3. Analiza empirycznego klasyfikatora Bayesa - asymptotyczna zgodność

dr inż. Urszula Libal

Politechnika Wrocławska

2015

1. Teoretyczny i empiryczny klasyfikator bayesowski

klasyfikator teoretyczny			
klasa 1	klasa 2	$\cdots$	klasa M
p_1	p_2	$\cdots$	p_M
$f_1(x)$	$f_2(x)$	$\cdots$	$f_M(x)$

klasyfikator empiryczny			
klasa 1	klasa 2	$\cdots$	klasa M
$\hat{p}_1$	$\hat{p}_2$	$\cdots$	$\hat{p}_M$
$\hat{f}_1(x)$	$\hat{f}_2(x)$	$\cdots$	$\hat{f}_M(x)$
N_1	N_2	$\cdots$	N_M
$\left\{X_j^{(1)}\right\}_{j=1}^{N_1}$	$\left\{X_j^{(2)}\right\}_{j=1}^{N_2}$	$\cdots$	$\left\{X_j^{(M)}\right\}_{j=1}^{N_M}$

Optymalny klasyfikator bayesowski w przypadku dwóch klas to reguła postaci

$$\Psi^*(x) = \begin{cases} 1, & \text{gdy } p_1 f_1(x) > p_2 f_2(x), \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{1}$$

Empiryczny klasyfikator bayesowski w przypadku dwóch klas ma postać

$$\Psi^*(x) = \begin{cases} 1, & \text{gdy } \hat{p}_1 \hat{f}_1(x) > \hat{p}_2 \hat{f}_2(x), \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{2}$$

2. Histogram jako nieparametryczny estymator funkcji gęstości

Jednym z prostszych nieparametrycznych estymatorów funkcji gęstości prawdopodobieństwa jest odpowiednio przeskalowany histogram

$$
\begin{aligned}
\hat{f}(x) &= \hat{f}(x;\Delta,N) && (3)\\
&= \frac{1}{N\Delta}\sum_{j=1}^{N} 1\left\{x < X_j \leq x+\Delta\right\} && (4)\\
&= \frac{1}{N\Delta}\#\left\{x < X_j \leq x+\Delta\right\}, && (5)
\end{aligned}
$$

którego kształt zależy od liczności N zbioru uczącego (próby) $\left\{X_j\right\}_{j=1}^{N}$ oraz szerokości przedziałów Δ.

3. Typy zbieżności probabilistycznych

Definicja 1. Ciąg $(\theta_N)_{N\in\mathbb{N}}$ jest *zbieżny z prawdopodobieństwem 1* (mocno) do θ, jeśli

$$P\left\{\lim_{N\to\infty}\theta_N=\theta\right\}=1. \tag{6}$$

Definicja 2. Ciąg $(\theta_N)_{N\in\mathbb{N}}$ jest *zbieżny według prawdopodobieństwa* (słabo) do θ, jeśli

$$\forall_{\varepsilon>0}\lim_{N\to\infty}P\{|\theta_N-\theta|<\varepsilon\}=1. \tag{7}$$

Definicja 3. Ciąg $(\theta_N)_{N\in\mathbb{N}}$ jest *zbieżny według rozkładu* do θ, jeśli

$$\lim_{N\to\infty}F_N(x)=F(x). \tag{8}$$

Definicja 4. Ciąg $(\theta_N)_{N\in\mathbb{N}}$ jest *zbieżny według średniej* z potęgą r do θ, jeśli

$$\lim_{N\to\infty} E|\theta_N - \theta|^r = 0. \tag{9}$$

W szczególności, ciąg $(\theta_N)_{N\in\mathbb{N}}$ jest *zbieżny średniokwadratowo* do θ, jeśli

$$\lim_{N\to\infty} E\left(\theta_N - \theta\right)^2 = 0. \tag{10}$$

4. Zależności między typami zbieżności probabilistycznych

$$\theta_N \stackrel{P1}{\to} \theta \;\;\Rightarrow\;\; \theta_N \stackrel{P}{\to} \theta, \tag{11}$$

$$\theta_N \stackrel{L_r}{\to} \theta \;\;\Rightarrow\;\; \theta_N \stackrel{P}{\to} \theta, \tag{12}$$

$$\theta_N \stackrel{P}{\to} \theta \;\;\Rightarrow\;\; \theta_N \stackrel{D}{\to} \theta. \tag{13}$$

5. Zgodność estymatora

Definicja. Estymator θ_N parametru θ nazywamy *zgodnym*, jeśli

$$\forall_{\varepsilon>0} \lim_{N\to\infty} P\{|\theta_N - \theta| < \varepsilon\} = 1. \tag{14}$$

Równoważnie warunek (14) można zapisać

$$\forall_{\varepsilon>0} \lim_{N\to\infty} P\{|\theta_N - \theta| > \varepsilon\} = 0. \tag{15}$$

Zachodzi również

$$\forall_{\varepsilon>0} P\{|\theta_N - \theta| > \varepsilon\} \leq \frac{1}{\varepsilon^2} E\left(\theta_N - \theta\right)^2. \tag{16}$$

6. Obciążenie i wariancja estymatora

$$
\begin{aligned}
E\left(\theta_N-\theta\right)^2 &= E\theta_N^2-2\theta E\theta_N+\theta^2 && (17)\\
&= E\theta_N^2-\left(E\theta_N\right)^2+\left(E\theta_N\right)^2-2\theta E\theta_N+\theta^2 && (18)\\
&= \underbrace{E\theta_N^2-\left(E\theta_N\right)^2}_{var(\theta_N)}+\underbrace{\left(E\theta_N-\theta\right)^2}_{bias^2(\theta_N)} && (19)
\end{aligned}
$$

7. Obciążenie estymatora funkcji gęstości

$$
\begin{aligned}
E\,\hat{f}(x) &= E\left(\frac{1}{N\Delta}\sum_{j=1}^{N} 1\left\{x < X_j \leq x+\Delta\right\}\right) && (20)\\
&= \frac{1}{N\Delta}\sum_{j=1}^{N} E1\left\{x < X_j \leq x+\Delta\right\} && (21)\\
&= \frac{1}{N\Delta} N E1\left\{x < X_j \leq x+\Delta\right\} && (22)\\
&= \frac{1}{\Delta} P\left\{x < X_j \leq x+\Delta\right\} && (23)\\
&= \frac{F(x+\Delta)-F(x)}{\Delta} && (24)\\
&\neq f(x) && (25)
\end{aligned}
$$

8. Wariancja estymatora funkcji gęstości

$$
\begin{aligned}
var\,\hat{f}(x) &= var\left(\frac{1}{N\Delta}\sum_{j=1}^{N} 1\left\{x < X_j \leq x+\Delta\right\}\right) && (26)\\
&= \frac{1}{N^2\Delta^2}\sum_{j=1}^{N} var 1\left\{x < X_j \leq x+\Delta\right\} && (27)\\
&= \frac{1}{N^2\Delta^2} N\, var 1\left\{x < X_j \leq x+\Delta\right\} && (28)\\
&= \frac{1}{N\Delta^2}\left[E1^2\left\{x < X_j \leq x+\Delta\right\} - \left(E1\left\{x < X_j \leq x+\Delta\right\}\right)^2\right] && (29)\\
&= \frac{1}{N\Delta^2}\left[P\left\{x < X_j \leq x+\Delta\right\} - \left(P\left\{x < X_j \leq x+\Delta\right\}\right)^2\right] && (30)\\
&= \frac{1}{N\Delta^2}\left[F(x+\Delta) - F(x) - (F(x+\Delta) - F(x))^2\right] && (31)\\
&= \frac{1}{N\Delta}\left(\frac{F(x+\Delta) - F(x)}{\Delta}\right)\left[1 - (F(x+\Delta) - F(x))\right] && (32)
\end{aligned}
$$

9. Asymptotyczna redukcja obciążenia i wariancji

Dla ustalonego N, przy $\Delta \to 0$

$$E\,\hat{f}(x) = \frac{F(x+\Delta) - F(x)}{\Delta} \to f(x), \tag{33}$$

czyli

$$bias\,\hat{f}(x) \to 0. \tag{34}$$

Wtedy

$$var\,\hat{f}(x) \quad = \quad \underbrace{\frac{1}{N\Delta}}_{\to\infty} \underbrace{\left(\frac{F(x+\Delta) - F(x)}{\Delta}\right)}_{\to f(x)} \underbrace{[1 - (F(x+\Delta) - F(x))]}_{\to 1} \to \infty. \tag{35}$$

Przy jednoczesnym $\Delta \to 0$ oraz $N\Delta \to \infty$

$$bias\, \hat{f}(x) \to 0 \tag{36}$$

oraz

$$var\, \hat{f}(x) \quad = \quad \underbrace{\frac{1}{N\Delta}}_{\to 0} \underbrace{\left(\frac{F(x+\Delta) - F(x)}{\Delta} \right)}_{\to f(x)} \underbrace{[1 - (F(x+\Delta) - F(x))]}_{\to 1} \to 0. \tag{37}$$

Wtedy także

$$E\,(\theta_N - \theta)^2 \to 0. \tag{38}$$

Histogram dany wzorem (4) jest *asymptotycznie zgodnym estymatorem* funkcji gęstości prawdopodobieństwa przy $N \to \infty$, jeśli jednocześnie

$$\Delta \to 0 \tag{39}$$

oraz

$$N\Delta \to \infty. \tag{40}$$

Szerokość przedziału Δ może być funkcją N, np. $\Delta_N = \frac{1}{\sqrt{N}}$.