

VI9 GRUPIRANJE

ZADACI ZA UČENJE:

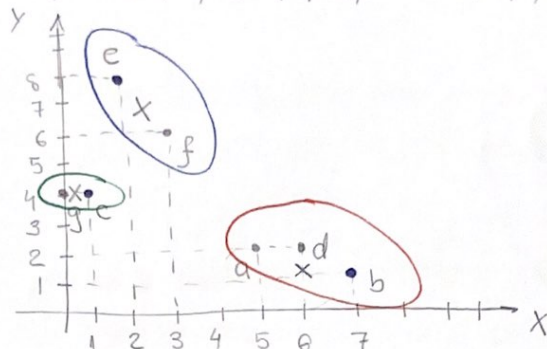
(ZAD 2.) $D = \{a = (5, 2), b = (7, 1), c = (1, 4), d = (6, 2), e = (2, 8), f = (3, 6), g = (0, 4)\}$

a) $K = 3$

$\mu_1 = b = (7, 1)$

$\mu_2 = c = (1, 4)$

$\mu_3 = e = (2, 8)$



→ za X_1 :

$$\|X_1 - \mu_1\| = \sqrt{(5-7)^2 + (2-1)^2} = 2,2361$$

$$\|X_1 - \mu_2\| = 4,4721$$

$$\|X_1 - \mu_3\| = 7,6158$$

$$\left. \begin{array}{l} \|X_1 - \mu_1\| = 2,2361 \\ \|X_1 - \mu_2\| = 4,4721 \\ \|X_1 - \mu_3\| = 7,6158 \end{array} \right\} \begin{array}{l} k = \arg \min_j \|X_1 - \mu_j\| = 1 \\ b_1^1 = 1 \\ b_2^1 = 0 \\ b_3^1 = 0 \end{array}$$

→ za X_2 :

$$\|X_2 = b = \mu_1\| \Rightarrow k = 1$$

$$b_1^2 = 1; b_2^2 = 0; b_3^2 = 0$$

→ za X_3 :

$$\|X_3 = c = \mu_2\| \Rightarrow k = 2$$

$$b_1^3 = 0; b_2^3 = 1; b_3^3 = 0$$

→ za X_4 :

$$\|X_4 - \mu_1\| = 1,414$$

$$\|X_4 - \mu_2\| = 5,38516$$

$$\|X_4 - \mu_3\| = 7,2111$$

$$\left. \begin{array}{l} \|X_4 - \mu_1\| = 1,414 \\ \|X_4 - \mu_2\| = 5,38516 \\ \|X_4 - \mu_3\| = 7,2111 \end{array} \right\} \begin{array}{l} k = 1 \\ b_1^4 = 1; b_2^4 = b_3^4 = 0 \end{array}$$

→ za X_5 :

$$X_5 = e = \mu_3 \rightarrow k = 3$$

$$b_1^5 = 0; b_2^5 = 0; b_3^5 = 1$$

→ za X_6 :

$$\left. \begin{aligned} \|X_6 - \mu_1\| &= 6,4 \\ \|X_6 - \mu_2\| &= 2,828 \\ \|X_6 - \mu_3\| &= 2,2361 \end{aligned} \right\} \begin{aligned} k &= 3 \\ b_1^6 &= b_2^6 = 0; b_3^6 = 1 \end{aligned}$$

→ za X_7 :

$$\left. \begin{aligned} \|X_7 - \mu_1\| &= 7,62 \\ \|X_7 - \mu_2\| &= 1 \\ \|X_7 - \mu_3\| &= 4,47 \end{aligned} \right\} \begin{aligned} k &= 2 \\ b_1^7 &= b_3^7 = 0; b_2^7 = 1 \end{aligned}$$

→ NOVI CENTROIDI:

$$\mu_1 = \frac{\sum_{i=1}^N b_i^{(1)} X_i}{\sum_{i=1}^N b_i^{(1)}} = \frac{X_1 + X_2 + X_4}{3} = \left(\frac{5+7+6}{3}, \frac{2+1+2}{3} \right)$$

$$\mu_1 = (6, 1,667)$$

$$\mu_2 = \left(\frac{1+9}{2}, \frac{4+4}{2} \right) = (5, 4)$$

$$\mu_3 = \left(\frac{2+3}{2}, \frac{6+8}{2} \right) = (2,5, 7)$$

b) Algoritam k-medoida vrlo je sličan algoritmu K-sredina, samo što radi s općenitom mjerom sličnosti / različitosti što otvara mogućnost da primjeri budu iz NEVEKTORSKEG prostora.

$$\left. \begin{aligned} \mu_1 &= b \\ \mu_2 &= c \\ \mu_3 &= e \end{aligned} \right\} \begin{aligned} &\text{isto grupiranje kao i u a) podzadatku:} \\ K=1 &: a, b, d \\ K=2 &: c, g \\ K=3 &: e, f \end{aligned}$$

MEDOIDI

→ NOVI ~~CENTROIDI~~ MEDOIDI: (razlika naspram K-sredina)

(*) pri određivanju novih medoida gleda se mjera sličnosti kandidata i ostalih $N-K$ primjera, naj zbroj koji je najmanji → NOVI MEDOID

$$\mu_1 = a$$

$$\mu_2 = g$$

$$\mu_3 = f$$

c) Računalna složenost:

(i) K-sredina - $O(NKT)$

(ii) K-medoida - $O(N^2KT)$

d) K-medoida:

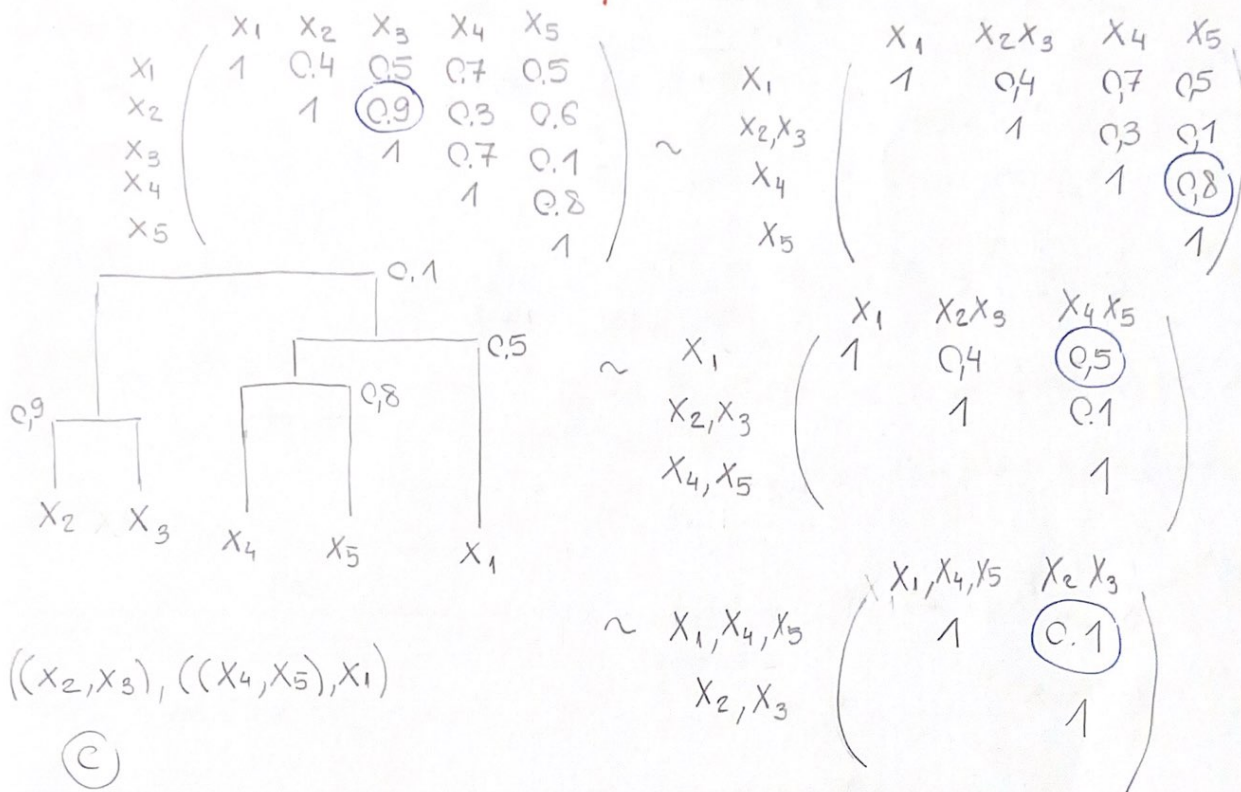
→ PREDNOSTI: brže izvršavanje, primjeri ne moraju biti iz Euklidskog prostora, robusan algoritam, neosjetljiv na stršće vrijednosti

→ NEDOSTACI: visoka vremenska složenost (nelinearna), nepogodan za klasificiranje proizvoljno oblikovanih (nesferičnih) grupa objekata

V20 GRUPIRANJE II.

Zadaci s ispita:

(8). HAC, $N=5$, potpuno povezivanje



V21 VREDNOVANJE MODELA

Zadaci za učenje:

(ZAD 1.)

$$\{y^{(i)}, h(\vec{x})^{(i)}\} = \{(\textcircled{1}, \textcircled{1}), (0, 2), (\textcircled{2}, \textcircled{2}), (1, 2), (\textcircled{1}, \textcircled{1}), (\textcircled{0}, \textcircled{0}), (\textcircled{1}, \textcircled{1}), (2, 1), (0, 1), (2, 0), (2, 1)\}$$

Confusion matrix:

	pred 0	pred 1	pred 2
true 0	0	1	2
true 1	4	0	3
true 2	2	1	2

$$Acc = \frac{1}{3} \cdot \frac{21}{11} = 63.64\%$$

$$a) Acc^M = Acc^U = \frac{1}{K} \sum_{j=1}^K Acc_j$$

$$Acc_0 = \frac{TP_0 + TN_0}{N} = \frac{1 + 7}{11} = \frac{8}{11}$$

$$Acc_1 = \frac{3 + 3}{11} = \frac{6}{11}$$

$$Acc_2 = \frac{1 + 6}{11} = \frac{7}{11}$$

b) MAKRO-prosjek :

$$P^M = \frac{1}{K} \sum_j P_j \Rightarrow P_0 = \frac{TP_0}{TP_0 + FP_0} = \frac{1}{1+1} = \frac{1}{2}$$

$$P_M = \frac{1}{3} \cdot \frac{4}{3} = 44,44\%$$

$$P_1 = \frac{3}{3+3} = \frac{3}{6} = \frac{1}{2}$$

$$P_2 = \frac{1}{1+2} = \frac{1}{3}$$

$$R^M = \frac{1}{K} \sum_{j=1}^K R_j \Rightarrow R_0 = \frac{TP_0}{TP_0 + FN_0} = \frac{1}{1+2} = \frac{1}{3}$$

$$R^M = 44,44\%$$

$$R_1 = \frac{3}{3+1} = \frac{3}{4}$$

$$R_2 = \frac{1}{1+3} = \frac{1}{4}$$

$$F_1^M = \frac{1}{K} \sum_{j=1}^K F_{1j} \Rightarrow F_{1,0} = \frac{2P_0R_0}{P_0 + R_0} = 0,4$$

$$F_1^M = 42,86\%$$

$$F_{1,1} = 0,6$$

$$F_{1,2} = \frac{2}{7}$$

MIKRO-prosjek :

$$P^M - R^M = F_1^M = \frac{TP}{TP + FP} = \frac{5}{11} = 45,45\%$$

$$TP = 5$$

$$FP = 6$$

Zadaci s ispita:

(2) MLR vs. RAND (nasumični klasifikator), $K=3$

	<u>true</u>		
	$y=1$	$y=2$	$y=3$
$y=1$	30	18	3
$y=2$	11	25	2
$y=3$	4	2	5

pred.

$$P(y=1)_{\text{RAND}} = \frac{30+11+4}{100} = 0,45\%$$

$$P(y=2)_{\text{RAND}} = 0,45\%$$

$$P(y=3)_{\text{RAND}} = 0,1\%$$

$$F_1^W(\text{MLR}) = ?$$

$$F_1^W = P^W = \frac{TP}{TP+FP}$$

$$\Rightarrow TP = TP_1 + TP_2 + TP_3 = 30 + 25 + 5 = 60$$

$$FP = FP_1 + FP_2 + FP_3 = 18 + 3 + 11 + 2 + 6 = 40$$

$$F_1^W(\text{MLR}) = \frac{60}{100} = 0,6\%$$

RAND KLASIFIKATOR:

(i) $y_{\text{pred}} = 1$ uz $y_{\text{true}} = 1$: $0,45 \cdot 0,45 = 0,2025\%$ (TP)

$y_{\text{pred}} = 1$ uz $y_{\text{true}} \neq 1$: $0,45 \cdot 0,55 = 0,2475\%$ (FP)

$y_{\text{pred}} \neq 1$ uz $y_{\text{true}} = 1$: $0,55 \cdot 0,45 = 0,2475\%$ (FN)

$y_{\text{pred}} \neq 1$ uz $y_{\text{true}} \neq 1$: $0,55 \cdot 0,55 = 0,3025\%$ (TN)

$y_{\text{true}} = 1$	$y_{\text{true}} \neq 1$	
0,2025	0,2475	$y_{\text{pred}} = 1$
0,2475	0,3025	$y_{\text{pred}} \neq 1$

(ii) $y_{\text{true}} = 2$ $y_{\text{true}} \neq 2$

$y_{\text{pred}} = 2$	0,2025	0,2475
$y_{\text{pred}} \neq 2$	0,2475	0,3025

(iii) $y_{\text{true}} = 3$ $y_{\text{true}} \neq 3$

$y_{\text{pred}} = 3$	0,01	0,09
$y_{\text{pred}} \neq 3$	0,09	0,81