4. Naiwny klasyfikator Bayesa

dr inż. Urszula Libal

Politechnika Wrocławska

2015

1. Wektory cech

Rozpoznawanie D-wymiarowych wektorów cech $\mathbf{x} = (x_1, x_2, \ldots, x_D)$.

Fisher's *Iris* Data

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	*I. setosa*
4.9	3.0	1.4	0.2	*I. setosa*
4.7	3.2	1.3	0.2	*I. setosa*
		⋮		
5.9	3.0	5.1	1.8	*I. virginica*

Rysunek 1. Zbiór danych *fisheriris*: 4 cechy (długość i szerokość działki kielicha, długość i szerokość płatka), 3 klasy (gatunki irysa).

Źródło: [1]

Rysunek 2. Klasyfikacja do 3 klas określonych przez gatunek irysa.

Źródło: [1]

2. Klasyfikator Bayesa - przypadek wielowymiarowy

W przypadku obrazów opisanych przez D-wymiarowe wektory cech

$$\mathbf{x} = (x_1, x_2, \ldots, x_D)\,, \tag{1}$$

klasyfikator Bayesa wskazuje na klasę $i \in \mathscr{M}$

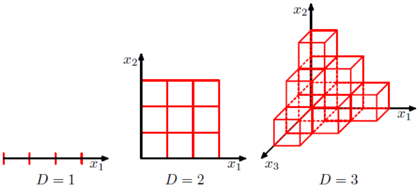
$$\Psi^*(\mathbf{x}) = i, \;\; \text{jeżeli } p_i f_i(\mathbf{x}) = \max_{k \in \mathscr{M}} p_k f_k(\mathbf{x}). \tag{2}$$

($\mathscr{M}$ - zbiór klas)

3. Estymacja funkcji gęstości a przekleństwo wymiarowości

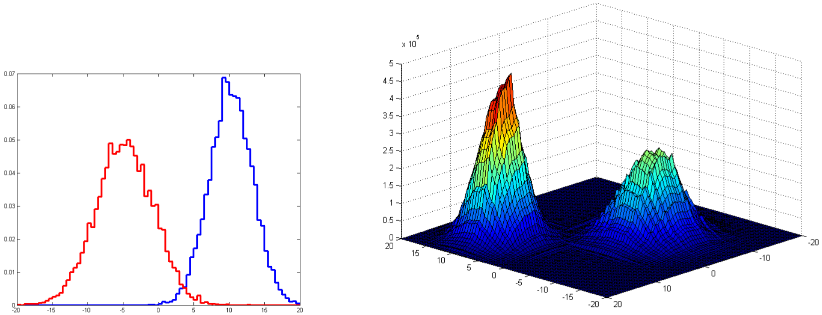
Przekleństwo wymiarowości (inaczej zwane *zjawiskiem pustej przestrzeni*)

- związane jest z wykładniczym wzrostem liczby D-wymiarowych kostek, stanowiących podział przestrzeni cech podczas nieparametrycznej estymacji funkcji gęstości, przy zwiększaniu rozmiaru D wektora cech.



Rysunek 3. Ilustracja przekleństwa wymiarowości.

Źródło: [2]



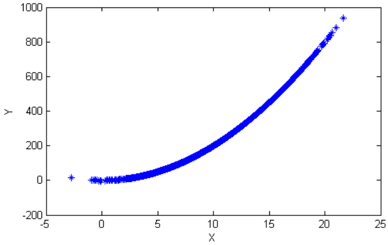
Rysunek 4. Nieparametryczna estymacja funkcji gęstości
dla liczby cech $D = 1$ oraz $D = 2$.

Źródło: opracowanie własne

4. Naiwny klasyfikator Bayesa

Naiwny klasyfikator Bayesa Ψ_{NB} to klasyfikator Bayesa Ψ^*,

dla którego zakłada się, że cechy $X_1, X_2, \ldots, X_D$ są wzajemnie niezależne!



Rysunek 5. Przykład zmiennych losowych zależnych X i $Y = 2X^2 - 1$.

Źródło: opracowanie własne

Definicja 1. Zmienne losowe $X_1, X_2, \ldots, X_D$ są *niezależne* wtedy i tylko wtedy, gdy

$$P\{X_1 < x_1, X_2 < x_2, \ldots, X_D < x_D\} = P\{X_1 < x_1\} P\{X_2 < x_2\} \ldots P\{X_D < x_D\}, \quad (3)$$

czyli

$$F_{X_1, X_2, \ldots, X_D}(\mathbf{x}) = \prod_{d=1}^{D} F_{X_d}(x_d). \quad (4)$$

Definicja 2. Zmienne losowe $X_1, X_2, \ldots, X_D$ są *niezależne* wtedy i tylko wtedy, gdy

$$f_{X_1, X_2, \ldots, X_D}(\mathbf{x}) = \prod_{d=1}^{D} f_{X_d}(x_d). \quad (5)$$

Naiwny klasyfikator Bayesa wskazuje na klasę $i \in \mathscr{M}$ na podstawie zaobserwowanego wektora cech $\mathbf{x} = (x_1, x_2, \ldots, x_D)$

$$\Psi_{NB}(\mathbf{x}) = i, \quad \text{jeżeli } p_i \prod_{d=1}^{D} f_i^{(d)}(x_d) = \max_{k \in \mathscr{M}} p_k \prod_{d=1}^{D} f_k^{(d)}(x_d). \tag{6}$$

Zasada działania pozostaje identyczna jak dla klasyfikatora Bayesa, tzn. maksymalizowane jest prawdopodobieństwo *a posteriori* - patrz wzór (21) z wykładu nr 1. Zakładając niezależność cech otrzymujemy, że funkcja gęstości f_k łącznego rozkładu w klasie $k \in \mathscr{M}$ to iloczyn gęstości brzegowych $f_k^{(d)}$, $d = 1, 2, \ldots, D$,

$$f_k(\mathbf{x}) = \prod_{d=1}^{D} f_k^{(d)}(x_d). \tag{7}$$

5. Naiwny klasyfikator Bayesa - przypadek dwóch klas

Naiwny klasyfikator Bayesa na podstawie zaobserwowanego wektora cech

$\mathbf{x} = (x_1, x_2, \ldots, x_D)$ wskazuje na klasę

$$\Psi_{NB}(\mathbf{x}) = \begin{cases} 1, & \text{gdy } p_1 \prod_{d=1}^{D} f_1^{(d)}(x_d) > p_2 \prod_{d=1}^{D} f_2^{(d)}(x_d), \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{8}$$

Warunek

$$p_1 \prod_{d=1}^{D} f_1^{(d)}(x_d) > p_2 \prod_{d=1}^{D} f_2^{(d)}(x_d) \tag{9}$$

można przekształcić na warunek równoważny:

$$\frac{p_1 \prod_{d=1}^{D} f_1^{(d)}(x_d)}{p_2 \prod_{d=1}^{D} f_2^{(d)}(x_d)} > 1 \tag{10}$$

$$\ln \frac{p_1 \prod_{d=1}^{D} f_1^{(d)}(x_d)}{p_2 \prod_{d=1}^{D} f_2^{(d)}(x_d)} > \ln 1 \tag{11}$$

$$\ln \frac{p_1}{p_2} + \ln \prod_{d=1}^{D} \frac{f_1^{(d)}(x_d)}{f_2^{(d)}(x_d)} > 0 \tag{12}$$

$$\ln \frac{p_1}{p_2} + \sum_{d=1}^{D} \ln \frac{f_1^{(d)}(x_d)}{f_2^{(d)}(x_d)} > 0 \tag{13}$$

Wyrażenie

$$\delta(\mathbf{x}) = \ln \frac{p_1}{p_2} + \sum_{d=1}^{D} \ln \frac{f_1^{(d)}(x_d)}{f_2^{(d)}(x_d)} \tag{14}$$

będziemy nazywać funkcją dyskryminacyjną między klasami 1 i 2.

Wtedy naiwny klasyfikator Bayesa można zapisać

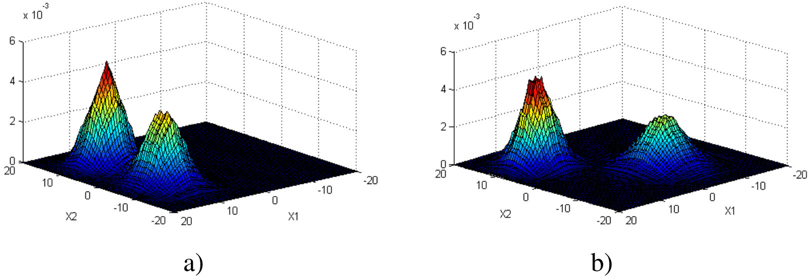
$$\Psi_{NB}(\mathbf{x}) = \begin{cases} 1, & \text{gdy } \delta(\mathbf{x}) > 0, \\ 2, & \text{w przeciwnym wypadku.} \end{cases} \tag{15}$$

Przykład: 2 klasy $\{1,2\}$ i 2 cechy $\mathbf{x} = (x_1, x_2)$,

prawdopodobieństwa a priori równe $p_1 = p_2 = 0.5$

(a) $f_1^{(1)} \sim \mathcal{N}(10,3)$, $f_1^{(2)} \sim \mathcal{N}(10,3)$, $f_2^{(1)} \sim \mathcal{N}(10,3)$, $f_2^{(2)} \sim \mathcal{N}(-5,4)$

(b) $f_1^{(1)} \sim \mathcal{N}(10,3)$, $f_1^{(2)} \sim \mathcal{N}(10,3)$, $f_2^{(1)} \sim \mathcal{N}(-5,4)$, $f_2^{(2)} \sim \mathcal{N}(-5,4)$



Rysunek 6. Przykład dyskryminacji między klasami.

Źródło: opracowanie własne

W przypadku (a) $p_1 = p_2 = 0.5$ oraz $f_1^{(1)}(x_1) = f_2^{(1)}(x_1)$ dla każdego x_1. Wtedy funkcja dyskryminacyjna otrzymuje postać

$$\begin{aligned} \delta(\mathbf{x}) &= \ln \frac{p_1}{p_2} + \sum_{d=1}^{D} \ln \frac{f_1^{(d)}(x_d)}{f_2^{(d)}(x_d)} && (16) \\ &= \ln \frac{f_1^{(2)}(x_2)}{f_2^{(2)}(x_2)}, && (17) \end{aligned}$$

ponieważ $\ln \frac{p_1}{p_2} = 0$ oraz $\ln \frac{f_1^{(1)}(x_1)}{f_2^{(1)}(x_1)} = 0$. Dyskryminacja między klasami odbywa się jedynie na podstawie wartości funkcji gęstości dla drugiej cechy x_2.

W przypadku (b) $p_1 = p_2 = 0.5$, więc $\ln \frac{p_1}{p_2} = 0$. Wtedy funkcja dyskryminacyjna ma postać

$$\delta(\mathbf{x}) \quad = \quad \ln \frac{f_1^{(1)}(x_1)}{f_2^{(1)}(x_1)} + \ln \frac{f_1^{(2)}(x_2)}{f_2^{(2)}(x_2)}. \tag{18}$$

Dyskryminacja między klasami odbywa się na podstawie wartości funkcji gęstości dla obu cech x_1 i x_2.

Literatura

[1] http://en.wikipedia.org/wiki/Iris_flower_data_set

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

[3] J. Koronacki, J. Ćwik, *Statystyczne systemy uczące się*, WNT, Warszawa (2005)