta\vec{\boldsymbol{w}}\right)^T \left(\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}_p\right)\right)^T \left(\vec{\boldsymbol{y}}_p - \widetilde{g}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}_p\right)\right) + o\left(\|\delta\vec{\boldsymbol{w}}\|\right). \end{aligned} \tag{6.5}$$

Člen $o\left(\|\delta\vec{\boldsymbol{w}}\|\right)$ můžeme vynechat a potom je podle 6.5 pro fixovanou hodnotu $\|\delta\vec{\boldsymbol{w}}\|$ dosaženo maximálního úbytku hodnoty $\delta\epsilon_p^2$, právě když bude platit

$$\delta\vec{\boldsymbol{w}} = \eta\left(\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}_p\right)\right)^T \left(\vec{\boldsymbol{y}}_p - \widetilde{g}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}_p\right)\right). \tag{6.6}$$

Tato formule představuje zobecněné delta pravidlo, čehož lze snadno nahlédnout, položíme-li $\widetilde{g}\left(\vec{\boldsymbol{w}}, \vec{\boldsymbol{x}}\right) = \vec{\boldsymbol{w}}^T\vec{\boldsymbol{x}}$. Po dosazení do 6.6 dostaneme původní delta pravidlo popsané výrazem 6.1.

Na základě vzorce 6.6 dostáváme iterační posloupnost pro nelineární delta pravidlo ve tvaru

$$\begin{aligned} \vec{\boldsymbol{w}}_{k+1} &= \vec{\boldsymbol{w}}_k + \delta\vec{\boldsymbol{w}}_k \\ &= \vec{\boldsymbol{w}}_k + \eta\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}_k, \vec{\boldsymbol{x}}_p\right)^T \left(\vec{\boldsymbol{y}}_p - \widetilde{g}\left(\vec{\boldsymbol{w}}_k, \vec{\boldsymbol{x}}_p\right)\right). \end{aligned}$$

Oproti lineárnímu případu, nemá chybová funkce $\widetilde{E}\left(\vec{\boldsymbol{w}}\right)$ obecně pouze jedno globální minimum, ale může mít lokální minima, což je obecně potvrzeno (např. při učení metodou zpětného šíření). Zaměřme se proto na chování posloupnosti $\vec{\boldsymbol{w}}_k$ poblíž některého lokálního minima. Budeme-li předpokládat spojitost a omezenost derivace $\boldsymbol{D}$ na nějakém okolí lokálního minima $\widehat{\vec{\boldsymbol{w}}}$ (což je zaručeno např. požadavkem, aby zobrazení $\widetilde{g}$ mělo spojité derivace do druhého řádu včetně), lze psát

$$\begin{aligned} \vec{\boldsymbol{w}}_{k+1} &= \vec{\boldsymbol{w}}_k + \eta\left(\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}_k, \vec{\boldsymbol{x}}_p\right)\right)^T \left(\vec{\boldsymbol{y}}_p - \widetilde{g}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right) - \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\left(\vec{\boldsymbol{w}}_k - \widehat{\vec{\boldsymbol{w}}}\right)\right) + o\left(\left\|\vec{\boldsymbol{w}}_k - \widehat{\vec{\boldsymbol{w}}}\right\|\right) \\ &= \left(\boldsymbol{I} - \eta\left(\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}_k, \vec{\boldsymbol{x}}_p\right)\right)^T \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\right)\vec{\boldsymbol{w}}_k + \eta\left(\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}_k, \vec{\boldsymbol{x}}_p\right)\right)^T \times \\ &\times \left(\vec{\boldsymbol{y}_p} - \widetilde{g}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right) + \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\widehat{\vec{\boldsymbol{w}}}\right) + o\left(\left\|\vec{\boldsymbol{w}}_k - \widehat{\vec{\boldsymbol{w}}}\right\|\right). \end{aligned}$$

Iterační matice $\left(\boldsymbol{I} - \eta\left(\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}_k, \vec{\boldsymbol{x}}_p\right)\right)^T \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\right)$ není obecně symetrická, ale pro $\vec{\boldsymbol{w}}_k \to \widehat{\vec{\boldsymbol{w}}}$ se symetrické matici blíží. Dále budeme předpokládat, že matice $\widetilde{\boldsymbol{D}}\left(\vec{\boldsymbol{w}}_k, \vec{\boldsymbol{x}}_p\right)$ je Lipstichovská vzhledem k proměnné $\vec{\boldsymbol{w}}$ a stejnoměrně spojitá vzhledem k proměnné $\vec{\boldsymbol{x}}$ na nějakém okolí bodu $\widehat{\vec{\boldsymbol{w}}}$. Za těchto předpokladů platí

$$\begin{aligned} \vec{\boldsymbol{w}}_{k+1} &= \left(\boldsymbol{I} - \eta\left(\widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\right)^T \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\right)\vec{\boldsymbol{w}_k} + \eta\left(\widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\right)^T \times \\ &\times \left(\vec{\boldsymbol{y}_p} - \widetilde{g}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right) + \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\widehat{\vec{\boldsymbol{w}}}\right) + o\left(\left\|\vec{\boldsymbol{w}}_k - \widehat{\vec{\boldsymbol{w}}}\right\|\right). \end{aligned}$$

Předpokládejme, že obdobně jako v lineárním případě i zde iterujeme cyklicky přes vektory $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_t$. Ukážeme, že i v tomto případě pro malé hodnoty η je lineární část (bez členu $o\left(\left\|\vec{\boldsymbol{w}}_k - \widehat{\vec{\boldsymbol{w}}}\right\|\right)$) zobrazení převádějící $\vec{\boldsymbol{w}}_k$ na $\vec{\boldsymbol{w}}_{k+t}$ kontrahující. Předně pro $p \in \{1, \ldots, t\}$ je matice

$$\widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)^T \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right) \tag{6.7}$$

pozitivně semidefinitní. Lze tedy volit η dostatečně malé tak, aby platilo

$$\left\| \boldsymbol{I} - \eta \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)^T \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right) \right\| \leq 1.$$

Protože matice 6.7 je pozitivně semidefinitní, můžeme rozložit prostor $\Re^n$ vstupních vektorů na direktní součet nulového prostoru operátoru daného maticí 6.7 a jeho ortogonálního doplňku. Složka vektoru $\vec{\boldsymbol{w}}_{k+1}$ ležící v nulovém prostoru je totožná s odpovídající složkou vektoru $\vec{\boldsymbol{w}}_k$. Stačí nám tedy ukázat, že operátor $\boldsymbol{I} - \eta \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)^T \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)$ zúžený na zmíněný ortogonální doplněk je kontrahující. K tomu stačí zvolit η tak, aby platilo

$$\eta < \rho\left(\widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)^T \widetilde{\boldsymbol{D}}\left(\widehat{\vec{\boldsymbol{w}}}, \vec{\boldsymbol{x}}_p\right)\right)^{-1}.$$

Tato volba zaručí, že v okolí lokálního minima chybové funkce $\widetilde{E}\left(\vec{\boldsymbol{w}}\right)$ se bude iterační posloupnost chovat obdobným způsobem, jako v lineárním případě, který je popsán na začátku této kapitoly.

Část II

Relevantní výsledky související s neuronovými sítěmi

(autor Martin Holeňa)

Celá druhá část je uvedena v originální tištěné verzi.

Literatura

[AB92] Martin Anthony and Norman Biggs. *Computational Learning Theory.* Press Syndicate of the University of Cambridge, 1992.

[BEHW89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis Dimension. *Journal of the Association for Computing Machinery*, 36:929–965, oct 1989.

[Blo93] Arthur Bloch. *Murphyho zákon.* Nakladatelství svoboda, Libertas, Praha, 1993.

[Ell94] S. W. Ellacott. Aspects of the numerical analysis of neural networks. *Acta Numerica*, pages 145–202, 1994.

[Fie81] M. Fiedler. *Speciální matice a jejich použití v numerické matematice.* Edice Teoretická knižnice inženýra. Státní nakladatelství technické literatury, 1981.

[GJ79] M. R. Garey and D. S. Johnson. *Computers and Intractability.* W.H.Freeman & Company, 1979.

[Hås94] Johan Håstad. On the size of weights for threshold gates. *Siam Journal Discrete Mathematics*, 7(3):484–492, aug 1994.

[Jar55] V. Jarník. *Integrální počet II.* NČAV Praha, 1955.

[Joh74] David S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9(3):256–278, dec 1974.

[KA77] L. V. Kantorovič and G. P. Akilov. *Funkcionalnyj analyz.* Nauka Moskva, 1977.

[Kol57] A. N. Kolmogorov. O predstavlenii nepreryvnych funkcij neskolkich peremennych v vide superpozicij nepreryvnych funkcij odnogo peremennogo i slozenija. *Doklady akademii nauk SSSR*, 114(5):953–956, 1957.

[Lej85] K. Lejchtvějs. *Vypuklije množestva.* Moskva Nauka, 1985.

[Mišík89] Ladislav Mišík. *Funkcionálna analýza.* ALFA Bratislava, 1989.

[Mur71] Saburo Muroga. *Threshold Logic and Its Applications.* John Wiley & Sons, 1971.

[RSO94] Vwani Roychowdhury, Kai-Yeung Siu, and Alon Orlitsky. *Theoretical Advances in Neural Computation and Learning.* Kluwer Academic Publishers, 1994.

[Sau72] N. Sauer. On the Density of Families of Sets. *Journal of the Association for Computing Machinery*, 13:145–147, feb 1972.

[ŠJ96] Miroslav Šnorek and Marcel Jiřina. *Neuronové sítě a neuropočítače*, chapter Kapitola 2. ČVUT, 1996.

[Smi65] David R. Smith. Bounds on the number of threshold functions. *IEEE Transactions on Electronic Computers*, pages 368–369, 1965.

[ŠN96] Jiří Šíma and Roman Neruda. *Teoretické otázky neuronových sítí*. MATFYZPRESS, MFF UK Praha, 1996.

[Spr93] David A. Sprecher. A universal mapping for Kolmogorov's superposition theorem. *Neural Networks*, 6:1089–1094, 1993.

[YI65] S. Yajima and T. Ibaraki. A lower bound of the number of threshold functions. *IEEE Transaction on Electronic Computers*, pages 926–929, 1965.

Rejstřík