# Nebayesovské úlohy statistického rozpoznávání

Martin Augustýn

8. června 2001

## Obsah

# 1 Bayesovské rozhodování (BR) - omezení

Omezení vyplývají přímo z definic klíčových pojmů:

- musím znát hodnou ztrátové funkce $\lambda(c_i, c_j) \in R$
- musím znát jednotlivé pravděpodobnosti p(c), p(x,c)

Pozn.: Bayesovskou úlohu nelze modelovat, nemám-li některé z pořebných veličin (ztrátovou matici nebo některé z pravděpodobností).

## 1.1 Bayesovské royhodování je minimalizací matematického očekávání rizika

Vyžaduje znalost hodnot rizika (pokuty) $\lambda(c_i, c_j) \in R$
Tyto hodnoty lze sčítat a násobit.

**Problém:**

- lékařská dianostika (smrt x cena léku)
- jaderná elektrárna (výbuch)
- soudní rozhodnutí (odsouzení nevinného)
- ztráta je v různých jednotkách

U podobných událostí je velmi obtížné přiřadit konkrétní události reálné číslo určující ztrátu mého rozhodnutí. Někdy dokonce není možné ztrátu vůbec určit.

## 1.2 Bayesovské rozhodování vyžaduje apriorní pravděpodobnosti $P(c_i)$

Nemá smysl hledat pravděpodobnost náhodného jevu (tj. četnost výskztu nekonverguje k žádnému číslu).

Příklad: letadla
c=1 ... vlastní letadlo
c=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.

**pseudořešení P(1)=P(2) není přípustné!**

## 1.3 Podmíněné pravděpodobnosti podléhají nenáhodnému zásahu

Příklad: Systém na rozpoznávání písma
Trénovací množinu připravili tři pisatelé ... Z={1,2,3}
$P(x \mid c, Z)$ ... pravděpodobnost, že naměříme x a jedná se o znak c napsaný pisatelem Z
Testování: jeden pisatel

následující řešení není možné:
$P(x|c) = \sum P(Z) \cdot P(x|c, Z)$

Z je totiž jeden konkrétní pisatel a ne jakýsi průměrem všech pisatelů.

## 2 Úloha Neymana-Pearsona

**Úvod:** předpokládáme, že máme dva druhy objektů (dvě třídy) C={1,2}. Přičmž klasifikace c=1 znamená, že se jedná o bezpečný objekt. Naopak klasifikace c=2 znamená, že se jedná o objekt nebezpečný. Dále jsme naměřili nmožinu příznaků $\overline{x}$. Při následné klasifikaci objektu může dojít ke dvěma druhům chyby:
Falešný poplach: objekt jsme klasifikovali jako nebezpečný ($c = 2$), ale ve skutečnosti se jedná o objekt ze třídy bezpečné ($c = 1$)
Přehlédnuté nebezpečí: objekt jsme klasifikovali jako bezpečný ($c = 1$), ale ve skutečnosti se jedná o objekt ze třídy nebezpečné ($c = 2$)

Pozn.: s každou třídou (bezpečná x nebezpečná) zacházím jinak. viz. 2.1 - Neyman-Personova strategie.

**Cíl:** určení třídy (stavu) c objektu, pokud víme, že jsme pozorovali příznak $\overline{x}$

**Předpoklady:**

- známe jednotlivé pravděpodobnosti $p(\overline{x}|c)\ldots$ provedli jsme mnohokrát měřeni a z něho jsme odvodili dané pravděpodobnosti.
- neexistuje bezchybná strategie, což znamená, že existuje takové měření x, které jsme získali jak pro $c = 1$, tak pro $c = 2$

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

**Strategie:** je charakterizována čísly
$\varepsilon_1 = \sum_{\overline{x}\in X(2)} p(x|1)$ ...falešný poplach
$\varepsilon_2 = \sum_{\overline{x}\in X(1)} p(x|2)$ ...přehlédnuté nebezpečí
kde:
$X(1) \subset X$ ...množina $\overline{x}$, kde rozhoduji c=1
$X(2) \subset X$ ...množina $\overline{x}$, kde rozhoduji c=2
$X(1) \cup X(2) = x$,
$X(1) \cap X(2) = 0$

## 2.1 Neyman-Personova strategie

Tato rozhodovací strategie se zkládá ze dvou kroků:
(1) $\varepsilon_2 = \sum_{x \in X(1)} p(\overline{x}|2) \leq \varepsilon_0$ ... pravděpodobnost přehlédnutého nebezpečí má být menší, než mnou zvolené $\varepsilon_0$
(2) minimalizuji pravděpodobnost falešného poplachu $\varepsilon_1$

**Řěšení:**
Ztrátová funkce $\underline{r = \varepsilon_1 + \mu(\varepsilon_2 - \varepsilon_0)}$,
kde $\mu$ je Lagrangův multiplikátor

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

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

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

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

Z této rovnice určíme $\mu$.

## 2.2 Neyman-Personova úloha - pokračování

Pozn.: Ve formulaci úlohy se nevyskytují apriorní pravděpodovnosti

Příklad: přátelské/nepřátelské letadlo

**Zobecnění:**
Máme množinu tříd (stavů) C={1,2,3...}, přičemž stavy $c = 1$ a $c = 2$ považujeme za nebezpečné.

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

**Řešení:** lye nalézt pomocí lineárního programování

# 3 Minimaxní úloha

C={1,2,3...} je množina stavů objektů (tříd)
$p(x|c)\ldots$ známo pro $c \in C$
$X(c) \subset X$, je-li $x \in X(c)$ rozhodujeme se pro třídu c.

**Ztrátová (optimalizační) funkce:**

$$\max_{c \in C} \varepsilon(c) \tag{2}$$

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

Rovnice (1) udává chybu klasifikátoru v případě, že $P(c_x) = 1$
Minimaxní úloha hledá rozklad X(c), který minimalizuje (1).

**Příklad:** Rozpoznávací algoritmus je hodnocen předběžnou a finální zkouškou. 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 $\varepsilon(1) = \varepsilon(2) = \varepsilon(3) = \ldots = \varepsilon(n)$, nemůže se mi stát, že vzroste chyba rozhodovámí, pokud budou přicházet pouze objekty z jedné třídy (té nejnevhodnější s největší chybou). Chyba je sice větší než chyba Bayesovského klasifikátoru, ale při změně $p(\omega_i)$ zůstává konstantní.

## 3.1 Minimaxní úloha - řešení

- rozhodování podle poměru $\frac{P(\overline{x}|1)}{P(\overline{x}|2)}$
- $\varepsilon_1 = \varepsilon_2$
- 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 $P(c_i)$

střední ztráta $\varepsilon$ (viz 2.přednáška)

$\varepsilon = \int_{x \in D_2} p(\omega_1, x)dx + \int_{x \in D_1} p(\omega_2, x)dx = P(\omega_1)\int_{x \in D_2} P(x|\omega_1)dx + P(\omega_2)\int_{x \in D_1} P(x|\omega_2)dx =$
$= P(\omega_1)\varepsilon_1 + P(\omega_2)\varepsilon_2 = P(\omega_1)\varepsilon_1 + (1 - P(\omega_1))\varepsilon_2 = \underline{\varepsilon_2 + P(\omega_1)(\varepsilon_1 - \varepsilon_2)}$

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

$\Rightarrow$ střední ztráta je přímo úměrná pravděpodobnosti, že nastane jev $\omega_1$ $P(\omega_1)$

Měním-li tedy pravděpodobnost $p(\omega_1)$, mění se po přímce i chyba Bayesovského klasifikátoru. Ochrana před nejhorším případem: $\varepsilon_1 = \varepsilon_2$, rozhodnutí podle $\frac{P(x|\omega_1)}{P(x|\omega_2)}$

# 5 Waldova úloha

Vyjděme z úllohy Neyman-Paersona

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

**Formulace:**
Množinu příynaků X rozkládáme na podmnožiny X(1),X(2),X(0)
je-li pozorování (příznak) $x \in X(1)$, rozhodujeme se pro c=1
je-li pozorování (příznak)$x \in X(2)$, rozhodujeme se pro c=2
je-li pozorování (příznak)$x \in X(0)$, rozhodujeme se, že pozorování x neposkytuje dostatek informace (možnost "nevím").

**Strategii charakterizují:**
$\varepsilon_1 = \sum_{x \in X(2)} p(x|1)$ ... případné chybné rozhodnutí o třídě (stavu) $c = 1$
$\varepsilon_2 = \sum_{x \in X(1)} p(x|2)$ ... případné chybné rozhodnutí o třídě (stavu) $c = 2$
$\chi_1 = \sum_{x \in X(0)} p(x|1)$ ... případné nerozhodnuté situace pro c=1
$\chi_2 = \sum p(x|2)$ ... případné nerozhodnuté situace pro c=2

**optimální strategie:**
(1) $\varepsilon_1 \leq \varepsilon$, $\varepsilon_2 \leq \varepsilon$ ... kladu omeující 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:**
opírá se o věrohodnostní poměr $\gamma(x) = \frac{p(x|1)}{p(x|2)}$

**prahy:** $\Theta_1, \Theta_1 \leq \Theta_2, \Theta_2$

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

3. $\gamma(x) > \Theta_2$

**Důkaz:**i pro obecný případ n tříd → minmalizační problém řešitelný lineárním programováním.

# 6 Rozhodování při náhodných zásazích (Linnikova úloha)

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

## 6.1 P(c) známe

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

**Hodnocení strategie:**
$\varepsilon^* = \max_{z \in Z} \varepsilon(z)$

Řešením je rozklad:
$(X^*(c), c \in C) = \arg\min_{X(c)} \max_{z \in Z} \sum_{c \in C} p(c) \sum_{x \notin X(c)} p(x|k, z)$

## 6.2 P(c) neznáme

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

**Hodnocení strategie:**
$\varepsilon^* = \max_{c \in C} \max_{z \in Z} \varepsilon(k, z)$

Řešením je rozklad:
$(X^*(c), c \in C) = \arg\min_{X(c)} \max_{k \in K} \max_{z \in Z} \sum_{x \notin X(c)} p(x, k, z)$

**Řešení:** pomocí úlohy lineárního programování

# 7 Použitá literatura

Jako podklady pro vypracování této přednášky jsem použil slidy z přednášky ing. Jiího Matase, dále pak své zápisky z této přednášky a jako poslední zdroj mi sloužil následující materiál v elektronické podobě:
http://cmp.felk.cvut.cz/cmp/courses/recognition/resources/PR-uni-arizona/lecture2.pdf