1 Aplikace

.

- Rozpoznávání řeči
- Identifikace mluvčího
- Optické rozpoznáváni znaků (OCR)
- Rozpoznávání otisků prstů
- Rozpoznávání oblyčejů
- Automatické třídění dřeva
- Detekce falešných bankovek
- Automatická analýza obrazů z družice
- Automatické prohledávání WWW stránek

2 Rozhodování podle kriteria minimální střední ztráty (Bayesovské rozhodování)

[?], strany 107–116
[?], kapitola 2.4, strany 21–28

Je dáno:

$\mathcal{X}$	příznakový prostor, $\mathbf{x} \in \mathcal{X}$
Ω	množina indikátorů tříd $\Omega = \{\omega_r\}, r = 1, \ldots, R$
$P(x, \omega)$	sdružená pravděpodobnost $\mathcal{X} \times \Omega \to \mathcal{R}$
$\mathcal{D}$	množina všech rozhodovacích pravidel $\mathcal{D} = \{d(\mathbf{x})\}$

$$\begin{bmatrix} \lambda(\omega_1, \omega_1) & \lambda(\omega_1, \omega_2) & \ldots & \lambda(\omega_1, \omega_R) \\ \vdots & & & \\ \lambda(\omega_R, \omega_1) & \lambda(\omega_R, \omega_2) & \ldots & \lambda(\omega_R, \omega_R) \end{bmatrix} \text{ matice ztrátových funkcí}$$

Je známa i apriorní pravděpodobnost tříd $P(\omega_r)$ a podmíněná pravděpodobnost $p(x|\omega_r)$ pro vsechny třídy, neboť je dána pravděpodobnost $p(x, \omega)$ a platí:

$$P(\omega_r) = \int_{x \in \mathcal{X}} p(\mathbf{x}, \omega_r) dx$$

$$p(\mathbf{x}|\omega) = \frac{p(\mathbf{x}, \omega_r)}{P(\omega_r)}$$

Střední ztrátu definujeme:

$$R(d) \quad = \sum_{s=1}^{R} \int_{\mathcal{X}} \lambda(d(\mathbf{x})|\omega_s) p(\mathbf{x}|\omega_s) P(\omega_s) d\mathbf{x} \tag{1}$$

Optimální rozhodovací pravidlo $d^*(\mathbf{x})$ minimalizuje střední ztrátu (riziko) $R(d)$

$$R(d^*) = \min_{d \in \mathcal{D}} R(d) \tag{2}$$

$$
\begin{aligned}
\min_{d\in\mathcal{D}} R(d) &= \min_{d\in\mathcal{D}} \int_{\mathcal{X}} \sum_{s=1}^{R} \lambda(d(\mathbf{x})|\omega_s) p(\mathbf{x}|\omega_s) P(\omega_s) d\mathbf{x} \\
&= \int_{\mathcal{X}} \min_{\mathbf{r}} \sum_{s=1}^{R} \lambda(\omega_r|\omega_s) p(\mathbf{x}|\omega_s) P(\omega_s) d\mathbf{x} \\
&= \int_{\mathcal{X}} \min_{\mathbf{r}} L_{\mathbf{x}}(\omega_r) d\mathbf{x}
\end{aligned} \tag{3}
$$

Funkce

$$L_{\mathbf{x}}(\omega_r) = \sum_{s=1}^{R} \lambda(\omega_r|\omega_s) p(\mathbf{x}|\omega_s) P(\omega_s)$$

se nazývá *ztráta obrazu* $\mathbf{x}$ *při klasifikaci do třídy* ω_r.

Optimální rozhodovací pravidlo $d^*(\mathbf{x})$ je takové, aby pro každé $\mathbf{x}$ platilo:

$$L_{\mathbf{x}}(d^*(\mathbf{x})) = \min_{r} L_{\mathbf{x}}(\omega_r) \tag{4}$$

Aniž bychom ovlivnili rozhodnutí, lze uvažovat s využitím Bayesova vztahu

$$P(\omega_s|\mathbf{x}) = p(\mathbf{x}|\omega_s) P(\omega_s) / p(\mathbf{x})$$

a minimalizovat

$$L'_{\mathbf{x}}(\omega_r) = L_{\mathbf{x}}(\omega_r)/p(\mathbf{x}) = \boxed{\sum_{s=1}^{R} \lambda(\omega_r|\omega_s) P(\omega_s|\mathbf{x})} \tag{5}$$

2.1 Rozhodování podle kritéria minimální chyby

Je speciálním případem Bayesovského rozhodovánᡠje-li matice ztrátových funkci ve tvaru

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ \vdots & & & \\ 1 & 1 & \dots & 0 \end{bmatrix}$$

Po dosazení do rovnice (5):

$$\begin{aligned}
L'_{\mathbf{x}}(\omega_r) &= \sum_{s=1}^{R} \lambda(\omega_r|\omega_s)P(\omega_s|\mathbf{x}) \\
&= \sum_{s=1}^{R} P(\omega_s|\mathbf{x}) - P(\omega_r|\mathbf{x}) \\
&= 1 - P(\omega_r|\mathbf{x})
\end{aligned} \tag{6}$$

Optimální je volba třídy

$$\begin{aligned}
\omega_s &= \arg\min_r L_{\mathbf{x}}(\omega_r) \\
&= \arg\min_r 1 - P(\omega_r|\mathbf{x}) \\
&= \boxed{\arg\max_r P(\omega_r|\mathbf{x})}
\end{aligned} \tag{7}$$

která má *největší aposteriorní pravděpodobnost*.

Klasifikaci do dvou tříd (dichotomie). Rozhodnutí lze učinit dle znamenka funkce

$$g(\mathbf{x}) = P(\omega_1|\mathbf{x}) - P(\omega_2|\mathbf{x})$$

$\rightarrow$ ADALINA

2.2 Diskriminační funkce

Rozhodujeme-li se pro ω_r takove, že

$$L_{\mathbf{x}}(\omega_r) \leq L_{\mathbf{x}}(\omega_s), s = 1, \ldots, R$$

pak transformaci

$$g(x) = F(L_{\mathbf{x}}(\omega_r))$$

zůstane nerovnost zachována, je-li $F(x)$ monotonní ($x \leq y \Rightarrow F(x) \leq F(q)$).

2.3 Bayesovské rozhodování, $p(\mathbf{x}|\omega_r)$ normální

$$p(\mathbf{x}|\omega_r) = N(\mu_r, \sigma_r) = \frac{1}{(2\pi)^{n/2}\sqrt{|\sigma_r|}} e^{\frac{1}{2}(\mathbf{x}-\mu_r)^T \sigma_r^{-1}(\mathbf{x}-\mu_r)}, r = 1, \ldots, R \qquad (8)$$

kde $|\sigma_r|$ je determinanat disperzní matice σ_r.
Uvažujeme jednotkovou matici ztrát a apriorní pravděpodobnosti tříd $P(\omega_r)$.
Pro zjednodušení výpočtu budeme uvažovat diskriminační funkci $g(\mathbf{x})$, která je logaritmem $L_{\mathbf{x}}(\omega_r)$:

$$\begin{aligned} g(\mathbf{x}) &= \ln(p(\mathbf{x}|\omega_r)) + \ln(P(\omega_r)) \\ &= \tfrac{1}{2}(\mathbf{x}-\mu_r)^T \sigma_r^{-1}(\mathbf{x}-\mu_r) - \tfrac{1}{2}\ln(|\sigma_r|) + \ln(P(\omega_r)) \qquad (9) \end{aligned}$$

Obecně jsou pro $N(\mu_r, \sigma_r)$ diskriminační funkce *kvadratické*. Mají-li navíc třídy shodné disperzní matice, pak lze po rozepsání pravé strany rovnice (9) vypustit členy $\frac{1}{2}\ln(|\sigma_r|)$, $\frac{1}{2}\mathbf{x}^T\sigma_r^{-1}\mathbf{x}$. Diskriminační funkce

$$g(\mathbf{x}) = \mu_r^T \sigma^{-1}\mathbf{x} + \tfrac{1}{2}\mu_r^T\sigma^{-1}\mu_r + \ln P(\omega_r) \qquad (10)$$

je nyní *lineární*.

→ lineární klasifikátor
→ prahový logický obvod, model neuronu

Navíc je-li $\sigma = k\mathbf{E}$ a třídy jsou stejně pravděpodobné, pak diskriminační funkce maji tvar

$$\omega g(\mathbf{x}) = \mu_r^T\mathbf{x} - \frac{1}{2}\mu_r^T\mu_r \qquad (11)$$

a představují klasifikace dle minima vzdálenosti od vektoru středních hodnot μ_r
→ klasifikace podle minima vzdálenosti

2.4 Bayesovské rozhodování s možností 'nevím' (reject option)

$\Omega = \{\omega_r\}, r = 1, \dots, R$	původní množina indikátorů tříd
$\Omega' = \Omega \cup \#$	Ω ozšířená o třídu 'nevím' označenou #

Ztrátovou funkci definujeme:

$$\lambda(\omega_r, \omega_s) = \begin{cases} 0 & : \quad \omega_s = \omega_r \\ 1 & : \quad \omega_s \neq \omega_r \wedge \omega_s \neq \# \\ \epsilon & : \quad \omega_s = \# \end{cases} \tag{12}$$

Optimální rozhodnutí:

$$\begin{aligned} \omega'_* &= \arg\min_{\omega_r \in \Omega'} \sum_{s=1}^{R} \lambda(\omega_r|\omega_s) P(\omega_s|\mathbf{x}) \\ &= \arg\min_{\omega_r \in \Omega'} L'_{\mathbf{x}}(\omega_r) \\ &= \begin{cases} \arg\min_{\omega_r \in \Omega} L'_{\mathbf{x}}(\omega_r) & : \quad \exists L_x(\omega) \leq L_x(\#) \\ \# & : \quad L_x(\#) < L_x(\omega), \forall \omega \in \Omega \end{cases} \end{aligned} \tag{13}$$

kde

$$L'_{\mathbf{x}}(\omega_r) = \sum_{s=1}^{R} \lambda(\omega_r|\omega_s) P(\omega_s|\mathbf{x}) \tag{14}$$

platí, že

$$\begin{aligned} \omega_* &= \arg\min_{\omega_r \in \Omega} L'_{\mathbf{x}}(\omega_r) \\ &= \arg\min_{\omega_r \in \Omega} \sum_{s=1}^{R} P(\omega_s|\mathbf{x}) \\ &= \arg\max_{\omega_r \in \Omega} P(\omega_r|\mathbf{x}) \end{aligned} \tag{15}$$

pro $L_x(\#)$ platí

$$\begin{aligned} L_x(\#) &= \sum_{s=1}^{R} \lambda(\#|\omega_s) P(\omega_s|\mathbf{x}) \\ &= \sum_{s=1}^{R} \epsilon P(\omega_s|\mathbf{x}) \\ &= \epsilon \end{aligned} \tag{16}$$

rozhodujeme se tedy pro třídu

$$\omega'_* = \begin{cases} \arg\max_{\omega_r \in \Omega} P(\omega_r|\mathbf{x}) & : \quad 1 - \max_{\omega_r \in \Omega} P(\omega_r|\mathbf{x}) \leq \epsilon \\ \# & : \quad 1 - \max_{\omega_r \in \Omega} P(\omega_r|\mathbf{x}) > \epsilon \end{cases} \tag{17}$$

Odkazy

[1] Zdeněk Kotek, Ivan Brůha, Vaclav Chalupa, and Jan ık Jelínek. *Adaptivní a ucici se systemy*. SNTL Praha, 1980.

[2] Zdeněk Kotek, Vladimír Mařík, Václav Hlaváč, Josef Psutka, and Zdeněk Zdráhal. *Metody rozpoznávání a jejich aplikace*. ACADEMIA Praha, 1993.