Support Vector Machine

Jan Faigl   Luboš Svoboda   Martin Žoldák

6. června 2001

Obsah

1 Úvod a motivace

Pro klasifikaci do dvou tříd mužeme použít klasifikatoru založeném na perceptronu. Jednou z nevýhod takového klasifikátoru je nalezení rozdělující nadplochy bez dalšího kritéria, výsledek je v podstatě závislý na průběhu učení, jakým způsobem jsou vybírány prvky z trénovací množiny. V případě klasifikátorů SVM je tato rozdělující nadplocha volena podle optimalizujícího kritéria.

V případě neseparabilních dat, lze použít stejného triku jako u učení klasifikátorů na principu perceptronu. Přechodem do vyšší dimenze prostoru, kde jsou data již separabilní.

Základní rozdělení SVM:

Lineární

- separabilní data
- neseparabilní data

Nelineární SVM

Další vlastností SVM je možnost vyčíslit chybu na testovací množině, pomocí „VC confidence“, z Vapnikovy teorie statistického učení.

2 Chyby klasifikátoru

Problém klasifikace do dvou tříd si můžeme představit jako nalezení funkce, která přiřadí naměřeným příznakům správnou třídu.

Množinu příznaků $\boldsymbol{x_i} \in R^n$, kde $i = 1, \ldots, l$ je počet prvků v trénovací množině klasifikujeme do dvou tříd $-1, 1$. Pro každé $\boldsymbol{x_i}$ známe y_i, což je třída, do které patří. Předpokládáme, že jednotlivá měření příznaků jsou nezávislá.

Úkolem učení klasifikátoru je nalezení transformační funkce z prostoru $\boldsymbol{x_i}$ do prostoru y_i.

Úkolem učení klasifikátoru je nalezení transformační funkce z prostoru Transformace $\boldsymbol{x_i} \to y_i$ je závislá na obecné hodnotě parametru α. Tento parametr je závislý na volbě klasifikátoru a také na metodě učení klasifikátoru.

$$\boldsymbol{x_i} \to f(\boldsymbol{x_i}, \alpha)$$

Bez znalosti chyby klasifikátorů je jeho použití obtížné, neboť nevíme jak moc se můžeme na výsledky spolehnout.

Očekávanou chybu klasifikátoru na skutečných datech po naučení lze vyjádřit:

$$R(\alpha) = \int \frac{1}{2}|y - f(\boldsymbol{x}, \alpha)| dP(\boldsymbol{x}, y) \tag{1}$$

Jesliže existuje sdružená hustota pravděpodobnosti pak

$$dP(\boldsymbol{x}, y) = p(\boldsymbol{x}, y) d\boldsymbol{x} dy$$

$R(\alpha)$ z (1) se nazývá očekávané riziko. Tato hodnota uvádí skutečnou chybu, ale pravděpodobnosti $P(\boldsymbol{x}, y)$ nebo její odhad, lze získat velmi těžko.

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^{l} |y_i - f(\boldsymbol{x_i}, \alpha)|. \tag{2}$$

$R_{emp}(\alpha)$ z (2) se nazývá empirická ztráta a udává chybu na trenovací množině, pro konečný počet pozorovaní. Hodnota $\frac{1}{2}|y_i - f(\boldsymbol{x_i}, \alpha)|$ se nazývá ztráta. Protože klasifikaci provádíme do tříd $-1, 1$ může nabývat pouze hodnot 0 nebo 1. Zvolíme-li η takové, že $0 \leq \eta \leq 1$, nabývá ztráta hodnoty η s pravděpodobností $1 - \eta$.

Samotná chyba na trenovací množině mnoho nevypovídá o chybě na testovacích (reálných) datech. Následující vztah dává odhad skutečné chyby větší než vztah (1), zato jej můžeme vyčíslit aniž bychom znali $P(\boldsymbol{x}, y)$

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\left(\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}\right)} \tag{3}$$

Kde h je celé nezáporné číslo nazývané VC (Vapnik-Chervonenkis) dimenze. Odmocnina ve vztahu (3) se nazývá „VC confidence“ (Vapnik 1995).

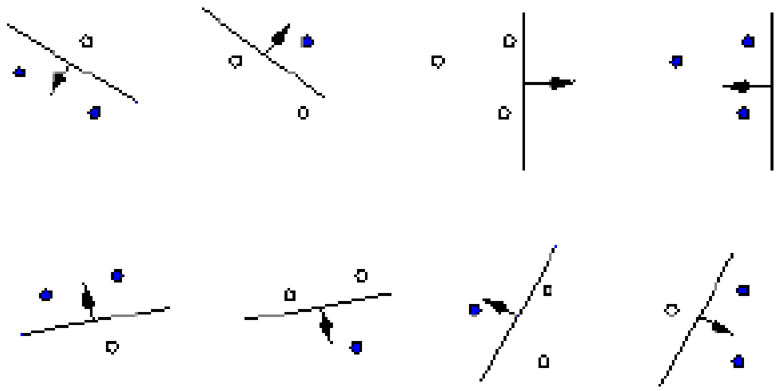
2.1 VC dimenze

VC dimenze je vlastnost množiny funkcí $f(\alpha)$; α představuje obecné parametry v závislosti na volbě funkce f. Pro klasifikaci do dvou tříd je VC dimenze definována jako maximální počet bodů, jež lze rozbít funkcí $f(\alpha)$.

Mějme množinu l bodů, kterým přiřazujeme označení $-1, 1$ existuje 2^l možných přiřazení (označení).

Jestliže pro každé označení l bodů existuje funkce z $f(\alpha)$ pak $f(\alpha)$ rozbíjí množinu bodů l.

VC dimence h zaručuje, že existuje alespoň jedna množina h bodů, které lze rozbít.



Obrázek 1: příklad rozbití množiny 3 bodů v rovině

2.2 VC dimenze a počet parametrů

Mohlo by se zdát, že hodnota VC dimenze závisí na počtu parametrů α rozbíjející funkce. Následující příklad ukáže funkci s jediným parametrem α a s hodnotu VC dimenze nekonečno.

Mějme prahovu funkci :

$$\theta(x), x \in (R) : \theta(x) = 1 \forall x > 0; \theta(x) = -1 \forall x \leq 0$$

a jedno parametrickou funkci

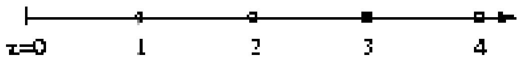
$$f(x, \alpha) = \theta(\sin(\alpha x)), x, \alpha \in \boldsymbol{R}.$$

Zvolíme nějaká čísla l a současně najdeme l bodů, které mohou být rozbity. Body vybereme podle předpisů:

$$x_i = 10^{-i}, i = 1, \ldots, l.$$

Bodům přiřadíme třídy $y_1, y_2, \ldots, y_l; y_i \in -1, 1.$

$f(x, \alpha)$ přiřazující bodům příslušnost ke třídě, a pro $\alpha = \pi(1+\sum_{i=1}^{l} \frac{(1-y_i 10^i}{2})$.
Takovýto klasifikátor rozbíjí nekonečně mnoho bodů a má tedy VC dimenzi nekonečno. Jistě lze najít množinu čísel, která touto funkcí nelze rozbít, bez ohledu na nekonečnou VC dimenzi. Ta nám však zaručuje, že existje alespoň jedna množina bodů jež lze rozbít.



Obrázek 2: příklad množiny čísel jež nelze rozbít funkcí $f(x, \alpha) = \theta(\sin(\alpha x)), x, \alpha \in \boldsymbol{R}$.

2.3 VC dimenze orientované nadroviny

Pro m bodů z $\boldsymbol{R}^n$ zvolme jeden bod za počátek souřadnic. m bodu lze rozbít orientovanými nadrovinami právě když zbylých $m-1$ bodů je lineárně nezavislých.

VC dimenze orientovaných nadrovin v $\boldsymbol{R}^n$ je $n+1$.

2.4 VC dimenze orientované nadroviny, separabilní data

Algoritmus SVM je založen na struktuře množin funkcí rozdělujících nadrovin. Popišme si množiny v prostoru Z. Množina vektorů $\boldsymbol{x_1}, \ldots, \boldsymbol{x_r} \in \boldsymbol{Z}$.

Každá nadrovina $\boldsymbol{x} \in \boldsymbol{Z} : (\boldsymbol{w} \cdot \boldsymbol{x} + b = 0)$ odpovídá dvojici $(\boldsymbol{w}, b) \in \boldsymbol{X}\mathrm{x}\boldsymbol{R}$.

Mějme podmínku na charakteristickou vzdálenost bodů $\boldsymbol{x_1}, \ldots, \boldsymbol{x_r}$ od rozdělovací nadroviny

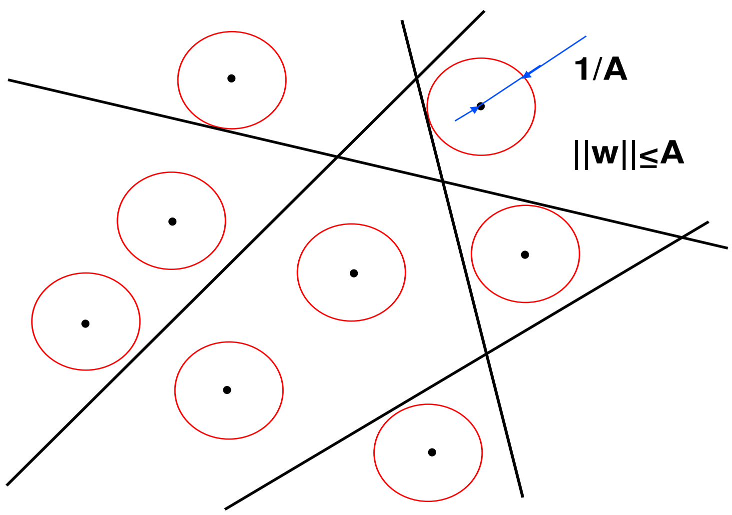
$$\max_{i=1,\ldots,r} |(\boldsymbol{w} \cdot \boldsymbol{x}) + b) = 1 \tag{4}$$

Nejmenší koule obsahující body $\boldsymbol{x_1}, \ldots, \boldsymbol{x_r}$, $B_{\boldsymbol{x_1},\ldots,\boldsymbol{x_r}} = \boldsymbol{x} \in \boldsymbol{Z} : |(\boldsymbol{w} - \boldsymbol{a}| \; < \; R$ $\boldsymbol{a} \in \boldsymbol{Z}$ a rozhodovací funkce definovaná na těchto bodech

$$f_{\boldsymbol{w},b} = sgn((\boldsymbol{w} \cdot \boldsymbol{x}) + b) \tag{5}$$

množina $f_{\boldsymbol{w},b} : |\boldsymbol{w}| \leq A$ má VC-dimenzi h (Vapnik, 1995) :

$$h \leq R^2 A^2. \tag{6}$$

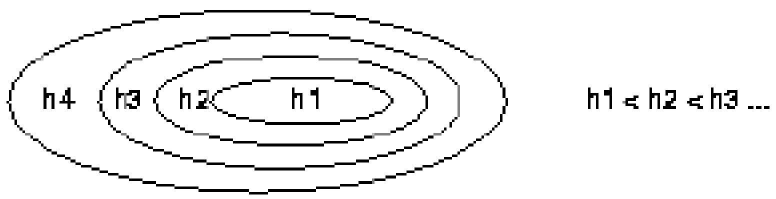


Obrázek 3: rozdělující nadroviny a koule obsahující body

2.5 Minimalizace strukturálního rizika

VC confidence v rovnici (3) závisí na volbě klasifikační funkce. Měli bychom vybrat takovou podmnožinu z množiny funkcí, pro kterou bude odhad chyby minimální, pro každou podmnožinu zjistíme její VC dimenzi. Minimalizace strukturálního rizika je hledání takové podmnožiny funkcí, aby hodnota empirického risku a hodnota VC confidence byla minimální.

Na obrázku 4 jsou zobrazeny podmnožiny s roustoucí VC dimenzí.

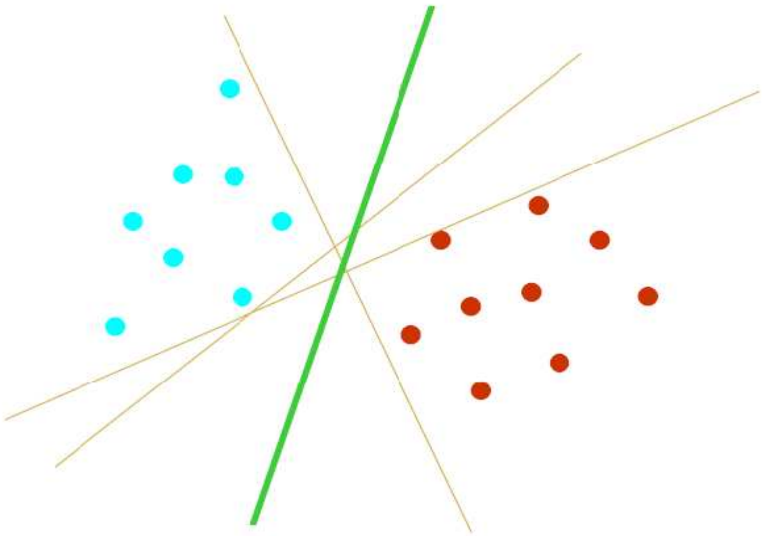


Obrázek 4: podmnožiny funkcí uspořádané podle VC dimenze

3 Separabilní data

Problém klasifikace je omezen na dvě třídy, jinak je obecný. Podívejme se na příklad na obr.5. Zde je mnoho schopných lineárních klasifikátorů, které umí rozdělit data, ale jen jeden maximalizuje „mezeru“ (tedy vzdálenost mezi nejbližšími body z obou tříd k rozdělující nadrovině). Tento lineární klasifikátor bývá nazýván optimální rozdělující nadrovinou.

Tedy hledáme lineární klasifikátor , jehož učení je založeno na minimalizaci strukturálního rizika, nalezením rozdělujícího pásu (mezery) maximální šířky.



Obrázek 5: Optimální rozdělující nadrovina

3.1 Optimální rozdělující nadrovina

Uvažujeme trénovací množinu ve tvaru $\{\boldsymbol{x_i}, y_i\}$, i=*1,...,l* , $y_i \in \{-1, 1\}$, $\boldsymbol{x_i} \in R^d$, kde y_i je třída (pozitivní a negativní vzorky) a $\boldsymbol{x_i}$ je vektor souřadnic d-dimenzionálního prostoru. Předpokládejme, že máme nějakou nadrovinu, která nám odděluje pozitivní vzorky od negativních (oddělující nadrovinu). Body (vzorky) $\boldsymbol{x}$ ležící na oddělující nadrovině vyhovují rovnici $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$, kde $\boldsymbol{w}$ je normálový vektor oddělující nadroviny (též si lze představit jako vektor vah) a $|b| / |\boldsymbol{w}|$ je vzdálenost od nadroviny k počátku souřadné soustavy, viz. obr.6.

Nechť $d_+(d_-)$ jsou nejkratší vzdálenosti od oddělující nadroviny k pozitivním (negativním) bodům.

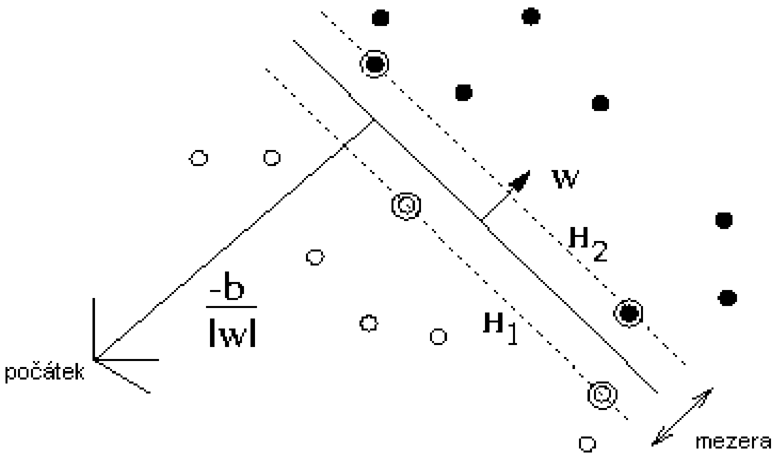
Definujme mezeru oddělující nadroviny: $\rho(\boldsymbol{w}, b) = d_+ + d_-$. Úlohu můžeme formulovat následovně: předpokládejme, že všechna trénovací data vyhovují následujícím podmínkám:

$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b \geq +1 \quad pro \quad y_i = +1 \tag{7}$$

$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b \leq -1 \quad pro \quad y_i = -1 \tag{8}$$

Tyto dvě podmínky lze též formulovat následovně:

$$y_i \cdot (\boldsymbol{x_i} \cdot \boldsymbol{w} + b) - 1 \geq 0 \quad \forall i \tag{9}$$



Obrázek 6: Lineární rozdělující nadrovina pro dvoudimenzionální prostor

Tedy vyrobíme dva pásy o šířce 1, a změnou parametru $\boldsymbol{w}$ budeme měnit šířku mezery, neboli měřítko námi zvolené soustavy. Je vhodné, aby všechna data z trénovací množiny spadala do tohoto měřítka. Nyní vezmeme v úvahu body, pro které platí podmínka (7).
Tyto body jsou odděleny nadrovinou H_1:

$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b = 1 \tag{10}$$

s normálou $\boldsymbol{w}$ a vzdáleností od počátku $|1 - b| \,/\, |\boldsymbol{w}|$. Podobně pro body splňující podmínku (8) platí pro H_2 :

$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b = -1 \tag{11}$$

Kde je vzdálenost od počátku rovna $|-1-b| / |\boldsymbol{w}|$. Poznamenejme, že H_1 a H_2 jsou paralelní (mají stejné normály), a že mezi nimi neleží žádná trénovací data. Body $\boldsymbol{x_i}$ splňující podmínky (9) resp. (10) leží na nadrovinách H_1 resp. H_2 se nazývají **support vectors** a jsou v obr.6 označeny kroužkem.Tedy hledáme body $\boldsymbol{x_i}$, které vyhovují podmínce

$$\min_{\boldsymbol{x_i}} |\boldsymbol{x} \cdot \boldsymbol{w} + b| = 1 \tag{12}$$

Vzdálenost $d(\boldsymbol{w}, b; \boldsymbol{x})$ od $\boldsymbol{x}$ k nadrovině $\boldsymbol{w}, b$ je

$$d(\boldsymbol{w}, b; \boldsymbol{x}) = \frac{|\boldsymbol{w} \cdot \boldsymbol{x} + b|}{|\boldsymbol{w}|} \tag{13}$$

Tedy optimální oddělující nadrovina maximalizující mezeru nám dává velikost mezery

$$\begin{aligned}
&\rho(\boldsymbol{w}, b) = d_+ + d_- = \min_{\{\boldsymbol{x_i}:y_i=1\}} d(\boldsymbol{w}, b; \boldsymbol{x_i}) + \min_{\{\boldsymbol{x_i}:y_i=-1\}} d(\boldsymbol{w}, b; \boldsymbol{x_i}) \\
&= \min_{\{\boldsymbol{x_i}:y_i=1\}} \tfrac{|\boldsymbol{w}\cdot\boldsymbol{x}+b|}{|\boldsymbol{w}|} + \min_{\{\boldsymbol{x_i}:y_i=-1\}} \tfrac{|\boldsymbol{w}\cdot\boldsymbol{x}+b|}{|\boldsymbol{w}|} \\
&= \tfrac{1}{|\boldsymbol{w}|} \left(\min_{\{\boldsymbol{x_i}:y_i=1\}} |\boldsymbol{w} \cdot \boldsymbol{x_i} + b| + \min_{\{\boldsymbol{x_i}:y_i=-1\}} |\boldsymbol{w} \cdot \boldsymbol{x_i} + b| \right) \\
&= \tfrac{2}{|\boldsymbol{w}|}
\end{aligned} \tag{14}$$

Chceme-li maximalizovat mezeru musíme maximalizovat výraz $^2/_{|\boldsymbol{w}|}$ nebo-li minimalizovat $|\boldsymbol{w}|$. Jak je z výše uvedených vztahů vidět, je potřeba pro nalezení optimální rozdělující nadroviny nalézt vázaný extrém, který je funkcí $\boldsymbol{w}, b$, pro již určené $\boldsymbol{x}$ jako support vectors. Použijeme minimalizačního kriteria Lagrangeových koeficientů pro nalezení optimální oddělující nadroviny:

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2} |\boldsymbol{w}|^2 - \sum_{i=1}^{l} \alpha_i y_i (\boldsymbol{x_i} \cdot \boldsymbol{w} + b) + \sum_{i=1}^{l} \alpha_i \tag{15}$$

Kde α_i jsou Lagrangeovy koeficienty (multiplikátory), pro které platí $\alpha_i > 0 \Leftrightarrow \boldsymbol{x_i}$ je „support vector“.

Teď si ještě naznačme řešení Lagrangeovy rovnice, předem je nutno podotknout, že se jedná o problém kvadratického programování. Hledáme minimum funkce $L(\boldsymbol{w}, b, \alpha)$ s ohledem ke koeficientům α_i. Tento problém bývá nazýván Wolfe dual (primárním problémem je označována rovnice (15) a sekundárním rovnice (16) viz. níže).

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \{ \min_{\boldsymbol{w},b} L(\boldsymbol{w}, b, \alpha) \} \tag{16}$$

Minimum $\min\limits_{\boldsymbol{w},b} L(\boldsymbol{w}, b, \alpha)$ určíme pomocí

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{17}$$

$$\frac{\partial L}{\partial \boldsymbol{w}} = 0 \quad \Rightarrow \quad \boldsymbol{w} = \sum_{i=1}^{l} \alpha_i y_i \boldsymbol{x_i} \tag{18}$$

Ze vztahů (16),(17) a (18) plyne

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i} \cdot \boldsymbol{x_j}) + \sum_{i=1}^{l} \alpha_i \tag{19}$$

Řešením rovnice (19) je

$$\bar{\alpha} = \arg\min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i} \cdot \boldsymbol{x_j}) - \sum_{i=1}^{l} \alpha_i \tag{20}$$

za podmínek (17) a $\alpha_i \geq 0, \quad i = 1, ..., l.$

4 Neseparabilní data

Pro neseparabilní data výše popsaný algoritmus nenajde vhodné řešení. Proto je třeba ho upravit tak, aby fungoval i pro neseparabilní data. Uvolníme proto omezení (7),(8) zavedením odchylky

$$\xi_i,\ i = 1 \dots l$$

Omezení definující oddělující nadrovinu potom přejdou v

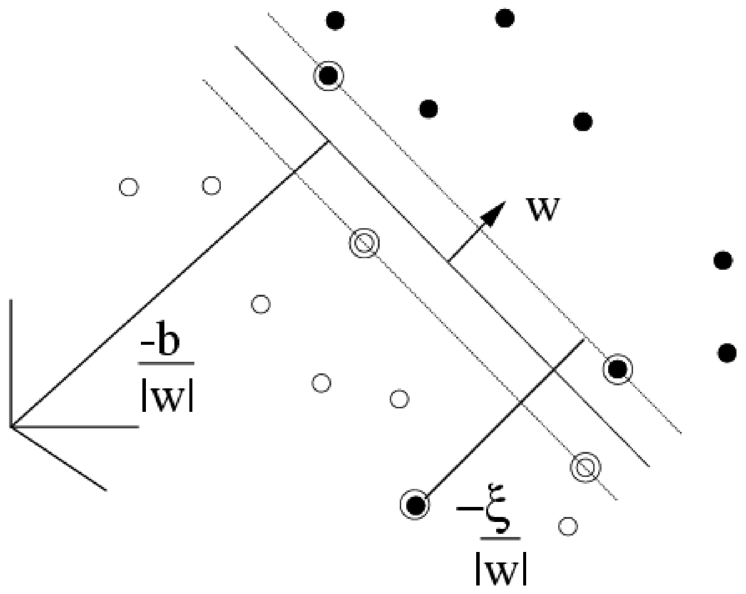
$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b \geq +1 - \xi_i \qquad \text{pro } y_i = +1 \tag{21}$$

$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b \leq -1 + \xi_i \qquad \text{pro } y_i = -1 \tag{22}$$

$$\xi_i \geq 0 \, \forall i. \tag{23}$$

Tím rozšíříme prostor příznaků a zároveň výraz $\sum\limits_i \xi_i$ je maximum chyby na trénovacích datech. Potom místo hledání minima $\frac{\|\boldsymbol{w}\|^2}{2}$ pro separabilní data budeme hledat minimum, kde parametr C volí uživatel a odpovídá pokutě za danou chybu.

$$\frac{\|\boldsymbol{w}\|^2}{2} + C \left(\sum_i \xi_i \right)^k \tag{24}$$



Obrázek 7: Lineární rozdělující nadroviny pro neseparabilní data

Vyřešení nerovnic (21) až (24) je konvexní minimalizační problém matematického programování. Pro k=1, k=2 přechází v problém kvadratického programování. Pro k=1 je výhodné, že ani ξ stejně jako Langrangeovy multiplikátory se neobjeví ve Wolfově duálním problému maximalizace

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j \boldsymbol{x_i} \cdot \boldsymbol{x_j} \tag{25}$$

$$0 \leq \alpha_i \leq C \tag{26}$$

$$\sum_i \alpha_i y_i = 0 \tag{27}$$

Řešením je:

$$\boldsymbol{w} = \sum_{i=1}^{N_s} \alpha_i y_i \boldsymbol{x_i}, \tag{28}$$

kde Ns je počet support vectors.

Další možná řešení lze najít pomocí:

- constrained confugate gradient descent

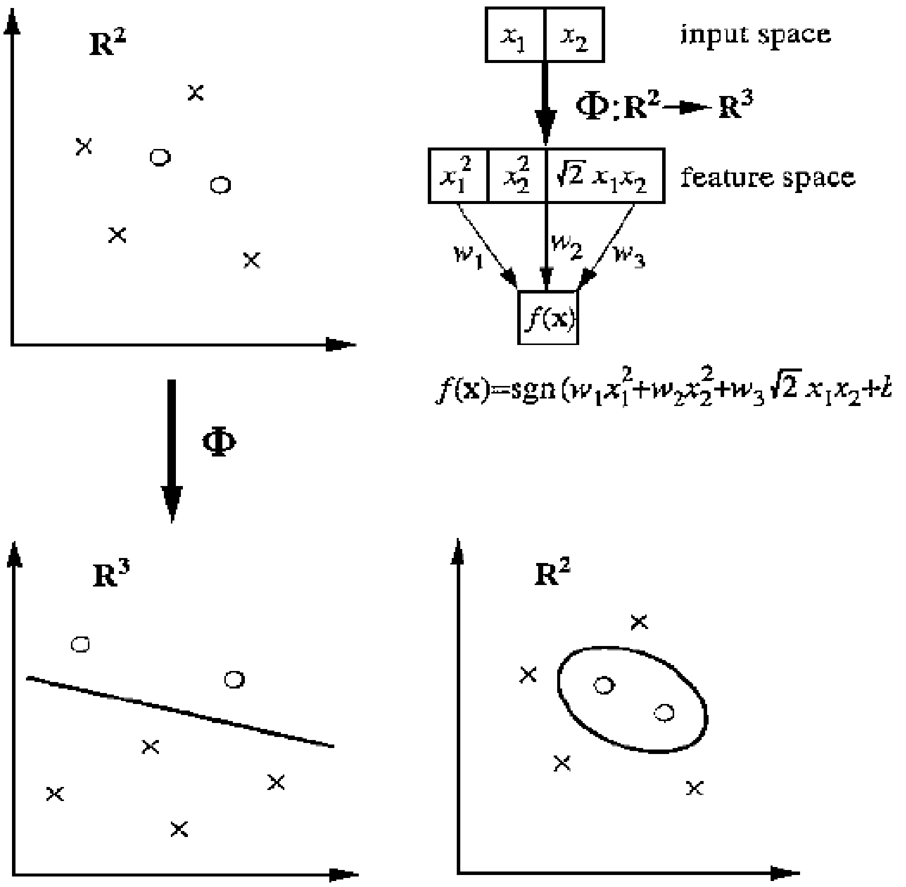
- interior point methods
- projection methods
- Bunhc Kaufman decomposition

5 Nelineární SVM

Nelineární SVM musíme použít v případě, když rozhodovací funkce není lineární. Boser, Guyon a Vapnik ukázali, že cesta vede přes vlastnosti skalárního součinu $\boldsymbol{x_i} \cdot \boldsymbol{y_i}$. Trénovací data jednoduše přetransformujeme z prostoru R^d do $\mathcal{H}$ pomocí zobrazení

$$\Phi : R^d \to \mathcal{H} \tag{29}$$

Řešení tedy vede na hledání v prostoru vyšší dimenze, kde jsou data separabilní a kde tedy můžeme použít postup z předchozího odstavce.



Obrázek 8: Transformace trénovacích dat (vlevo nahoře) nelineárně přes Φ do vícedimenziálního prostoru R^3 a konstrukce oddělující nadroviny (dole vlevo), SV klasifikátor (vpravo nahoře) odpovídající nelineární rozhodovací oblasti ve vstupním prostoru (vpravo dole).

Trénovací algoritmus závísí pouze na datech ze skalárního součinu v $\mathcal{H}$(tj. $\Phi(x_i) \cdot \Phi(x_j)$). Pokud tedy existuje Kernel funkce:

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \Phi(\boldsymbol{x_i}) \cdot \Phi(\boldsymbol{x_j}) \tag{30}$$

nemusíme funkci Φ počítat ani ji nemusíme znát. Ne vždy lze jednoduše určit, zda existuje pár $\{\mathcal{H},\Phi\}$. Odpověd lze určit pomocí Mercerovy podmínky:

existuje

$$K(\boldsymbol{x_i}, \boldsymbol{y_i}) = \sum_i \Phi(\boldsymbol{x})_i \Phi(\boldsymbol{y})_i \tag{31}$$

a jakákoliv funkce $g(x)$ taková, že integrál 32 je konečný.

$$\int g(\boldsymbol{x})^2 d\boldsymbol{x} \tag{32}$$

potom pár $\{\mathcal{H},\Phi\}$ existuje pouze když platí

$$\int K(\boldsymbol{x},\boldsymbol{y})g(\boldsymbol{x})^2)g(\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y} \geq 0 \tag{33}$$

Příklad: (ukázka povoleného kernelu, pro který můžeme určit Φ)
Trénovací data jsou v dimenzi R^2 a jako kernel vyberem $K(\boldsymbol{x_i},\boldsymbol{x_j}) = (\boldsymbol{x_i},\boldsymbol{x_j})^2$. Potom není těžké najít prostor $\mathcal{H} = \mathcal{R}^{\ni}$ a transformaci Φ z R^2 do $\mathcal{H}$ takovou, že

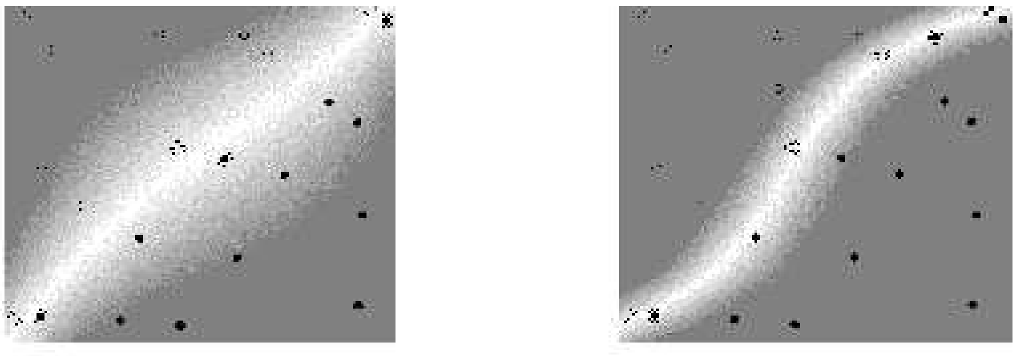
$(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}\cdot\boldsymbol{y})$ a

$$\Phi(\boldsymbol{x}) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}$$

Rozhodovací funkce klasifikátoru pro nelineární SVM se počítá dle

$$f(\boldsymbol{x}) = \sum_{i=1}^{N_s} \alpha_i y_i K(\boldsymbol{s_i},\boldsymbol{x_i}) + b, \tag{34}$$

kde $\boldsymbol{s_i}$ jsou support vectors.



Obrázek 9: Příklad nelineární SVM, kde kernel je kubický polynom 3. stupně

6 Generalizace SVM

- VC dimenze množiny funkcí rozdělujících nadrovin je

$$n + 1$$

- podmnožina orientovaných nadrovin

$$f(x, \omega, b) = sign(\omega \cdot x) + b$$

  definovaná na $\boldsymbol{X^r} = x_1, x_2, \ldots, x_r; |x_i - a| \leq R$ a splnující $\omega| \leq A$ má VC dimenzi omezenou nerovností

$$h \leq min(R^2 A^2)$$

  výše dva zmíněné případy lze shrnout a omezit VC dimenzi:

$$h \leq min(R^2 A^2, n) + 1 \tag{35}$$

- pro separabilní data (Vapnik, 1995) přináší druhou možnost jak odhadnout skutečnou chybu klasifikátoru SVM

$$E[P(erorr)] \leq \frac{E[\text{number of SUPPORT VECTORS}]}{\text{Number of training vectors}} \tag{36}$$

  kde $P(error)$ je hodnota skutečného rizika, klasifikátoru naučeného z $l - 1$ vzoru. $E[P(error)]$ očekavaní skutečného rizika přes všechny možnosti trénovací množiny $l - 1$. $E[Number\ of\ support\ vectors]$ je očekávání z počtu support vectors přes všechny možnosti trenovacích množin velikosti l.

7 Závěr

Metoda SVM má velmi jednouduchou geometrickou interpretaci. Učení SVM klasifikátoru vždy nalezne globalní minimum.

Reference

[1] CMP: *http://cmp.felk.cvut.cz/cmp/courses/recognition/resources/svmBurges.ps.gz*

[2] CMP: *http://cmp.felk.cvut.cz/cmp/courses/recognition/resources/svm_soton.ps.gz*

[3] CMP: *http://cmp.felk.cvut.cz/cmp/courses/recognition/resources/AIM_1599.ps*