SPMD – TEST 1 (29 listopada 2020)

Dane są próbki dwóch klas A i B. Pierwsza klasa składa się z czterech próbkę:

$$x_1^A = \begin{bmatrix}1\\1\end{bmatrix} \quad x_2^A = \begin{bmatrix}2\\2\end{bmatrix} \quad x_3^A = \begin{bmatrix}0\\3\end{bmatrix} \quad x_4^A = \begin{bmatrix}3\\0\end{bmatrix}$$

O klasie B wiadomo, że wektor średni próbek tej klasy i jej macierz kowariancji to:

$$\underline{\mu}^B = \begin{bmatrix}-2\\-2\end{bmatrix} \qquad C^B = \begin{bmatrix}3 & 1\\1 & 1\end{bmatrix}$$

Która jednowymiarowa przestrzeń cech będzie lepsza z punktu widzenia możliwość poprawnej klasyfikacji próbek? Jaki będzie wynik klasyfikacji próbki $\underline{x} = \begin{bmatrix}0\\0\end{bmatrix}$ w tej przestrzeni?

$$\underline{x}^A = \begin{bmatrix}1 & 2 & 0 & 3\\1 & 2 & 3 & 0\end{bmatrix}$$

$$\underline{\mu}^A = \frac{1}{4} * \begin{bmatrix}1+2+0+3\\1+2+3+0\end{bmatrix} = \begin{bmatrix}\frac{1+2+0+3}{4}\\\frac{1+2+3+0}{4}\end{bmatrix} = \begin{bmatrix}\frac{6}{4}\\\frac{6}{4}\end{bmatrix} = \begin{bmatrix}1{,}5\\1{,}5\end{bmatrix}$$

Macierz odchylenia (pomocnicza)

$$\left(\underline{x}^A - \underline{M}^A\right) = \left(\begin{bmatrix}1 & 2 & 0 & 3\\1 & 2 & 3 & 0\end{bmatrix} - \begin{bmatrix}1{,}5 & 1{,}5 & 1{,}5 & 1{,}5\\1{,}5 & 1{,}5 & 1{,}5 & 1{,}5\end{bmatrix}\right) = \begin{bmatrix}-0{,}5 & 0{,}5 & -1{,}5 & 1{,}5\\-0{,}5 & 0{,}5 & 1{,}5 & -1{,}5\end{bmatrix}$$

Macierz kowariancji

$$C^A = \frac{1}{4} * \left(\underline{x}^A - \underline{\mu}^A\right) * \left(\underline{x}^A - \underline{\mu}^A\right)^T = \frac{1}{4} * \begin{bmatrix}-0{,}5 & 0{,}5 & -1{,}5 & 1{,}5\\-0{,}5 & 0{,}5 & 1{,}5 & -1{,}5\end{bmatrix} * \begin{bmatrix}-0{,}5 & -0{,}5\\0{,}5 & 0{,}5\\-1{,}5 & 1{,}5\\1{,}5 & -1{,}5\end{bmatrix} =$$

$$= \frac{1}{4} * \begin{bmatrix}5 & -4\\-4 & 5\end{bmatrix} = \begin{bmatrix}1{,}25 & -1\\-1 & 1{,}25\end{bmatrix}$$

Selekcja cech:

$$F = \frac{|\mu^A - \mu^B|}{\delta^A + \delta^B}$$

$$F_1 = \frac{|{\mu_1}^A - {\mu_1}^B|}{\delta_1^A + \delta_1^B} = \frac{|1{,}5 - (-2)|}{\sqrt{1{,}25} + \sqrt{3}} \approx 1{,}228$$

$$F_2 = \frac{|{\mu_2}^A - {\mu_2}^B|}{\delta_2^A + \delta_2^B} = \frac{|1{,}5 - (-2)|}{\sqrt{1{,}25} + \sqrt{1}} \approx 1{,}652$$

Z punktu widzenia możliwość poprawnej klasyfikacji próbek lepszą cechą jest cecha 2 ponieważ $F_2 > F_1$

$$d\left(x, \{\mu_2^A, \delta_2^A\}\right) = \frac{\left|x - \mu_2^A\right|}{\delta_2^A} = \frac{|0 - 1{,}5|}{\sqrt{1{,}25}} \approx 1{,}342$$

$$d(x, \{\mu_2^B, \delta_2^B\}) = \frac{\left|x - \mu_2^B\right|}{\delta_2^B} = \frac{|0 - (-2)|}{\sqrt{1}} = 2$$

Z obliczonych odległości na podstawie najlepszej cechy wynika, że próbka $\underline{x}$ należy do klasy A.

KOL2 – bez treści

Klasa A:

$$\underline{\boldsymbol{x}}^{\boldsymbol{A}} = \begin{bmatrix} \mathbf{2} & \mathbf{4} & \mathbf{6} & \mathbf{4} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{2} \end{bmatrix}$$

$$\underline{\boldsymbol{\mu}}^{\boldsymbol{A}} = \frac{1}{4} * \begin{bmatrix} 2+4+6+4 \\ 1+0+1+2 \end{bmatrix} = \begin{bmatrix} \frac{2+4+5+4}{4} \\ \frac{1+0+1+2}{4} \end{bmatrix} = \begin{bmatrix} \frac{16}{4} \\ \frac{4}{4} \end{bmatrix} = \begin{bmatrix} \mathbf{4} \\ \mathbf{1} \end{bmatrix}$$

Macierz odchylenia (pomocnicza):

$$\left(\underline{x}^A - \underline{M}^A\right) = \left(\begin{bmatrix} 2 & 4 & 6 & 4 \\ 1 & 0 & 1 & 2 \end{bmatrix} - \begin{bmatrix} 4 & 4 & 4 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix}\right) = \begin{bmatrix} -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

Macierz kowariancji:

$$\boldsymbol{C}^{\boldsymbol{A}} = \frac{1}{4}\left(\underline{x}^A - \underline{\mu}^A\right)\left(\underline{x}^A - \underline{\mu}^A\right)^T = \frac{1}{4} * \begin{bmatrix} -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} -2 & 0 \\ 0 & -1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{4} * \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0,5} \end{bmatrix}$$

Klasa B:

$$\underline{\boldsymbol{\mu}}^{\boldsymbol{B}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad \boldsymbol{C}^{\boldsymbol{B}} = \begin{bmatrix} \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{2} \end{bmatrix}$$

Selekcja cech (współczynnik Fishera liniowej dyskryminacji):

$$F_1 = \frac{|\mu_1{}^A - \mu_1{}^B|}{\delta_1^A + \delta_1^B} = \frac{|4-0|}{\sqrt{2}+\sqrt{2}} = \sqrt{2} \approx 1{,}414$$

$$F_2 = \frac{|\mu_2{}^A - \mu_2{}^B|}{\delta_2^A + \delta_2^B} = \frac{|1-0|}{\sqrt{0{,}5}+\sqrt{2}} = \frac{\sqrt{2}}{3} \approx 0{,}471$$

Z punktu widzenia możliwość poprawnej klasyfikacji próbek lepszą cechą jest cecha 1 ponieważ F_1 > F_2

Odległość Mahalanobisa dla wybranej cechy (punkt do klasyfikacji $\underline{x} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$):

$$d(x_1, \{\mu_1^A, \delta_1^A\}) = \frac{\left|x_1 - \mu_1^A\right|}{\delta_1^A} = \frac{|4-4|}{\sqrt{2}} = 0$$

$$d(x_1, \{\mu_1^B, \delta_1^B\}) = \frac{\left|x_1 - \mu_1^B\right|}{\delta_1^B} = \frac{|4-0|}{\sqrt{2}} = \frac{4}{\sqrt{2}} \approx 2{,}828$$

Z obliczonych odległości na podstawie najlepszej cechy wynika, że próbka $\underline{x}$ należy do klasy A.

KOL3

Dane są parametry statystyczne dwóch klas:

$$\underline{\mu}_A = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \qquad \underline{\mu}_B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$C_A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad C_B = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Do której klasy należą punkty $\underline{x} = \begin{bmatrix} 0,5 \\ 0 \end{bmatrix}$?

Odległość Mahalanobisa:

$$d\left(\underline{x}, \left\{\underline{\mu}_A, C^A\right\}\right) = \left(\underline{x} - \underline{\mu}_A\right)^T * C_A^{-1} * \left(\underline{x} - \underline{\mu}_A\right) = \left(\begin{bmatrix} 0,5 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ -1 \end{bmatrix}\right)^T * \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} * \left(\begin{bmatrix} 0,5 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ -1 \end{bmatrix}\right) =$$

$$= \frac{1}{3} * [1,5 \quad 1] * \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} * \begin{bmatrix} 1,5 \\ 1 \end{bmatrix} = \frac{1}{3} * [2 \quad 0,5] * \begin{bmatrix} 1,5 \\ 1 \end{bmatrix} = \frac{1}{3} * \frac{7}{2} = \frac{7}{6} \approx 1,167$$

$$d\left(\underline{x}, \left\{\underline{\mu}_B, C_B\right\}\right) = \left(\underline{x} - \underline{\mu}_B\right)^T * C_B^{-1} * \left(\underline{x} - \underline{\mu}_B\right) = \left(\begin{bmatrix} 0,5 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)^T * \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}^{-1} * \left(\begin{bmatrix} 0,5 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) =$$

$$= \frac{1}{3} * [-0,5 \quad -1] * \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} * \begin{bmatrix} -0,5 \\ -1 \end{bmatrix} = \frac{1}{3} * [-2 \quad -2,5] * \begin{bmatrix} -0,5 \\ -1 \end{bmatrix} = \frac{1}{3} * \frac{7}{2} = \frac{7}{6} \approx 1,167$$

Odległość Euklidesa:

$$d_A = \sqrt{(x_1 - \mu_{A1})^2 + (x_2 - \mu_{A2})^2} = \sqrt{\left(0,5 - (-1)\right)^2 + \left(0 - (-1)\right)^2} = \sqrt{2,25 + 1} = \sqrt{3,25} \approx 1,802$$

$$d_B = \sqrt{(x_1 - \mu_{B1})^2 + (x_2 - \mu_{B2})^2} = \sqrt{(0,5 - 1)^2 + (0 - 1)^2} = \sqrt{0,25 + 1} = \sqrt{1,25} \approx 1,118$$

Współczynnik Fishera:

$$F_1 =$$

??

KOL4

Dane są parametry statystyczne dwóch klas:

$$\underline{\mu}_A = \begin{bmatrix} 0 \\ -1 \\ -8 \end{bmatrix} \qquad \underline{\mu}_B = \begin{bmatrix} 5 \\ 1 \\ -1 \end{bmatrix}$$

$$C_A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 4 \end{bmatrix} \quad C_B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Która jednowymiarowa przestrzeń cech będzie najlepsza z punktu widzenia możliwości przeprowadzenia liniowej separacji próbek tych klas?

Współczynnik Fishera 1D

$$F_1 = \frac{|\mu_{A1} - \mu_{B1}|}{\delta_{A1} + \delta_{B1}} = \frac{|0-5|}{\sqrt{1} + \sqrt{2}} \approx 2{,}071$$

$$F_2 = \frac{|\mu_{A2} - \mu_{B2}|}{\delta_{A2} + \delta_{B2}} = \frac{|-1-1|}{\sqrt{2} + \sqrt{1}} \approx 0{,}828$$

$$F_3 = \frac{|\mu_{A3} - \mu_{B3}|}{\delta_{A3} + \delta_{B3}} = \frac{|-8-(-1)|}{\sqrt{4} + \sqrt{2}} = \sqrt{2} \approx 2{,}05$$

Najlepszą cechą do dyskryminacji jest cecha 1 ponieważ F1 >F3> F2

Jak wyznaczyć najlepszą przestrzeń dwuwymiarową?

W celu wyznaczenia najlepszej przestrzeni dwuwymiarowej należy wyznaczyć wartość współczynnika Fishera dla każdej możliwej kombinacji par współczynników.

Współczynnik Fishera 2D

$$F_{12} = \frac{\sqrt{(\mu_{A1} - \mu_{B1})^2 + (\mu_{A2} - \mu_{B2})^2}}{\delta_{A11} + \delta_{A22} + \delta_{B11} + \delta_{B22}} = \frac{\sqrt{(0-5)^2 + \left((-1)-1\right)^2}}{1+2+2+1} = \frac{\sqrt{29}}{6} \approx 0{,}898$$

$$F_{13} = \frac{\sqrt{(\mu_{A1} - \mu_{B1})^2 + (\mu_{A3} - \mu_{B3})^2}}{\delta_{A11} + \delta_{A33} + \delta_{B11} + \delta_{B33}} = \frac{\sqrt{(0-5)^2 + \left((-8)-(-1)\right)^2}}{1+4+2+2} = \frac{\sqrt{74}}{9} \approx 0{,}956$$

$$F_{23} = \frac{\sqrt{(\mu_{A2} - \mu_{B2})^2 + (\mu_{A3} - \mu_{B3})^2}}{\delta_{A22} + \delta_{A33} + \delta_{B22} + \delta_{B33}} = \frac{\sqrt{\left((-1)-1\right)^2 + \left((-8)-(-1)\right)^2}}{2+4+1+2} = \frac{\sqrt{53}}{9} \approx 0{,}808$$

Najlepszą parą cech do dyskryminacji są cechy 1 i 3 ponieważ F13 >F12> F23