2. Empiryczna wersja klasyfikatora bayesowskiego

dr inż. Urszula Libal

Politechnika Wrocławska

2015

1. Brak pełnej informacji probabilistycznej

Klasyfikator bayesowski wymaga pełnej informacji probabilistycznej, tzn. muszą być znane prawdopodobieństwa *a priori* klas oraz funkcje gęstości prawdopodobieństwa w klasach.

Nieznane rozkłady prawdopodobieństwa w klasach $1, 2, \ldots, M$ można estymować na podstawie M ciągów uczących.

Każdy ciąg uczący zawiera N_k obserwacji $\left\{X_j^{(k)}\right\}_{j=1}^{N_k}$ z klasy $k \in \{1, 2, \ldots, M\}$.

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

2. Estymacja prawdopodobieństw a priori klas

Prawdopodobieństwa a priori klas estymujemy za pomocą *częstości* ich występowania, tj.

$$\hat{p}_k = \frac{N_k}{N_1 + N_2 + \cdots + N_M}. \tag{1}$$

dla każdej klasy $k \in \mathscr{M}$, $\mathscr{M} = \{1, 2, \ldots, M\}$.

3. Estymacja funkcji gęstości prawdopodobieństwa

— *Metody parametryczne* - zakładamy pewien rozkład prawdopodobieństwa i estymujemy jego parametry.

— *Metody nieparametryczne* - umożliwiają estymację dowolnego rozkładu.

4. Histogram

Standardowy *histogram* dzieli przestrzeń cech $\mathscr{X}$ na przedziały o szerokości Δ, a następnie zlicza liczbę n_i obserwacji, które wpadły do i-tego przedziału $I_i = (t_i, t_i + \Delta]$, tzn.

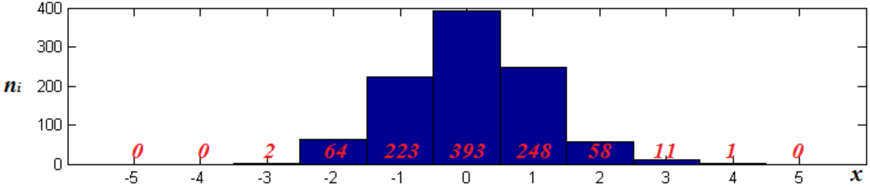
$$
\begin{aligned}
n_i &= \#\left\{X_j \in I_i\right\} = \#\left\{X_j \in (t_i, t_i + \Delta]\right\} && (2)\\
&= \#\left\{X_j \le t_i + \Delta\right\} - \#\left\{X_j < t_i\right\} && (3)\\
&= \sum_{j=1}^{N} \left(1\left\{X_j \le t_i + \Delta\right\} - 1\left\{X_j < t_i\right\}\right). && (4)
\end{aligned}
$$

Histogram to funkcja schodkowa

$$h(x;\Delta,N)=\sum_i n_i 1\{x\in I_i\}, \tag{5}$$

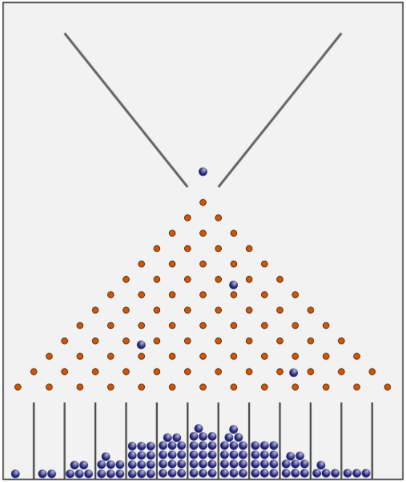
która w punkcie x należącym do i-tego przedziału przyjmuje wartość n_i

(pozostałe elementy sumy są wtedy zerowe).



Rysunek 1. Przykładowy histogram dla ciągu uczącego o liczności $N = 1000$.

Źródło: opracowanie własne



Rysunek 2. Deska Galtona - przykład nieparametrycznej estymacji funkcji gęstości dla rozkładu normalnego.

Źródło: [2]

5. Nieparametryczna estymacja funkcji gęstości prawdopodobieństwa

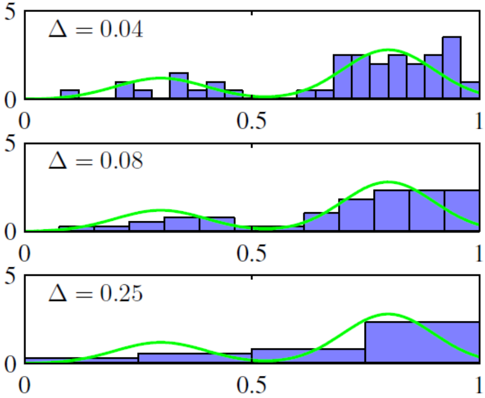
Funkcja gęstości prawdopodobieństwa $f(x)$ spełnia warunek

$$\int_{\mathscr{X}} f(x) = 1. \tag{6}$$

Aby histogram estymował funkcję gęstości prawdopodobieństwa, należy go znormalizować poprzez podzielenie liczby n_i obserwacji przez całkowitą liczbę obserwacji N oraz szerokość przedziału Δ.

Nieparametryczny estymator funkcji gęstości prawdopodobieństwa przyjmuje formę:

$$\hat{f}(x) = \hat{f}(x; \Delta, N) = \sum_i \frac{n_i}{N\Delta} 1\left\{x \in I_i\right\}. \tag{7}$$



Rysunek 3. Nieparametryczna estymacja gęstości dla różnych szerokości przedziałów histogramu.

Źródło: [1]

5. Empiryczna wersja klasyfikatora bayesowskiego

Korzystając z zasady *plug-in*, wstawiamy do algorytmu bayesowskiego odpowiednie estymatory:

— w miejsce prawdopodobieństw *a priori* wystąpienia klas p_k - ich częstości $\hat{p}_k$, $k \in \mathscr{M}$,

— w miejsce funkcji gęstości prawdopodobieństwa w klasach $f_k(x)$ - ich nieparametryczne estymatory $\hat{f}_k(x)$, $k \in \mathscr{M}$.

Empiryczny klasyfikator bayesowski w przypadku dwóch klas przyjmuje wtedy postać

$$\Psi^*(x) = \begin{cases} 1, & \text{gdy } \hat{p}_1 \hat{f}_1(x) > \hat{p}_2 \hat{f}_2(x), \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{8}$$

Jeżeli w obu klasach przyjęto identyczny podział przestrzeni cech na przedziały I_i, to empiryczny klasyfikator bayesowski w przypadku dwóch klas otrzymuje postać

$$\Psi^*(x) = \begin{cases} 1, & \text{gdy } \frac{N_1}{(N_1+N_2)} \sum_i \frac{n_i^{(1)}}{N_1 \Delta} 1\{x \in I_i\} > \frac{N_2}{(N_1+N_2)} \sum_i \frac{n_i^{(2)}}{N_2 \Delta} 1\{x \in I_i\}, \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{9}$$

Regułę decyzyjną (9) można uprościć. Jeżeli zaobserwowano cechę o wartości x z i-tego przedziału histogramu, tzn. $x \in I_i$, to empiryczny klasyfikator bayesowski bazuje jedynie na licznościach obserwacji w tym przedziale w klasie 1 i klasie 2

$$\Psi^*(x) = \begin{cases} 1, & \text{gdy } n_i^{(1)} > n_i^{(2)}, \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{10}$$

Przy założeniu, że analizowany obraz

$$x \in (t_i, t_i + \Delta], \tag{11}$$

empiryczny klasyfikator bayesowski przyjmuje ostatecznie postać opartą w jawny sposób o ciągi uczące $\left\{X_j^{(1)}\right\}_{j=1}^{N_1}$ oraz $\left\{X_j^{(2)}\right\}_{j=1}^{N_2}$:

$$\Psi^*(x) = \begin{cases} 1, & \text{gdy } \sum_{j=1}^{N_1}\left(1\left\{X_j^{(1)} \le t_i + \Delta\right\} - 1\left\{X_j^{(1)} < t_i\right\}\right) \\ & \quad > \sum_{j=1}^{N_2}\left(1\left\{X_j^{(2)} \le t_i + \Delta\right\} - 1\left\{X_j^{(2)} < t_i\right\}\right), \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{12}$$

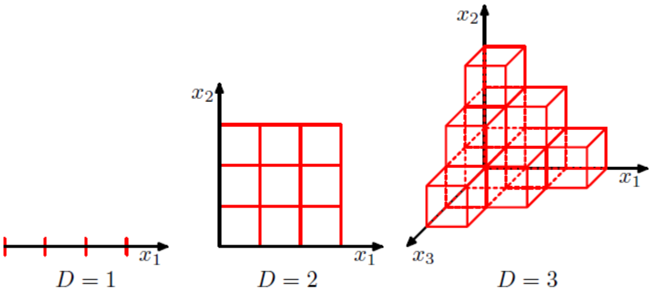
6. Zjawisko pustej przestrzeni

Zjawisko pustej przestrzeni, inaczej zwane *przekleństwem wymiarowości*,

występuje w przypadku:

— dużej liczby cech (dużego wymiaru zadania $D \gg 1$),

— zbyt małej liczby obserwacji (liczności ciągów uczących N_k w klasach $k \in \mathscr{M}$).



Rysunek 4. Ilustracja przekleństwa wymiarowości, obrazująca wykładniczy wzrost liczby obszarów, na które podzielono przestrzeń cech w zadaniu estymacji nieparametrycznej.

Źródło: [1]

Literatura

[1] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer Series: Information Science and Statistics (2006).

[2] http://pl.wikipedia.org/wiki/Deska_Galtona

[3] http://www.youtube.com (szukaj pod hasłem: *Galton board*)