Cvičení z RPZ – Bayesovské rozhodování

Vojtěch Franc, Martin Urban

19. října 2005

1 Bayesovské rozhodování

1.1 Bayesovské riziko

Nechť X značí množinu pozorování (příznakový prostor), Y množinu skrytých (odhadovaných) parametrů, $d : X \to Y$ rozhodovací strategii, a $W : Y \times Y \to R$ ztrátovou/pokutovou funkci.

Bayesovské riziko zvolené strategie $d(x)$ se definuje vztahem[1]:

$$
\begin{aligned}
R(d(x)) &= \int_X \sum_y p(x,y) W(d(x),y) dx && (1)\\
&= \int_X p(x) \sum_y p(y|x) W(d(x),y) dx\,, \quad x \in X, y \in Y && (2)
\end{aligned}
$$

1.2 Bayesovské rozhodování

Bayesovské rozhodování předpokládá znalost apriorních pravděpodobností $p(y)$ a podmíněných pravděpodobností $p(x|y)$ (tj. pravděpodobnosti pozorování x při stavu y).

Cílem bayesovského rozhodování je nalézt takovou strategii $d_b(x)$, která minimalizuje bayesovské riziko $R(d(x))$:

$$d_b(x) = \operatorname*{argmin}_{d(x)} R(d(x)) \qquad (3)$$

Pro jednodušší odvození optimální strategie $d_b(x)$ předpokládejme dále *zero-one* ztrátovou funkci W:

$$W(d(x),y) = 0 \Leftrightarrow d(x) = y \text{ (tj. pro správné rozhodnutí)}$$

[1] Volba integračních symbolů $\int$ a $\sum$ ve vztahu (1) implicitně předpokládá x ze spojité domény a y z diskrétní. Tato volba odpovídá příkladu řešeném v tomto cvičení.

$$W(d(x), y) = 1 \Leftrightarrow d(x) \neq y \text{ (tj. pro chybné rozhodnutí)}$$

Dosazením *zero-one* W do rovnice (1) získáme vztah

$$R(d(x)) = \int_X p(x) \sum_{y:\ y \neq d(x)} p(y|x) dx \ , \tag{4}$$

z kterého je již zřejmé, že optimální strategie $d_b(x)$ rozhoduje na základě maxima pravděpodobnosti $p(y|x)$:

$$d_b(x) = \operatorname*{argmax}_y p(y|x) \tag{5}$$

(Dk.: $p(x)$ je pro všechny strategie stejné, minimum rizika (4) je tedy určeno sumou. Pokud rozhodujeme dle maxima, tj. $d(x) = y_{max} = \operatorname{argmax}_y p(y|x)$, neobjeví se tato maximalní $p(y_{max}|x)$ ve zmíněné sumě a tudíž je tato suma nejmenší možná. Při jakémkoliv jiném rozhodnutí $d(x) \neq y_{max}$, bude suma zahrnovat $p(y_{max}|x)$ a bude tedy zákonitě vždy větší (či alespoň rovna té minimální).)

Ukázali jsme, že optimální bayesovská strategie $d_b(x)$ rozhoduje na základě maxima aposteriorní pravděpodobnosti $p(y|x)$

$$d_b(x) = \operatorname*{argmax}_y p(y|x) \ .$$

V bayesovské rozhodovací úloze se ale znalost aposteriorního rozdělení $p(y|x)$ nepředpokládá. Předpokládá se pouze znalost $p(y)$ a $p(x|y)$. Vyjádřeme tedy $d_b(x)$ pomocí apriorních $p(y)$ a $p(x|y)$. Pro přehlednost se omezíme pouze na dvě třídy $y \in \{1, 2\}$.

Dle optimální strategie $d_b(x)$ bude pozorování x přiřazen skrytý stav 1, právě tehdy když

$$p(1|x) \geq p(2|x) \ . \tag{6}$$

Jelikož $p(y|x) = p(x|y)p(y)/p(x)$ (bayesův vzorec), lze (6) snadno převést na

$$p(x|1)p(1) \geq p(x|2)p(2) \ ,$$

což je ekvivalentní zápisu

$$\frac{p(x|1)}{p(x|2)} \geq \frac{p(2)}{p(1)} \ . \tag{7}$$

Shrnutí: Pokud známe apriorní pravděpodobnosti $p(1)$, $p(2)$ a rozdělení v rámci tříd $p(x|1)$, $p(x|2)$, potom umíme na základě nerovnice (7) optimálně klasifikovat pozorování $x \in X$ do tříd $k \in \{1, 2\}$. (Optimálně ve smyslu minimilizace bayesovské chyby (1) při *zero-one* ztrátové funkci W).

1.3 Risk bayesovského rozhodování

Risk optimální strategie $d_b(x) = \text{argmax}_y\, p(y|x)$ za předpokladu *zero-one* ztrátové funkce W lze vyjádřit jako

$$
\begin{aligned}
R(d_b(x)) &= \int_X p(x) \sum_{k \neq \text{argmax}_y\, p(y|x)} p(y|x)dx = && (8)\\
&= \int_X p(x)[1 - \max_y p(y|x)]dx = && (9)\\
&= 1 - \int_X p(x) \max_y p(y|x)dx = && (10)\\
&= 1 - \int_X \max_y p(y)p(x|y)dx && (11)
\end{aligned}
$$

Úloha

Bayesovský odhad vyzkoušíme na jednoduchém příkladě. Jsme postaveni před problém navrhnout klasifikátor umístěný u vchodu do budovy FEL, který bude rozpoznávat pohlaví přicházející osoby (žena/muž) na základě jediného příznaku (např. délka vlasů). Budeme předpokládat, že od studijního oddělení známe počet studentek i studentů, (tj. známe apriorní pravděpodobnosti $p(\text{žena})$, $p(\text{muž})$) a dále podmíněné rozdělení pravděpodobností $p_{x|y}(x|\text{žena})$, $p_{x|y}(x|\text{muž})$, tj. pravděpodobnost délky vlasů x u žen a u mužů. Cílem je navrhnout klasifikátor s minimální bayesovskou chybou.

Podmíněné pravděpodobnosti $p_{x|y}(x|\text{žena})$, $p_{x|y}(x|\text{muž})$ budou gausovská rozdělení a budou zadány střední hodnotou a rozptylem.

1. Ukažte, že diskriminační funkce bayesovské strategie, za předpokladu, že $p_{x|y}(x|\text{žena})$, $p_{x|y}(x|\text{muž})$ jsou gausovského rozdělení, může být vyjádřena jako kvadratická funkce.

2. Pro

$$
\begin{aligned}
p(\text{žena}) &= 0.25,\\
p(\text{muž}) &= 0.75,\\
p_{x|y}(x|\text{žena}) &= N(25, 10),\\
p_{x|y}(x|\text{muž}) &= N(10, 5),
\end{aligned}
$$

vypočtěte parametry optimální kvadratické diskriminační funkce a bayesovský risk. K výpočtu bayesovského risku použijte funkci `bayeserr.m`. Ručně vypočtené parametry diskriminační funkce ověřte pomocí funkce `bayesdf.m`.

3. Na simulovaných datech ověřte funkčnost nalezeného klasifikátoru. Porovnejte chybu naměřenou na datech s bayesovským riskem z bodu 2. Porovnejte chybu optimální strategie s chybou jiné rozhodovací funkce (např. mírně pozměňte parametry optimální strategie).

   Testovací data vygenerujte funkcí `gmmsamp.m`. Syntaxe je `data = gmmsamp(struct('Mean',[m1 m2],'Cov',[c1 c2],'Prior',[p1 p2]),N)`, kde `m1, m2` jsou střední hodnoty a `c1, c2` jsou variance hustot pravděpodobnosti $p_{x|y}(x|1)$, $p_{x|y}(x|2)$; `p1, p2` jsou apriorní pravděpodobnosti a `N` je počet vzorků, které se vygenerují. Proměná `data.X` obsahuje generované pozorování a `data.y` odpovídající hodnoty skrytého parameteru.

4. Vypočtěte risk optimální bayesovské strategie pro apriorní pravděpodobnosti $p(\text{žena}) = 0.1, 0.2, ..., 0.9$. Výsledky zakreslete do grafu.