Nebayesovské úlohy statistického rozpoznávání

Luděk Kopáček, Tomáš Vítek

28. května 2002

Obsah

1 Bayesovské rozhodování

1.1 Bayesovské rozhodování - opakování

Z důvodů rozdílného značení jednotlivých veličin si na začátek dovolím zopakovat základní pojmy z minulé přednášky. Bayesovské rozhodování vychází z minimalizace bayesovského rizika definovaného:

$R(d) = \sum_{x \in \mathcal{X}} \sum_{k \in \mathcal{K}} p(x,k) W(k, d(x))$

kde:
x **- pozorovatelné příznaky** ($x \in \mathcal{X}$ - množina možných pozorování)

k **- skryté parametry** ($k \in \mathcal{K}$ - množina hodnot skrytého parametru)

d **- rozhodnutí** ($d \in \mathcal{D}$ - množina rozhodnutí)

$p(x,k)$ - určuje pravděpodobnost, že objekt je ve stavu k a příznak nabývá hodnoty x. Vztah mezi pozorováním x a stavem k je tedy dán pravděpodobnostní funkcí.

$W(k, d(x))$ - pokutová funkce označuje ztrátu vyvolanou rozhodnutím d, pokud je objekt ve stavu k.

Riziko $R(d)$ je střední hodnota penalizační funkce, kterou platíme při našem rozhodování.

Výsledkem bayesovské strategie je funkce $d = d(x)$, který pozorování x přiřadí rozhodnutí d. Bayesovská strategie rozkládá pravděpodobnostní prostor na konvexní kužely.

1.2 Omezení bayesovského rozhodování

Bayesovský přístup přináší značnou obecnost přístupu ke klasifikaci a dává optimální klasifikátor. Ale musíme znát další informace upřesňující úlohu, které nejsou vždy k dispozici. Tato apriorní znalost o problému přináší úskalí Bayesovského přístupu. Rozdělme ji na tři okruhy:

- znalost ztrátové funkce $W(k, d(x))$: $K \times D \to R$
- znalost apriorní pravděpodobnosti $p(k)$
- znalost podmíněných pravděpodobností pozorování pro daný stav $p(x\,|k)$

Pozn.: Bayesovské rozhodování nelze použít, nemám-li některé z potřebných veličin (ztrátovou matici nebo některé z pravděpodobností).

Jednotlivé body rozebereme podrobněji v následujících kapitolách.

1.2.1 Bayesovské rozhodování pokutová funkce

Bayesovské rozhodování je minimalizací matematického očekávání rizika (pokuty). Vyžaduje tedy apriorní znalost pokutové funkce $W(k, d(x))$. Tato funkce předpokládá definici na upořádané množině hodnot, kde je definováno porovnávání, násobení a sčítání. Většinou je ztrátová funkce definována na množině reálných čísel. Tedy ve tvaru $W(k, d(x)) \in R$. Pokud na této množině nelze porovnávat, není možné sestavit pokutovou funkci.

Uveďme několik případů:

Př. 1. - lékařská diagnostika:
Mějme množinu měřitelných hodnot X - teplota, rengen, vyšetření krve, apod.

skryté stavy K - nemocný a zdravý
množina rozhodnutí D - léčit a neléčit

Řekněme, že chceme použít bayesovské strategie. Množina skrytých stavů a množina rozhodnutí dávají 4 kombinace. Stanovme nyní ztrátovou funkci pro tento případ.

W	zdravý	nemocný
neléčit	0	(?)
léčit	1000	0

Správná klasifikace je charakterizována v pokutové funkci. V případě, že je pacient zdravý a my léčíme, tak příjdeme například o cenu léku. Jaká hodnota má být ale nastavena v případě, že je pacient nemocný a my to nepoznáme? Tento případ může skončit také smrtí. Jak tedy ohodnotit tento stav?

Problém spočívá ve snaze o porovnání ceny léku a ceny lidského života.

Př. 2. - jaderná elektrárna:
Vezměme jadernou elektrárnu. Opět mějme množinu měřitelných hodnot X, což jsou ruzná měření a parametry.

skryté stavy K vše vpořádku nebo skrytá vada
množina rozhodnutí D nic nedělat nebo poslat kontrolu, případně provést opravu.

Pokud je elektrárna v pořádku, tak jsem udělal jen něco zbytečně, ale nic se nestalo. Přišel jsem o peníze na kontrolu. Ale pokud jsem nepoznal skrytou vadu a neprovedl opravu, tak elektrárna může až vybouchnout. Opět je zde problém, jak porovnat tyto dva stavy, kdy došlo ke španému rozhodnutí.

Př. 3. - odsouzení nevinného člověka:
Vezměme případ u soudu.

- Obžalovaný může být uznán viným v plném rozsahu obžaloby a může být odsouzen na 5 let vězení
- Obžalovaný může být odsouzen pouze k peněžitému trestu
- Obžalovaný může být osvobozen

Ve všech těchto případech byl výsledek v **jiných jednotkách**. Museli-bychom porovnávat pokutu a 5 let vězení.

Všechny tyto tři případy si vyžadovali porovnání hodnot, které byly neporovnatelné, či byli v jiných jednotkách. Tyto důvody znemožnily sestavení pokutové funkce a tím využití klasické Bayesovské strategie. Pro tyto úlohy je nutné formulovat jiné rozpoznávací úlohy (viz. dále).

1.2.2 Bayesovské rozhodování vyžaduje apriorní pravděpodobnosti $P(k)$

K Bayesovské startegii potřebujeme také apriorní pravděpodobnost skrytého stavu $P(k)$ pro všechny stavy $k \in K$. Ta ale nemusí být vždy dostupná. Existují dva důvody nedostupnosti apriorní pravděpodobnost.

- $P(k)$ existuje, ale my ji dosud neznáme. Řešením je buď úlohu pečlivěji prozkoumat a nalézt odhad pro apriorní pravděpodobnosti skrytého stavu, nebo formulovat úlohu jako nebayesovkou a vyhnout se tak použití $P(k)$ (viz. dále).
- Skrytý stav není náhodná veličina. Skrytý stav je deterministicky určen. Řešením je formulace úlohy takovým způsobem, aby nepotřebovala $P(k)$ (opět viz. dále).

Nemá smysl hledat pravděpodobnost nenáhodného (deterministického) jevu (tj. četnost výskytu nekonverguje k žádnému číslu).

Př. 4. - : letadla
Chceme sestrojit přístroj, který bude určovat, zdali letadlo, které se k nám blíží je přátelské nebo ne. Klasifikujeme tedy do dvou stavů :
k=1 ... vlastní letadlo
k=2 ... nepřátelské letadlo
x ... pozorování (příznaky)
Určení pravděpodobností $P(\overline{x}|1), P(\overline{x}|2)$ je sice složité, ale principielně zjistitelné

Pravděpodobnosti P(1), P(2) však vůbec neexistují. Zdali přiletí přátelské nebo nepřátelské letadlo není náhodný jev. Velmi to závisí na politické situaci. Pokud je mír, tak můžeme téměř s jistotou říci, že nepřiletí nepřátelské letadlo. Za válečné situace je již pravděpodobnost toho, že přiletí nepřátelské letadlo větší. Těžko lze charakterizovat výskyt letadla jedním číslem v intervalu $< 0, 1 >$

pseudořešení P(1)=P(2) není přípustné!
Proč?

Rovnost P(1)=P(2) by znamenala, že výskyt nepřátelského letadla je stejný jako výskyt přátelského letadla, což není pravda.

1.2.3 Podmíněné pravděpodobnosti podléhají nenáhodnému zásahu

Posledním připadem, kdy nelze použít Bayesovskou strategi, je případ, kdy neznáme podmíněné pravděpodobnosti pozorování za podmínky, že jsme v nějakém stavu $P(x|k)$ (potřebujeme ji pro určení pravděpodobnosti $P(x, k)$). Existují případy, kdy tato pravděpodobnost závisí na jevu, který není náhodný

Př. 5 - : Systém na rozpoznávání písma

Systém má rozpoznávat písmeno napsané uživatelem (OCR systém). K sytému budou přistupovat tři uživatelé Z={1,2,3}. Množina skrytých stavů a množina rozhodnutí je stejná a tvoří ji abeceda (jedná se tedy o klasifikaci). Je dána četnost výstytu jednotlivých písmen (tedy apriorní pravděpodobnost $P(k)$). Trénovací množinu připravili tři pisatelé ... Z={1,2,3}. Tato množina tvoří tzv. nenáhodný zásah.
Nyní nám zbývá určit podmíněnou pravděpodobnost $P(x|k, z)$ toho, že přistoupil-li uživatel z a napsal písmeno k, tak pozorujeme stav x.
Pokud by jev z byl náhodný zásah a byla známá pravděpodobnost $P(z)$ tohoto zásahu, pak by platilo:

$P(x|k) = \sum P(z) \cdot P(x|k, z)$
kde:
$P(x \mid k, z)$ ... pravděpodobnost, že naměříme x, když píše pisatelel z.

Toto řešení není v tomto případě možné, protože výskyt pisatele není náhodný jev. Četnost výskytu se mění například podle toho, zda je nekterý uživatel na dovolené, nebo nemocen.

V předchozích kapitolách byly uvedeny důvody, proč nelze ve všech případech použít klasickou Bayesovskou strategii, a proč se hledají nové formulace úloh rozpoznávání.

2 Úloha Neymana-Pearsona

Charakteristika řešeného problému: Mějme systém a v něm dva možné stavy (dvě třídy). Přiřadme těmto stavům označení K={1,2}, kde klasifikací k=1 je označen bezpečný stav a naopak klasifikace k=2 znamená, že se jedná o stav nebezpečí.
K rozhodnutí v jakém stavu se nacházíme máme množinu naměřených příznaků $\overline{x}$. Jedná se tedy o klasifikaci do dvou tříd při němž může dojít ke dvěma druhům chyby:

Falešný poplach: stav jsme klasifikovali jako nebezpečný ($k = 2$), ale ve skutečnosti se jedná o objekt ze třídy bezpečné ($k = 1$).

Přehlédnuté nebezpečí: stav jsme klasifikovali jako bezpečný ($k = 1$), ale ve skutečnosti se jedná o objekt ze třídy nebezpečné ($k = 2$).

Modelovým představitelem takové úlohy je problém spojený s odstávkami atomové elektrárny. Hrozí-li nebezpečí je nutné elektrárnu odstavit, ale jak poznat z naměřených příznaků jestli je odstávka nutná, nebo jestli jen zbytečně zastaví štěpnou reakci.

Požadavek na řešení : určení stavu (třídy) k objektu, pokud víme, že jsme pozorovali příznak $\overline{x}$

Předpoklady:

- Známe hodnoty pravděpodobnosti $p(\overline{x}|k)\ldots$ provedli jsme mnohokrát měřeni za obou stavů (bezpečného i nebezpečného) a z těchto měření jsme odvodili dané pravděpodobnosti.
  Jedná se o problém odhadování parametrů ze souboru dat. Odhadujeme s jakou pravděpodobností přijde např. v bezpečném stavu příznak x s určitou hodnotou. Je to vlastně odhad průběhu pravděpodobnostní funkce. (viz. kurz M4)

- Navíc víme, že v systému není žadná hodnota příznaku, která by přímo indikovala chybu, ale všechny hodnoty se vyskytují v obou stavech, jen s různými pravděpodobnostmi.
  Neexistuje tedy bezchybná strategie, což znamená, že každou hodnotu příznaku $\overline{x}$, jsme získali jak pro $k = 1$, tak i pro $k = 2$.

$(\exists\overline{x} : p(\overline{x}|1) > 0, p(\overline{x}|2) > 0)$

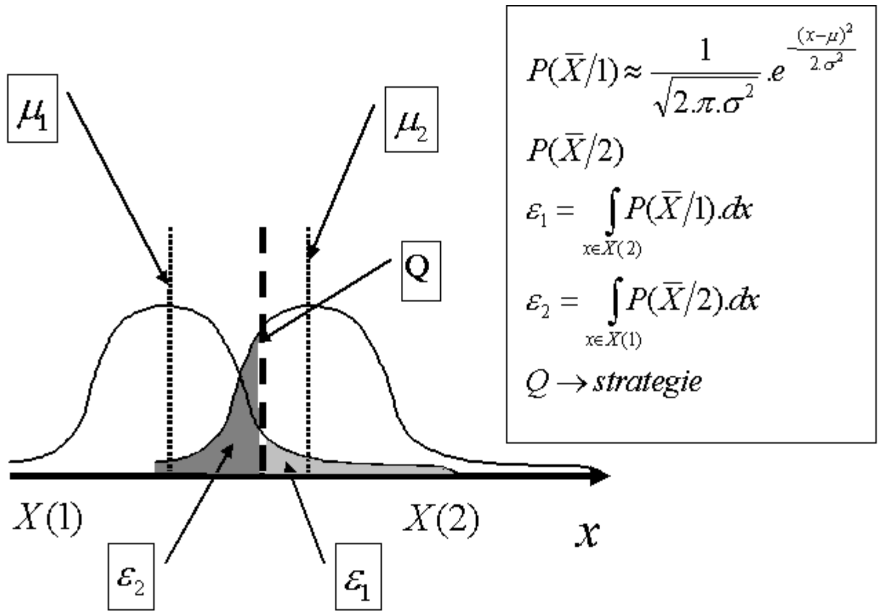
Strategie rozděluje příznakový prostor na dva podprostory oblasti pro první a druhou třídu. $X(1)...k = 1, X(2)...k = 2$
Pro určení takového rozdělení je nutné dát si předem nějaké hranice pro

nás únosného rizika a to zejména přehlédnutého nebezpečí. Toto riziko jde vyjádřit pomocí podmíněné pravděpodobnostní funkce:

$\varepsilon_1 = \sum_{x \in X(2)} p(x|1) \ldots$ falešný poplach

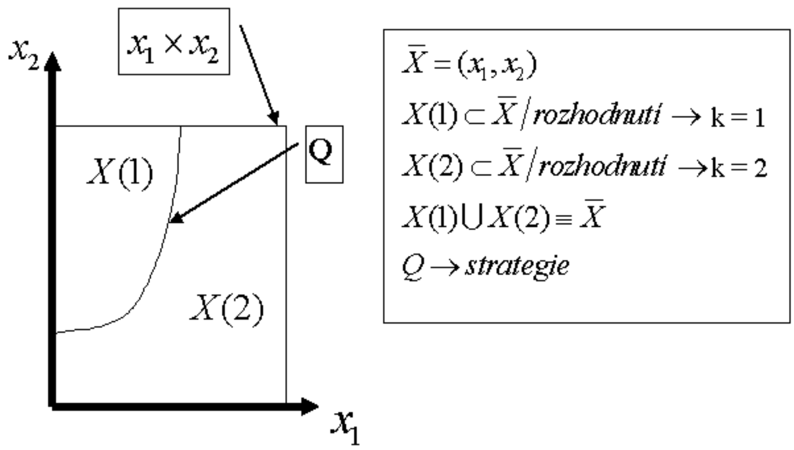
$\varepsilon_2 = \sum_{x \in X(1)} p(x|2) \ldots$ přehlédnuté nebezpečí

Následující obrázek ilustruje hodnoty ε_1 a ε_2 pro jednorozměrný spojitý případ.



Obrázek 1: Ilustrace rizik ve spojitém 1D případě

kde:
$X(1) \subset X \ldots$ množina $\overline{x}$, kde rozhoduji k=1
$X(2) \subset X \ldots$ množina $\overline{x}$, kde rozhoduji k=2
$X(1) \cup X(2) = X$,
$X(1) \cap X(2) = \emptyset$



Obrázek 2: Rozdělení příznakového prostoru

Celá strategie by měla být určena čísly ε_1 a ε_2. Chceme tedy najít takové rozdělení příznakového prostoru, že ε_2 (přehlédnutí) je méně pravděpodobné než nějaké malé, předem zvolené číslo a zároveň chceme minimalizovat počet případů kdy odstavíme zbytečně.

2.1 Definice úlohy Neymana-Pearsona

Tato rozhodovací strategie plní tedy dva požadavky:

(1) $\varepsilon_2 = \sum_{x \in X(1)} p(\overline{x}|2) \leq \varepsilon_0$ ... pravděpodobnost přehlédnutého nebezpečí je menší, než mnou zvolené ε_0

(2) minimalizuje pravděpodobnost falešného poplachu ε_1 za podmínky (1)

Řěšení:

Úloha lze prevést na problém lineárního programování (viz. kurz SDU), která minimalizuje funkci za omezení. Strategie vede na poměr věrohodností což chceme ukázat.

Zavedeme ztrátovou funkci:

$$\underline{r = \varepsilon_1 + \mu(\varepsilon_2 - \varepsilon_0)},$$

(μ je Lagrangůvmultiplikátor)

Za ε_1, ε_2 můžeme dosadit výrazy obsahující podmíněné pravděpodobnosti a následně využít faktu, že stavy 1,2 vyplňují celý příznakový prostor a pravděpodobnosti jdou převést přes pravděpodobnost doplňkového jevu.

$r = \sum_{x \in X(2)} P(x|1) + \mu(\sum_{x \in X(1)} P(x|2) - \varepsilon_0) =$

$= 1 - \sum_{x \in X(1)} P(\overline{x}|1) - \varepsilon_0\mu + \mu \sum_{x \in X(1)} P(\overline{x}|2) =$

$= 1 - \varepsilon_0\mu + \sum_{x \in X(1)} (\mu P(x|2) - P(x|1))$

$\mu P(\overline{x}|2) - P(\overline{x}|1) \lesseqgtr 0$

Protože chceme předchozí výraz minimalizovat, je nutné aby argument sumy mel co nejmenší hodnotu. To se dá vyjádřit požadavkem:

$$\frac{P(\overline{x}|1)}{P(\overline{x}|2)} \lesseqgtr \mu \qquad (1)$$

$\varepsilon_2 = \sum_{X(1)} P(\overline{x}|2) = \varepsilon_0$

Z této rovnice určíme μ.

2.2 Neyman-Personova úloha - poznámky a zobecnění

Ve formulaci úlohy se nevyskytují apriorní pravděpodovnosti, a proto s ní jdou vyřešit i další problémy popsane v kapitole 1. Příkladem je problem přátelské/nepřátelské letadlo.(Př.4)

Zobecnění:
Máme množinu tříd (stavů) K={1,2,3...n}, přičemž stavy $k = 2$ až $k = n$ považujeme za nebezpečné, a stav $k = 1$ za bezpečný. Chceme aby pravděpodobnost přehlédnutého nebezpečí všech těchto stavů byla menší než nějaké ε_0 a zaroveň minimalizujem pravděpodobnost falešného poplachu.

Požadavky:
(1) $\sum_{X(1)} P(\overline{x}|2) \leq \varepsilon_0$
$\sum_{X(2)} P(\overline{x}|3) \leq \varepsilon_0$
$\vdots$
(2) minimalizace $\sum_{X(2) \cup X(3) \cup \ldots} P(\overline{x}|1)$

Řešení: lze nalézt pomocí lineárního programování.

3 Minimaxní úloha

Strategie opět rozkladá příznakový prostor na disjunktní podmnožiny. K=$\{$1,2,3...$\}$ je množina stavů objektu (tříd).
$p(x|k)\ldots$ známo pro $k \in K$
$X(k) \subset X$, je-li $x \in X(k)$ rozhodujeme se pro třídu k.

Pro všechny třídy $k \in K$ můžeme spočítat hodnotu $\varepsilon(k)$, která je pravděpodobností toho, že byl-li objekt v daném stavu, tak my jsme klasifikovali do jiného (neboli pravděpodobnost špatné klasifikace).

Máme-li definovány hodnoty pro $\varepsilon(k)$, tak vybereme to, které je největší (reprezentuje třídu, pro kterou je největší chyba klasifikace).Cílem je najít takové rozdělení, aby prátě toto číslo bylo co nejmenší.

Definujme úlohu:

$$\max_{k \in K} \varepsilon(k) \tag{2}$$

kde $\varepsilon(k) = \sum_{x \notin X(k)} p(x|k)$

Cílem je nalézt rozklad X(k), který minimalizuje (2) za podmínky(3). Pro dvě třídy vede také tato strategie na výpočet poměru pravděpodobností a porovnání s nějakým prahem. Tento práh je určen jinak, je to jiná úloha, ale přesto je tvar rozhodování stejný.

Příklad:
Rozpoznávací algoritmus je hodnocen dvěma zkouškami - předběžnou a finální. Při finální zkoušce se experimentuje jen s objekty z nejtěžší třídy.
Při předběžné zkoušce jsou rovnoměrně vybírány objekty ze všech tríd. Celková chyba je pak vlastně průměrem všech chyb. Pokud však při finální zkoušce budou vybírány pouze objekty ze třídy, která má největší chybu, vzroste i celková chyba.
Zajistím-li, že $max\varepsilon(k)$ je nejmenší možná, a proto nemúže se mi stát, že vzroste chyba rozhodovámí, pokud budou přicházet pouze objekty z jedné třídy (té nejnevhodnější s největší chybou). Chyba je sice větsí než chyba Bayesovského klasifikátoru, ale při změně $p(\omega_i)$ zůstává konstantní.

3.1 Minimaxní úloha - řešení

- rozhodování podle poměru $\frac{P(\overline{x}|1)}{P(\overline{x}|2)}$
- pro případ dvou stavů vede tato strategie na $\varepsilon_1 = \varepsilon_2$, neboli pravděpodobnost toho, že objekt byl v 1. stavu a byla zařazen do druhého stavu je stejná, jako pravděpodobnost, že objekt byl ve 2. stavu a byl klasifikován do prvního stavu.
- důkaz $\longrightarrow$ lineární programování

Pozn.: na rozdíl od Neyman-Personovy úlohy zacházím s každou třídou stejně.

4 Chyba Bayesovské strategie při proměnné apriorní pravděpodobnosti tříd $P(k)$

Chceme navrhnout Bayesovský klasifikátor, ale nemáme apriorní pravděpodobnosti. Chceme proto určit apriorní pravděpodobnosti tak aby minimalizovali bayesovské riziko. Vše budeme demonstrovat na příkladu klasifikace do dvou tříd. Máme penalizační funkci W zavedenou jako:

d/k	1	1
1	0	1
2	1	0

Připomeňme si výraz pro bayesovké riziko pomozí ztrátové funkce W :

$$R(Q) = \sum_{\forall x} \sum_{\forall k} P(x,k).W(Q(x),k)$$

Rozepišme pro dva stavy:

$$R(Q) = \sum_{x\in(X1\cup X2)} \sum_{k\in\{1,2\}} P(x|1).W(Q(x),k)+$$
$$+\sum_{x\in(X1\cup X2)} \sum_{k\in\{1,2\}} P(x|2).W(Q(x),k)$$

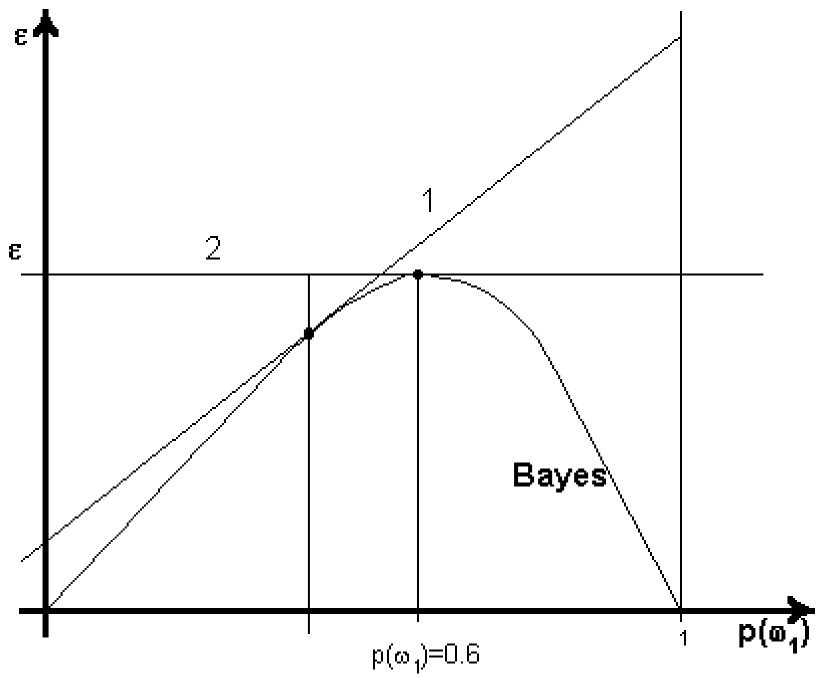
Přejdeme do vyjádření pomocí spojitého rozdělení a potom je R(Q) střední ztrátou ε:

$$\varepsilon = \int_{x\in X(2)} p(1,x)dx + \int_{x\in X(1)} p(2,x)dx =$$
$$= P(\omega_1)\int_{x\in D_2} P(x|\omega_1)dx + P(\omega_2)\int_{x\in D_1} P(x|\omega_2)dx$$

Pomocí integrálů podmíněných právděpodobností jsou však definována $\varepsilon_1, \varepsilon_2$:
$\varepsilon = P(\omega_1)\varepsilon_1 + P(\omega_2)\varepsilon_2 = P(\omega_1)\varepsilon_1 + (1 - P(\omega_1))\varepsilon_2 = \underline{\varepsilon_2 + P(\omega_1)(\varepsilon_1 - \varepsilon_2)}$

$D_i = \{x \in X,$kde se rozdodujeme pro $\omega_i\}$

Opět se snažíme, aby střední ztráta měla co nejmenší hodnotu a vím, že střední ztráta je přímo úměrná pravděpodobnosti jevu ω_1 $P(\omega_1)$. Bayesovské riziko, které optimalizujeme, je tedy funkce apriorní pravděpodobnosti $P(\omega_1)$. Situaci pro dvě třídy ilustruje následující obrázek.



Obrázek 3: Chyba Bayesovského klasifikátoru

Pokud bychom znali apriorní pravděpodobnosti $P(\omega_1)$, pak by chyba bayesovského klasifikátou v závislosti na $p(\omega_1)$ byla dána křivkou označenou jako Bayes (Pro $P(\omega_1) = 0$ je ten stav nemožný a vždy klasifikuju do druhé třídy a chyba je nulová. Podobně pro $P(\omega_1) = 1$, kdy klasifikuji vždy do první třídy.). Např. pro $P(\omega_1) = 0,6$ je chyba Bayesovského klasifikátoru největší.

Nemáme-li hodnoty $P(\omega_1)$, tak spočítáme hodnoty ε_1 a ε_2, ze kterých získáme chybu klasifikace. Pokud jsme se trefili, je chyba stejná jako Bayesovská. V opačném případě může být chyba klasifikace mnohem větší

než Bayesovská, protože se pohybujeme po přímce (viz přímka 1). Ochrana před nejhorším případem je tedy co jejvíce zmenšit hodnotu ε což platí při: $\varepsilon_1 = \varepsilon_2$. (viz přímka 2)
Výsledkem celého odvození je nejhorší chyba pro Bayese, ale přesto máme zaručeno, že pravděpodobnost chyby nebude větší než ε. (hledali jsme ohodnocení tak, že ε_1=ε_2)

5 Waldova úloha

Tato úloha je podobná úloze Neyman-Paersona, která hledala strategii takovou, která by zaručila, že ε_2 bude menší než stanovený práh. To znamená, že pravděpodobnost klasifikace do 1.třídy, byl-li objekt ve skutečnosti v 2.třidě je menší než stanovený práh. Mezi všemi těmito strategiemi jsme brali tu, která minimalizuje ε_1.

Liší se všajk v tom,že můžeme požadovat, aby pravděpodobnost falešného poplachu, stejně tak i pravděpodobnost přehlédnutého nebezpečí byla menší než stanovený práh. Této formulace vede na neřešitelnou úlohu, protože minimalizace ε_1 může jít proti minimalizaci ε_2.Pokud posouváme práh na jednu stranu, tak jedno ε klesá a druhé stoupá.

- asymetrie $\varepsilon_1, \varepsilon_2$
- nelze klást omezení na ε_1(falešný poplach), což může být nepřijatelné
- ovšem $\varepsilon_1 \leq \varepsilon$, $\varepsilon_2 \leq \varepsilon$ může být nesplnitelné

Formulace:
Waldova úloha neklasifikuje do dvou tříd, ale přidává třetí třídu X(0), která odpovídá stavu, kdy nemám dostatek informací pro rozhodnutí. Množinu příznaků X rozkládáme na podmnožiny X(1),X(2),X(0):

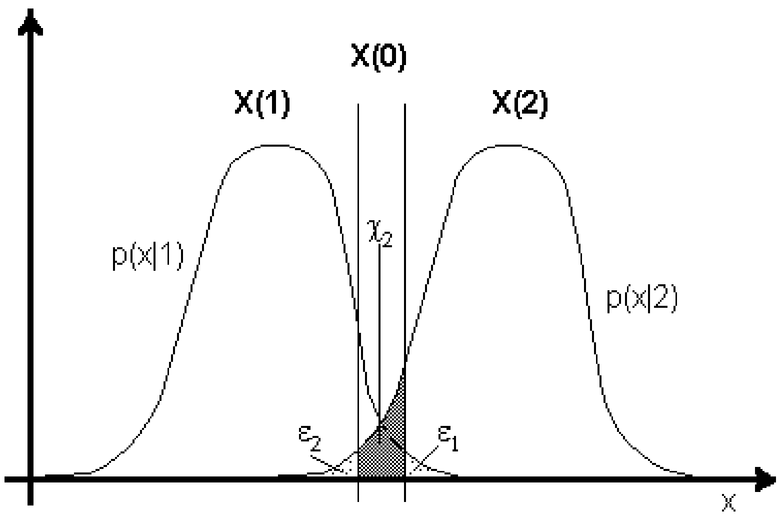
je-li pozorování (příznak) $x \in X(1)$, rozhodujeme se pro k=1
je-li pozorování (příznak)$x \in X(2)$, rozhodujeme se pro k=2
je-li pozorování (příznak)$x \in X(0)$, rozhodujeme se pro možnost "nevím".

Příklad : Letadla
Klasifikátor rozhoduje podle radaru, zda-li se jedná o přátelské nebo nepřátelské letadlo, ale pokud nemá dostatek infomací pro klasifikaci, tak se rozhodne pro třídu X(0) a je nutné použít jinou metody pro identifikaci letadla.

Př. 7 - : Lékařství
Klasifikátor posuzuje rentgenové snímky, podle kterých se má určit, zda-li pacient má rakovinu nebo ne. Dejme tmu, že má klasifikovat 100 snímků. Některé snímky jsou tak jasné, že nemá potíže s jejich klasifikací. Pak jsou tam další snímky, jejich zařazení je obtížnější. V tomto případě nechá rozhodnutí na expertovi, který může nařídit další upřesňující vyšetření, apod. Výsledkem je, že z velkého množství snímků posuzuje expert jen sporné případy.

Strategii charakterizují:
Definice ε_1 a ε_2 zůstávají stejné

$\varepsilon_1 = \sum_{x \in X(2)} p(x|1)$ ... chybné rozhodnutí o třídě (stavu) $k = 1$, neboli pravděpodobnost klasifikace do třídy 2, když objekt je 1. stavu.
$\varepsilon_2 = \sum_{x \in X(1)} p(x|2)$ ... chybné rozhodnutí o třídě (stavu) $k = 2$, neboli pravděpodobnost klasifikace do třídy 1 když objekt je 2. stavu.
Dále je úloha charakterizována dvěma čísly:
$\chi_1 = \sum_{x \in X(0)} p(x|1)$ ... případné nerozhodnuté situace pro k=1 - pravděpodobnost klasifikace do X(0) (nevim), když objekt byl ve stavu k=1
$\chi_2 = \sum_{x \in X(0)} p(x|2)$ ... případné nerozhodnuté situace pro k=2 - pravděpodobnost klasifikace do X(0) (nevim), když objekt byl ve stavu k=2



Obrázek 4: Ilustrace rizik pro spojitý 1D případ

Optimální strategie:

Hledáme strategii, pro kterou platí:
(1) $\varepsilon_1 \leq \varepsilon$, $\varepsilon_2 \leq \varepsilon$ ... kladu omezující požadavky na chybu rozhodnutí pro obě třídy (jak bezpečnou, tak nebezpečnou).
(2) minimalizuji $max(\chi(1), \chi(2))$, aby klasifikátor stále jen neodpovídal "nevím", což by vyhovovalo omezujícím podmínkám z bodu (1).

Řešení:
Strategie se opírá o věrohodnostní poměr $\gamma(x) = \frac{p(x|1)}{p(x|2)}$, který je porovnán s dvěma prahy$\Theta_1 a \Theta_2$, které rozdělí množinu příznaků na tři podmnožiny.
prahy: $\Theta_1, \Theta_2, (\Theta_1 \leq \Theta_2)$

1. $\gamma(x) < \Theta_1$
2. $\gamma(x) \in (\Theta_1, \Theta_2)$
3. $\gamma(x) > \Theta_2$

Důkaz: pro obecný případ n tříd → minimalizační problém řešitelný lineárním programováním.

6 Rozhodování při náhodných zásazích (Linnikova úloha)

Problém je popsán v kapitole 1.4 jako problém rozpoznávání při závislosti na nenáhodnem parametru (Příklad 5)

$X, K, Z; p(x|K, Z)$ ... známé
$z \in Z$ ... množina všech zásahů (nenáhodných)
$X(k), k \in K$ ... rozklad X, který určuje třídu

Rozhodování závisí navíc na tom zda předem známe apriorní pravděpodobnosti jednolivých nenáhodných parametrů.

6.1 P(k) známe

Pokud známe apriorní pravděpodobnost k můžeme spočítat střední hodnotu chybn0 klasifikace při daném nenáhodném zásahu z:

$$\varepsilon(z) = \sum_{k\in K} p(k) \sum_{x\in X(k)} p(x|k,z),$$

kde $\sum_{x\in X(k)} p(x|k,z)$ je pravděpodobnost toho, že při zásahu z_1 se vyskytl jev k, bylo klasifikována jiné d.
$p(k)$ charakterizuje pravděpodobnost výskytu jednotlivých k v textu příslušného jazyka.

Strategii hodnotíme podle chyby při nejhorším zásahu:

$$\max_{z\in Z} \varepsilon(z)$$

Chceme najít takovou strategii, která minimalizuje $\max_{z\in Z} \varepsilon(z)$:

$(X^*(k), k \in K) = \arg\min_{X(k)} \max_{z\in Z} \sum_{k\in K} p(k) \sum_{x\notin X(k)} p(x|k,z)$
Pozn.: lze převést na úlohu lineárního programování.

6.2 P(k) neznáme

Neznáme apriorní pravděpodobnosti $p(k)$, a proto je ε závislé na dvou parametrech a to nenáhodném zásahu z a skrytém stavu k:

$$\varepsilon(k,z) = \sum_{x\notin X(x)} p(x\,|k,z)\,,$$

což odpovídá tomu, že přistoupil-li uživatel z_2 a jev k se vyskytl, bylo přesto k špatně klasifikováno. Mezi těmito čísli se vybere to, které odpovídá největší chybě.(V příkladu 5 uživatel 2 píše strašné písmeno "a", pak $\varepsilon^* = \varepsilon(''a'', z_2)$)

Hodnocení strategie opět probíhá podle nejhoršího případu:

$$\max_{k\in K} \max_{z\in Z} \varepsilon(k,z)$$

Řešením je rozklad:

$(X^*(k), k \in K) = \arg\min_{X(k)} \max_{k\in K} \max_{z\in Z} \sum_{x\notin X(k)} p(x,k,z)$
Řešení:lze převézt na úlohu lineárního programování

7 Shrnutí

1) Baysovké rozhodování není jediná cesta: nemohu minimalizovat riziko neznám-li všechny pravděpodobnosti potřebné pro řešení. Pak musím úlohu definovat jinak.
2) Strategie v bayesovském i nebayesovském rozhodování dělí prostor pravděpodobností na konvexní kužely - opírají výpočet věrohodnostního poměru (strategie jsou si hodně podobné a liší se pouze v určení rozhodovacích prahů).
3) Rozhodovací strategie pro nebayesovské úlohy lze hledat pomocí lineárního programování.

8 Použitá literatura

http://cmp.felk.cvut.cz/cmp/courses/recognition/resources/PR-uni-arizona /lecture2.pdf
Schesinger, M., Hlaváč, V., Deset přednášek z teorie statistického a strukturního rozpoznávání, ČVUT Praha 1999