Statystyczna Eksploracja Danych

Wykład 2 - metoda Fishera dla $g > 2$, klasyfikator Bayesa, metody LDA i QDA

dr inż. Julian Sienkiewicz

12 marca 2021

Mamy próby należące do dwóch klas $y \in \{A, B\}$: $(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), ..., (\mathbf{x_n}, y_n)$, gdzie $\mathbf{x_i}$ to i-ta obserwacja, a y_i to jej przynależność do jednej z klas.

Każdy element $\mathbf{x}$ jest wektorem **kolumnowym** o długości p

$$\bar{\mathbf{x}}_\mathbf{i} = \begin{bmatrix} x_i^1 \\ x_i^2 \\ ... \\ x_i^p \end{bmatrix}$$

Wartość p określa wymiar układu (ogólnie niekoniecznie przestrzeń fizyczna).

W każdej klasie możemy wyznaczyć wartość średnią

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{i=n} \mathbf{x_i}$$

oraz macierz kowariancji

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{i=n} (\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T$$

Główne założenie: macierze kowariancji **w obu klasach** są takie same. W efekcie macierz kowarinacji wewnątrzgrupowej **W** wyraża się wzorem:

$$\mathbf{W} = \frac{1}{n-2}\sum_{k=1}^{i=2}(n_k - 1)\mathbf{S_k}$$

$\mathbf{S_k}$ - macierz kowariancji w klasie k
$n = n_1 + n_2$ - liczba danych w obu klasach

(Pierwszy) wektor kanoniczny

$$\widetilde{\mathbf{a}} \sim \mathbf{W}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)$$

(Pierwszy) wektor kanoniczny

$$\widetilde{\mathbf{a}} \sim \mathbf{W}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)$$

Czyli obserwacja **x** spełniająca warunek

Warunek dyskryminacyjny

$$|\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_1)| < |\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_2)|$$

zostanie uznana za przynależną do klasy A.

(Pierwszy) wektor kanoniczny

$$\widetilde{\mathbf{a}} \sim \mathbf{W}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)$$

Czyli obserwacja **x** spełniająca warunek

Warunek dyskryminacyjny

$$|\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_1)| < |\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_2)|$$

zostanie uznana za przynależną do klasy A. Natomiast hiperpłaszczyzna dyskryminacyjna jest dana równaniem

Hiperpłaszczyzna dyskryminacyjna

$$(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T\mathbf{W}^{-1}\left[\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\right] = 0$$

Można zauważyć, że

$$\left(\mathbf{a}^T\bar{\mathbf{x}}_{\mathbf{2}} - \mathbf{a}^T\bar{\mathbf{x}}_{\mathbf{1}}\right)^2 = \mathbf{a}^T(\bar{\mathbf{x}}_{\mathbf{2}} - \bar{\mathbf{x}}_{\mathbf{1}})(\bar{\mathbf{x}}_{\mathbf{2}} - \bar{\mathbf{x}}_{\mathbf{1}})^T\mathbf{a} = \tag{1}$$

$$= \frac{n_1 + n_2}{n_1 n_2}\mathbf{a}^T \sum_{k=1}^{2} n_k(\bar{\mathbf{x}}_{\mathbf{k}} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{\mathbf{k}} - \bar{\mathbf{x}})^T\mathbf{a}$$

gdzie $\bar{\mathbf{x}}$ to średnia ogólna ze wszystkich obserwacji.

Można zauważyć, że

$$\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2 = \mathbf{a}^T(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T\mathbf{a} = \tag{1}$$

$$= \frac{n_1 + n_2}{n_1 n_2}\mathbf{a}^T \sum_{k=1}^{2} n_k(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T\mathbf{a}$$

gdzie $\bar{\mathbf{x}}$ to średnia ogólna ze wszystkich obserwacji. Możemy dzięki temu zdefiniować macierz charakteryzującą **zmienność międzygrupową**

Macierz kowariancji międzygrupowej

$$\mathbf{B} = \sum_{k=1}^{2} n_k(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T$$

W efekcie maksymalizacja ze względu na **a** ilorazu

$$\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

W efekcie maksymalizacja ze względu na **a** ilorazu

$$\frac{\left(\mathbf{a}^T\bar{\mathbf{x_2}} - \mathbf{a}^T\bar{\mathbf{x_1}}\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

jest równoznaczna maksymalizacji wyrażenia

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}},$$

W efekcie maksymalizacja ze względu na **a** ilorazu

$$\frac{\left(\mathbf{a}^T\bar{\mathbf{x}}_2 - \mathbf{a}^T\bar{\mathbf{x}}_1\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

jest równoznaczna maksymalizacji wyrażenia

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}},$$

co daje możliwość uogólnienia metody Fishera na liczbę $g > 2$ klas.

Problem dyskryminacyjny $g > 2$

Znajdź kierunek $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}}$$

Problem dyskryminacyjny $g > 2$

Znajdź kierunek $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

gdzie

Macierz kowariancji międzygrupowej

$$\mathbf{B} = \frac{1}{g-1}\sum_{k=1}^{g} n_k(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T$$

Problem dyskryminacyjny $g > 2$

Znajdź kierunek $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}}$$

gdzie

Macierz kowariancji międzygrupowej

$$\mathbf{B} = \frac{1}{g-1} \sum_{k=1}^{g} n_k (\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{k} - \bar{\mathbf{x}})^T$$

i

Macierz kowariancji wewnątrzgrupowej

$$\mathbf{W} = \frac{1}{n-g} \sum_{k=1}^{g} n_k (n_k - 1) \mathbf{S}_k$$

Rozwiązanie $g > 2$

Wektor $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}}$$

jest **wektorem własnym** macierzy $\mathbf{W}^{-1}\mathbf{B}$, odpowiadającym największej **wartości własnej** tej macierzy

Rozwiązanie $g > 2$

Wektor $\widetilde{\mathbf{a}}$ maksymalizujący wyrażenie

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

jest **wektorem własnym** macierzy $\mathbf{W}^{-1}\mathbf{B}$, odpowiadającym największej **wartości własnej** tej macierzy

Reguła decyzyjna

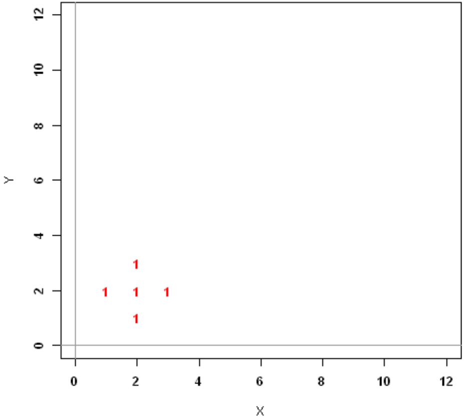
Obserwację $\mathbf{x}$ przypisujemy do klasy j ,jeżeli

$$|\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_\mathbf{j})| < |\widetilde{\mathbf{a}}^T(\mathbf{x} - \bar{\mathbf{x}}_\mathbf{k})|$$

dla wszystkich $k \neq j$.

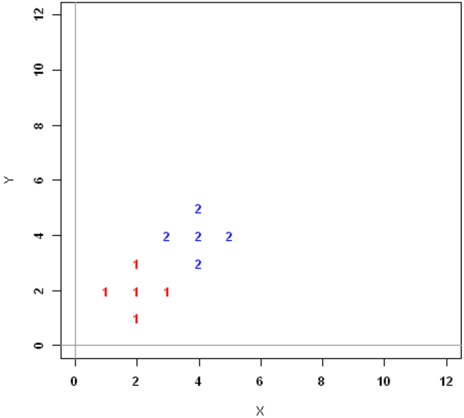
Przykład 4 - $g = 5$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$



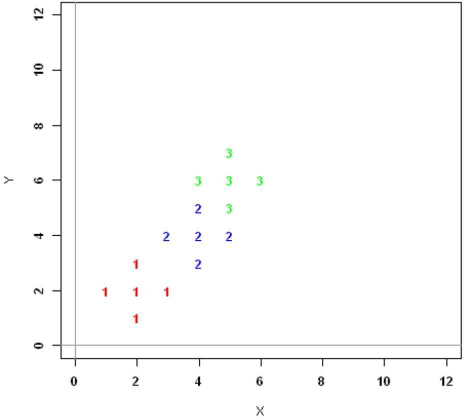
Przykład 4 - $g = 5$

$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ $\bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$



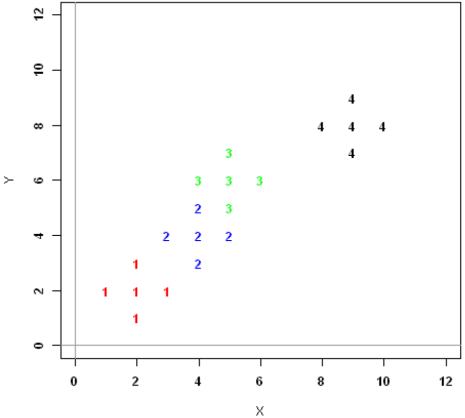
Przykład 4 - $g = 5$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$



Przykład 4 - $g = 5$

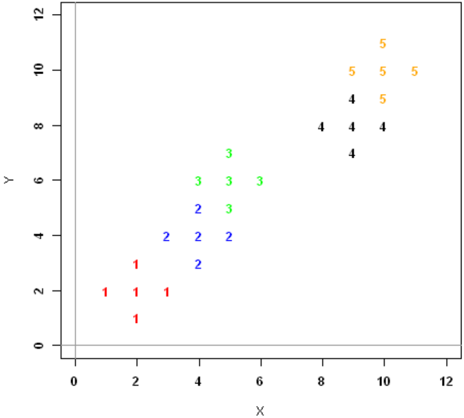
$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$
$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$



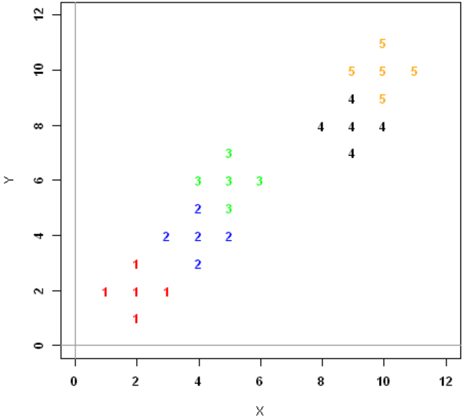
Przykład 4 - $g = 5$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \quad \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$



Przykład 4 - $g = 5$



$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

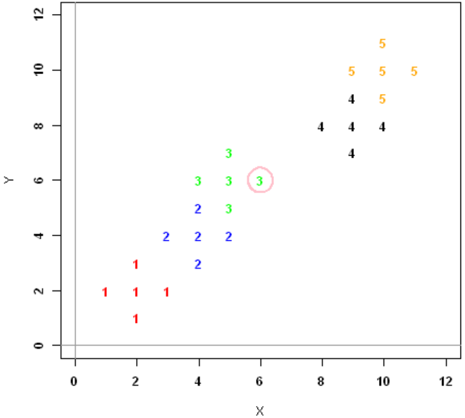
Przykład 4 - $g = 5$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

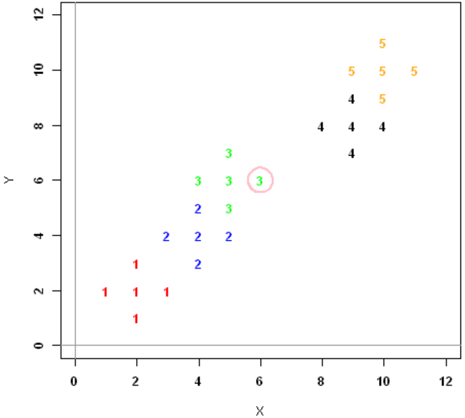
$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \quad \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$



Przykład 4 - $g = 5$



$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

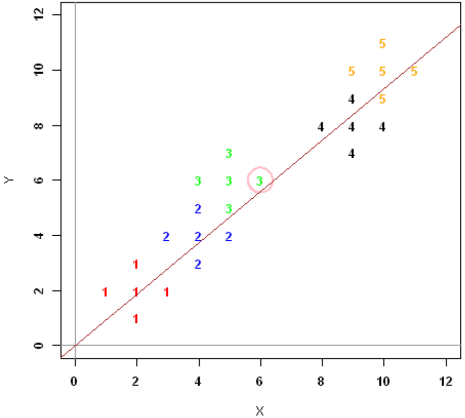
$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$

$$\lambda_1 \approx \frac{425}{2} \; \mathbf{p}_1 = \begin{bmatrix} \frac{15}{14} \\ 1 \end{bmatrix}$$

$$\lambda_2 \approx \frac{5}{2} \; \mathbf{p}_2 = \begin{bmatrix} -\frac{13}{14} \\ 1 \end{bmatrix}$$

Przykład 4 - $g = 5$



$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \quad \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

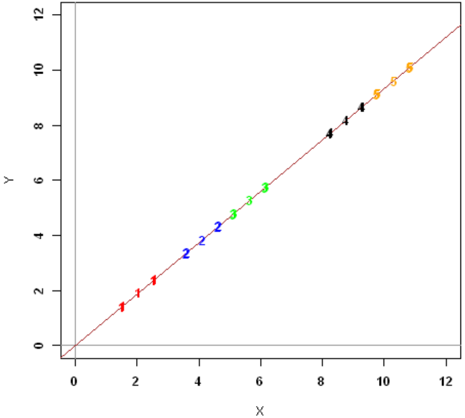
$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$

$$\lambda_1 \approx \frac{425}{2} \quad \mathbf{p}_1 = \begin{bmatrix} \frac{15}{14} \\ 1 \end{bmatrix}$$

$$\lambda_2 \approx \frac{5}{2} \quad \mathbf{p}_2 = \begin{bmatrix} -\frac{13}{14} \\ 1 \end{bmatrix}$$

$$y = \frac{14}{15} x$$

Przykład 4 - $g = 5$



$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \bar{\mathbf{x}}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_4 = \begin{bmatrix} 9 \\ 8 \end{bmatrix} \bar{\mathbf{x}}_5 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

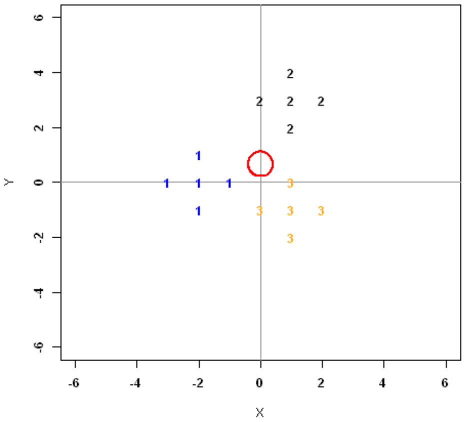
$$\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \mathbf{B} = \begin{bmatrix} \frac{115}{2} & \frac{105}{2} \\ \frac{105}{2} & 50 \end{bmatrix}$$

$$\lambda_1 \approx \frac{425}{2} \; \mathbf{p}_1 = \begin{bmatrix} \frac{15}{14} \\ 1 \end{bmatrix}$$

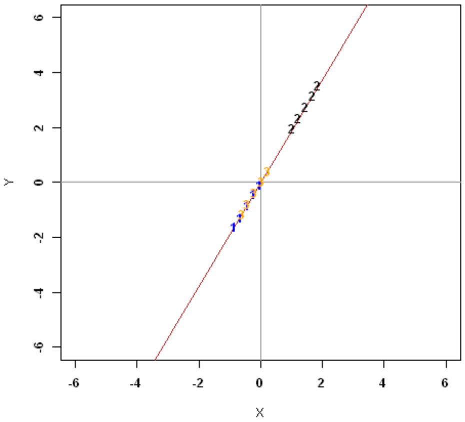
$$\lambda_2 \approx \frac{5}{2} \; \mathbf{p}_2 = \begin{bmatrix} -\frac{13}{14} \\ 1 \end{bmatrix}$$

$$y = \frac{14}{15} x$$

Przykład 5 - $g = 3$



Przykład 5 - $g = 3$



Klasyfikator bayesowski

Zaobserwowany wektor $\mathbf{x}$ klasyfikujemy jako pochodzący z tej klasy k, dla której wartość

$$p(k|\mathbf{x}) \qquad k = 1, ..., g$$

jest największa.

$p(k|\mathbf{x})$ - prawdopodobieństwo, że obserwacja $\mathbf{x}$ pochodzi z klasy k

Klasyfikator bayesowski

Zaobserwowany wektor $\mathbf{x}$ klasyfikujemy jako pochodzący z tej klasy k, dla której wartość

$$p(k|\mathbf{x}) \qquad k = 1, ..., g$$

jest największa.

$p(k|\mathbf{x})$ - prawdopodobieństwo, że obserwacja $\mathbf{x}$ pochodzi z klasy k

Twierdzenie Bayesa

$$p(k|\mathbf{x}) = \frac{\pi_k p(\mathbf{x}|k)}{\sum_{r=1}^{g} \pi_r p(\mathbf{x}|r)} = \frac{\pi_k p(\mathbf{x}|k)}{p(\mathbf{x})}$$

Klasyfikator bayesowski

Zaobserwowany wektor $\mathbf{x}$ klasyfikujemy jako pochodzący z tej klasy k, dla której wartość

$$p(k|\mathbf{x}) \qquad k = 1, ..., g$$

jest największa.

$p(k|\mathbf{x})$ - prawdopodobieństwo, że obserwacja $\mathbf{x}$ pochodzi z klasy k

Twierdzenie Bayesa

$$p(k|\mathbf{x}) = \frac{\pi_k p(\mathbf{x}|k)}{\sum_{r=1}^{g} \pi_r p(\mathbf{x}|r)} = \frac{\pi_k p(\mathbf{x}|k)}{p(\mathbf{x})}$$

- $p(\mathbf{x}|k)$ - rozkład obserwacji $\mathbf{x}$ z klasy k,
- π_k - prawdopodobieństwo *a priori*, że obserwacja pochodzi z klasy k,
- $p(\mathbf{x})$ - prawdopodobieństwo pojawienia się obserwacji $\mathbf{x}$.

Dla dwóch klas ($g = 2$)

Dla dwóch klas ($g = 2$)

$$\frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \frac{\pi_2 p(\mathbf{x}|2)}{\pi_1 p(\mathbf{x}|1)}$$

Dla dwóch klas ($g = 2$)

$$\frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \frac{\pi_2 p(\mathbf{x}|2)}{\pi_1 p(\mathbf{x}|1)}$$

Załóżmy teraz, że składowe wektora $\mathbf{x} = [x^1, x^2, ..., x^p]^T$ są niezależne

Dla dwóch klas ($g = 2$)

$$\frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \frac{\pi_2 p(\mathbf{x}|2)}{\pi_1 p(\mathbf{x}|1)}$$

Załóżmy teraz, że składowe wektora $\mathbf{x} = [x^1, x^2, ..., x^p]^T$ są niezależne

$$p(\mathbf{x}|k) = \prod_{i=1}^{p} p(x^{(i)}|k)$$

Dla dwóch klas ($g = 2$)

$$\frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \frac{\pi_2 p(\mathbf{x}|2)}{\pi_1 p(\mathbf{x}|1)}$$

Załóżmy teraz, że składowe wektora $\mathbf{x} = [x^1, x^2, ..., x^p]^T$ są niezależne

$$p(\mathbf{x}|k) = \prod_{i=1}^{p} p(x^{(i)}|k)$$

Wtedy

$$\frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \frac{\pi_2 \prod_{i=1}^{p} p(x^i|2)}{\pi_1 \prod_{i=1}^{p} p(x^i|1)}$$

Po zlogarytmowaniu

Po zlogarytmowaniu

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\pi_2}{\pi_1} + \sum_{i=1}^{p} \ln \frac{p(x^i|2)}{p(x^i|1)}$$

Po zlogarytmowaniu

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\pi_2}{\pi_1} + \sum_{i=1}^{p} \ln \frac{p(x^i|2)}{p(x^i|1)}$$

Prowadzi to do następującej reguły klasyfikacyjnej:

Po zlogarytmowaniu

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\pi_2}{\pi_1} + \sum_{i=1}^{p} \ln \frac{p(x^i|2)}{p(x^i|1)}$$

Prowadzi to do następującej reguły klasyfikacyjnej:

Naiwny klasyfikator bayesowski dla $g = 2$

Jeżeli

$$\ln \frac{\pi_2}{\pi_1} + \sum_{i=1}^{p} \ln \frac{p(x^i|2)}{p(x^i|1)} > 0$$

sklasyfikuj obserwacje $\mathbf{x}$ do klasy 2, w przeciwnym razie - do klasy 1.

Cechy naiwnego klasyfikatora bayesowskiego:

- **addytywny** wpływ kolejnych składowych wektora obserwacji $\mathbf{x}$

Cechy naiwnego klasyfikatora bayesowskiego:

- **addytywny** wpływ kolejnych składowych wektora obserwacji $\mathbf{x}$
- wartość prawdopodobieństw $p(x^i|k)$ są nieznane i estymujemy je jako

Cechy naiwnego klasyfikatora bayesowskiego:

- **addytywny** wpływ kolejnych składowych wektora obserwacji $\mathbf{x}$
- wartość prawdopodobieństw $p(x^i|k)$ są nieznane i estymujemy je jako

$$\hat{p}(x^i = l|k) = \frac{n_{ik}(l)}{n_{ik}}$$

Cechy naiwnego klasyfikatora bayesowskiego:

- **addytywny** wpływ kolejnych składowych wektora obserwacji $\mathbf{x}$
- wartość prawdopodobieństw $p(x^i|k)$ są nieznane i estymujemy je jako

$$\hat{p}(x^i = l|k) = \frac{n_{ik}(l)}{n_{ik}}$$

  - n_{ik} - liczba wszystkich obserwacji w próbie uczącej z klasy k, dla których została zmierzona wartość i-tej składowej,
  - $n_{ik}(l)$ - liczba wszystkich obserwacji w próbie uczącej z klasy k, dla których została zmierzona wartość i-tej składowej, która przyjęła poziom l

Cechy naiwnego klasyfikatora bayesowskiego:

- **addytywny** wpływ kolejnych składowych wektora obserwacji $\mathbf{x}$
- wartość prawdopodobieństw $p(x^i|k)$ są nieznane i estymujemy je jako

$$\hat{p}(x^i = l|k) = \frac{n_{ik}(l)}{n_{ik}}$$

  - n_{ik} - liczba wszystkich obserwacji w próbie uczącej z klasy k, dla których została zmierzona wartość i-tej składowej,
  - $n_{ik}(l)$ - liczba wszystkich obserwacji w próbie uczącej z klasy k, dla których została zmierzona wartość i-tej składowej, która przyjęła poziom l

- podobnie $\hat{\pi}_k$ estymujemy jako $\frac{n_k}{n}$, gdzie n liczba wszystkich obserwacji, n_k - liczba obserwacji z klasy k

Kodowanie

- Wilgotność powietrza:

Kodowanie

- Wilgotność powietrza:
  - 0-30 %  (mała),
  - 31-70 %  (średnia),
  - 71-100 %  (duża),

Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:

Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),

Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:

Naiwny klasyfikator bayesowski - przykład

Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),

dzień	wilgotność	ciśnienie	klasa

Naiwny klasyfikator bayesowski - przykład

Kodowanie

- Wilgotność powietrza:
  - 0-30 %  (mała),
  - 31-70 %  (średnia),
  - 71-100 %  (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa  (niskie),
  - 990-1010 hPa  (średnie),
  - $>$ 1010 hPa  (duże),
- Klasy:
  - klasa 1  (pogoda),
  - klasa 2  (deszcz),

dzień	wilgotność	ciśnienie	klasa
1			

Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),

dzień	wilgotność	ciśnienie	klasa
1			
2			

Kodowanie

- Wilgotność powietrza:
  - 0-30 %  (mała),
  - 31-70 %  (średnia),
  - 71-100 %  (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa  (niskie),
  - 990-1010 hPa  (średnie),
  - $>$ 1010 hPa  (duże),
- Klasy:
  - klasa 1  (pogoda),
  - klasa 2  (deszcz),

dzień	wilgotność	ciśnienie	klasa
1			
2			
3			

Kodowanie

- Wilgotność powietrza:
  - 0-30 %  (mała),
  - 31-70 %  (średnia),
  - 71-100 %  (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa  (niskie),
  - 990-1010 hPa  (średnie),
  - $>$ 1010 hPa  (duże),
- Klasy:
  - klasa 1  (pogoda),
  - klasa 2  (deszcz),

dzień	wilgotność	ciśnienie	klasa
1			
2			
3			
4			

Naiwny klasyfikator bayesowski - przykład

Kodowanie

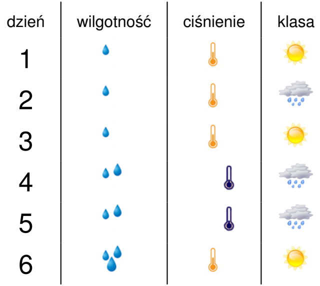
- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),

dzień	wilgotność	ciśnienie	klasa
1			
2			
3			
4			
5			

Naiwny klasyfikator bayesowski - przykład

Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),



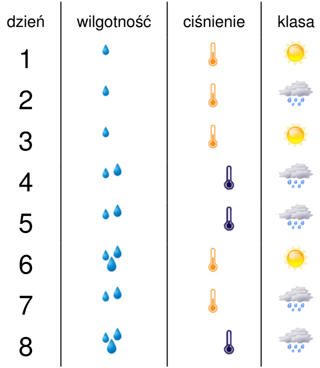
Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),

Naiwny klasyfikator bayesowski - przykład

Kodowanie

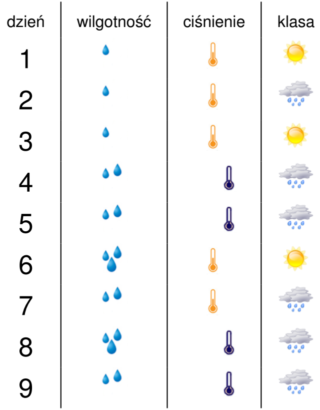
- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),



Naiwny klasyfikator bayesowski - przykład

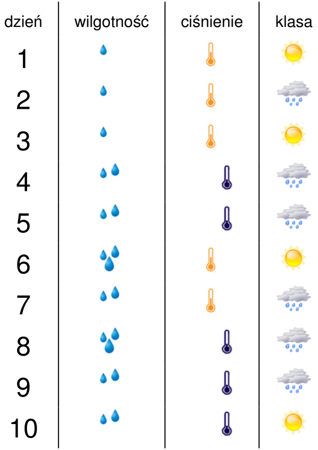
Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),



Kodowanie

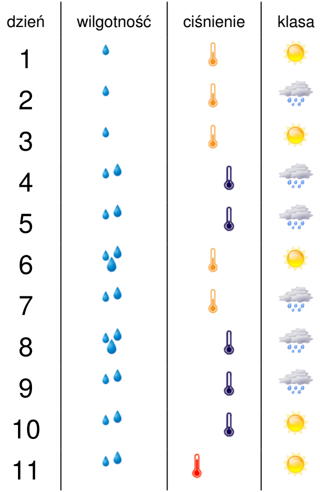
- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),



Naiwny klasyfikator bayesowski - przykład

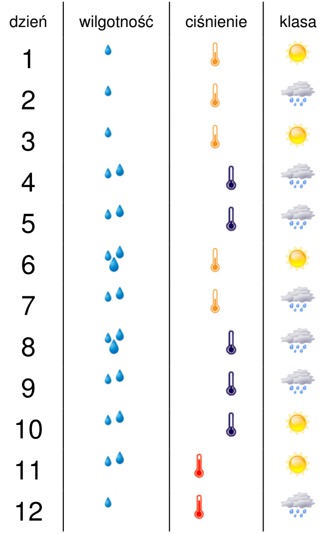
Kodowanie

- Wilgotność powietrza:
  - 0-30 % (mała),
  - 31-70 % (średnia),
  - 71-100 % (duża),
- Ciśnienie atmosferyczne:
  - $<$ 990 hPa (niskie),
  - 990-1010 hPa (średnie),
  - $>$ 1010 hPa (duże),
- Klasy:
  - klasa 1 (pogoda),
  - klasa 2 (deszcz),



Kodowanie

- Wilgotność powietrza:
    - 0-30 % (mała),
    - 31-70 % (średnia),
    - 71-100 % (duża),
- Ciśnienie atmosferyczne:
    - $<$ 990 hPa (niskie),
    - 990-1010 hPa (średnie),
    - $>$ 1010 hPa (duże),
- Klasy:
    - klasa 1 (pogoda),
    - klasa 2 (deszcz),



Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = \mathit{niska}, p = \mathit{wysokie}]^T$. Do której klasy należy ją sklasyfikować?

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = \mathit{niska}, p = \mathit{wysokie}]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} =$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
	☀	🌧	☀	🌧
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
	☀	🌧	☀	🌧
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.

Pojawia się nowa obserwacja $\mathbf{x} = [H = \textit{niska}, p = \textit{wysokie}]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} =$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\frac{7}{12}}{\frac{5}{12}}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\frac{7}{12}}{\frac{5}{12}} + \ln \frac{\frac{2}{7}}{\frac{2}{5}}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\frac{7}{12}}{\frac{5}{12}} + \ln \frac{\frac{2}{7}}{\frac{2}{5}} + \ln \frac{\frac{1}{7}}{\frac{1}{5}}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\frac{7}{12}}{\frac{5}{12}} + \ln \frac{\frac{2}{7}}{\frac{2}{5}} + \ln \frac{\frac{1}{7}}{\frac{1}{5}} = \ln \frac{7}{5}\frac{2}{7}\frac{5}{2}\frac{1}{7}\frac{5}{1}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\frac{7}{12}}{\frac{5}{12}} + \ln \frac{\frac{2}{7}}{\frac{2}{5}} + \ln \frac{\frac{1}{7}}{\frac{1}{5}} = \ln \frac{7}{5}\frac{2}{7}\frac{5}{2}\frac{1}{7}\frac{5}{1} = \ln \frac{5}{7}$$

Naiwny klasyfikator bayesowski - przykład

	wilgotność		ciśnienie	
				
niskie	2	2	1	4
średnie	2	4	3	2
wysokie	1	1	1	1
Σ	5	7	5	7

	wilgotność		ciśnienie	
				
niskie	2/5	2/7	1/5	4/7
średnie	2/5	4/7	3/5	2/7
wysokie	1/5	1/7	1/5	1/7
Σ	1	1	1	1

Estymacja prawdopodobieństw a priori: $\hat{\pi}_1 = \frac{5}{12}$, $\hat{\pi}_2 = \frac{7}{12}$.
Pojawia się nowa obserwacja $\mathbf{x} = [H = niska, p = wysokie]^T$. Do której klasy należy ją sklasyfikować?

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\hat{\pi}_2}{\hat{\pi}_1} + \ln \frac{\hat{p}(H = niska|2)}{\hat{p}(H = niska|1)} + \ln \frac{\hat{p}(p = wysokie|2)}{\hat{p}(p = wysokie|1)}$$

$$\ln \frac{p(2|\mathbf{x})}{p(1|\mathbf{x})} = \ln \frac{\frac{7}{12}}{\frac{5}{12}} + \ln \frac{\frac{2}{7}}{\frac{2}{5}} + \ln \frac{\frac{1}{7}}{\frac{1}{5}} = \ln \frac{7}{5}\frac{2}{7}\frac{5}{2}\frac{1}{7}\frac{5}{1} = \ln \frac{5}{7} < 0$$ 

Rozważmy k klas i obserwacje $\mathbf{x}$ pochodzące z wielowymiarowego rozkładu Gaussa $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S})$, czyli:

Rozważmy k klas i obserwacje $\mathbf{x}$ pochodzące z wielowymiarowego rozkładu Gaussa $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S})$, czyli:

- $p(\mathbf{x}|k) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{S}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}_k)\right]$

Rozważmy k klas i obserwacje $\mathbf{x}$ pochodzące z wielowymiarowego rozkładu Gaussa $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S})$, czyli:

- $p(\mathbf{x}|k) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{S}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}_k)\right]$
- każda klasa k jest scharakteryzowana wartością oczekiwaną $\mathbf{m}_k$

Rozważmy k klas i obserwacje $\mathbf{x}$ pochodzące z wielowymiarowego rozkładu Gaussa $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S})$, czyli:

- $p(\mathbf{x}|k) = \frac{1}{(2\pi)^{\frac{p}{2}}|\mathbf{S}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \mathbf{S}^{-1}(\mathbf{x} - \mathbf{m}_k)\right]$
- każda klasa k jest scharakteryzowana wartością oczekiwaną $\mathbf{m}_k$
- wszystkie klasy mają taką samą macierz kowariancji $\mathbf{S}$

Rozważmy k klas i obserwacje $\mathbf{x}$ pochodzące z wielowymiarowego rozkładu Gaussa $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S})$, czyli:

- $p(\mathbf{x}|k) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{S}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}_k)\right]$
- każda klasa k jest scharakteryzowana wartością oczekiwaną $\mathbf{m}_k$
- wszystkie klasy mają taką samą macierz kowariancji $\mathbf{S}$

Korzystając z reguły bayesowskiej można określić **funkcję dyskryminacyjną** dla $g = 2$

Rozważmy k klas i obserwacje $\mathbf{x}$ pochodzące z wielowymiarowego rozkładu Gaussa $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S})$, czyli:

- $p(\mathbf{x}|k) = \frac{1}{(2\pi)^{\frac{p}{2}}|\mathbf{S}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_k)^T\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m}_k)\right]$
- każda klasa k jest scharakteryzowana wartością oczekiwaną $\mathbf{m}_k$
- wszystkie klasy mają taką samą macierz kowariancji $\mathbf{S}$

Korzystając z reguły bayesowskiej można określić **funkcję dyskryminacyjną** dla $g = 2$

$$\delta_{12}(\mathbf{x}) = \ln\frac{p(1|\mathbf{x})}{p(2|\mathbf{x})} = \ln\frac{\pi_1}{\pi_2} + \ln\frac{p(\mathbf{x}|1)}{p(\mathbf{x}|2)}$$

czyli

Rozważmy k klas i obserwacje $\mathbf{x}$ pochodzące z wielowymiarowego rozkładu Gaussa $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S})$, czyli:

- $p(\mathbf{x}|k) = \frac{1}{(2\pi)^{\frac{p}{2}}|\mathbf{S}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}_k)\right]$
- każda klasa k jest scharakteryzowana wartością oczekiwaną $\mathbf{m}_k$
- wszystkie klasy mają taką samą macierz kowariancji $\mathbf{S}$

Korzystając z reguły bayesowskiej można określić **funkcję dyskryminacyjną** dla $g = 2$

$$\delta_{12}(\mathbf{x}) = \ln \frac{p(1|\mathbf{x})}{p(2|\mathbf{x})} = \ln \frac{\pi_1}{\pi_2} + \ln \frac{p(\mathbf{x}|1)}{p(\mathbf{x}|2)}$$

czyli

$$\delta_{12}(\mathbf{x}) = \ln \frac{\pi_1}{\pi_2} + \frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{S}^{-1} (\mathbf{m}_1 + \mathbf{m}_2) + (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{S}^{-1} \mathbf{x}$$

Jeżeli $\delta_{12}(\mathbf{x}) > 0$, to obserwacja **x** zostaje sklasyfikowana jako pochodząca z klasy 1, jeśli $\delta_{12}(\mathbf{x}) < 0$, to **x** jest z klasy 2.

Jeżeli $\delta_{12}(\mathbf{x}) > 0$, to obserwacja $\mathbf{x}$ zostaje sklasyfikowana jako pochodząca z klasy 1, jeśli $\delta_{12}(\mathbf{x}) < 0$, to $\mathbf{x}$ jest z klasy 2.

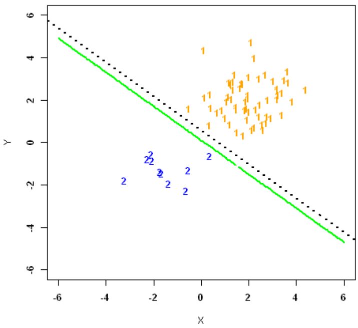
Równanie hiperpłaszczyzny dyskryminacyjnej

$$\ln \frac{\pi_1}{\pi_2} + \frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{S}^{-1}(\mathbf{m}_1 + \mathbf{m}_2) + (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{S}^{-1}\mathbf{x} = 0$$

Jeżeli $\delta_{12}(\mathbf{x}) > 0$, to obserwacja $\mathbf{x}$ zostaje sklasyfikowana jako pochodząca z klasy 1, jeśli $\delta_{12}(\mathbf{x}) < 0$, to $\mathbf{x}$ jest z klasy 2.

Równanie hiperpłaszczyzny dyskryminacyjnej

$$\ln \frac{\pi_1}{\pi_2} + \frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{S}^{-1}(\mathbf{m}_1 + \mathbf{m}_2) + (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{S}^{-1}\mathbf{x} = 0$$



Próba losowa z $n_1 = 60$, $n_2 = 10$.

$\mathbf{S} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\mathbf{m}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$\mathbf{m}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

W przypadku gdy $\mathbf{W} = \mathbf{S}$ oraz $\pi_1 = \pi_2$ otrzymujemy dyskryminacje Fishera (kreskowana linia).

Jeżeli $g > 2$ można podać równania hiperpłaszczyzn rozdzielających **pary klas**.

Jeżeli $g > 2$ można podać równania hiperpłaszczyzn rozdzielających **pary klas**.

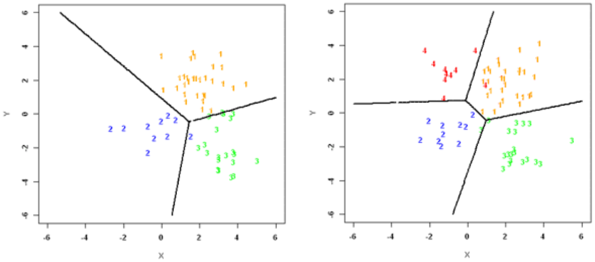
Równanie hiperpłaszczyzny dyskryminacyjnej dla LDA

$$\delta_{kl}(\mathbf{x}) = \ln \frac{\pi_1}{\pi_2} + \frac{1}{2}(\mathbf{m}_k - \mathbf{m}_l)^T \mathbf{S}^{-1}(\mathbf{m}_k + \mathbf{m}_l) + (\mathbf{m}_k - \mathbf{m}_l)^T \mathbf{S}^{-1}\mathbf{x} = 0$$

Jeżeli $g > 2$ można podać równania hiperpłaszczyzn rozdzielających **pary klas**.

Równanie hiperpłaszczyzny dyskryminacyjnej dla LDA

$$\delta_{kl}(\mathbf{x}) = \ln\frac{\pi_1}{\pi_2} + \frac{1}{2}(\mathbf{m}_k - \mathbf{m}_l)^T\mathbf{S}^{-1}(\mathbf{m}_k + \mathbf{m}_l) + (\mathbf{m}_k - \mathbf{m}_l)^T\mathbf{S}^{-1}\mathbf{x} = 0$$



W przypadku metody QDA (*quadratic discriminant analysis*) zakładamy różne macierze kowariancji $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S}_k)$.

W przypadku metody QDA (*quadratic discriminant analysis*) zakładamy różne macierze kowariancji $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S}_k)$.

Równanie hiperpłaszczyzny dyskryminacyjnej dla QDA

$$\delta_{kl}(\mathbf{x}) = \ln \frac{\pi_1}{\pi_2} + \frac{1}{2} \ln \frac{|\mathbf{S}_k|}{|\mathbf{S}_l|} + \mathbf{x}^T(\mathbf{S}_l^{-1}\mathbf{m}_l - \mathbf{S}_k^{-1}\mathbf{m}_k) - \frac{1}{2}\mathbf{x}^T(\mathbf{S}_l^{-1} - \mathbf{S}_k^{-1})\mathbf{x} - \frac{1}{2}\mathbf{m}_l^T\mathbf{S}_l^{-1}\mathbf{m}_l + \frac{1}{2}\mathbf{m}_k^T\mathbf{S}_k^{-1}\mathbf{m}_k = 0$$

W przypadku metody QDA (*quadratic discriminant analysis*) zakładamy różne macierze kowariancji $p(\mathbf{x}|k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S}_k)$.

Równanie hiperpłaszczyzny dyskryminacyjnej dla QDA

$$\delta_{kl}(\mathbf{x}) = \ln \frac{\pi_1}{\pi_2} + \frac{1}{2} \ln \frac{|\mathbf{S}_k|}{|\mathbf{S}_l|} + \mathbf{x}^T(\mathbf{S}_l^{-1}\mathbf{m}_l - \mathbf{S}_k^{-1}\mathbf{m}_k) - \frac{1}{2}\mathbf{x}^T(\mathbf{S}_l^{-1} - \mathbf{S}_k^{-1})\mathbf{x} - \frac{1}{2}\mathbf{m}_l^T\mathbf{S}_l^{-1}\mathbf{m}_l + \frac{1}{2}\mathbf{m}_k^T\mathbf{S}_k^{-1}\mathbf{m}_k = 0$$