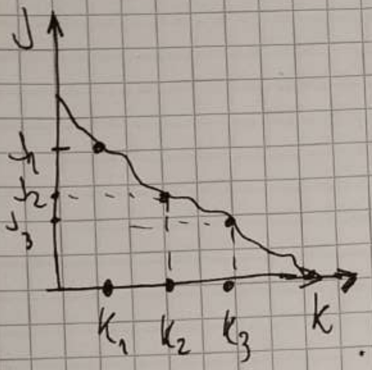
VI9

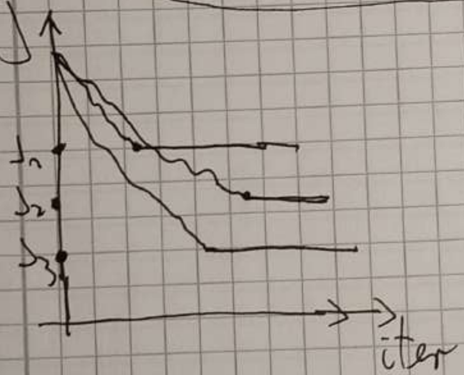
1.1 a)



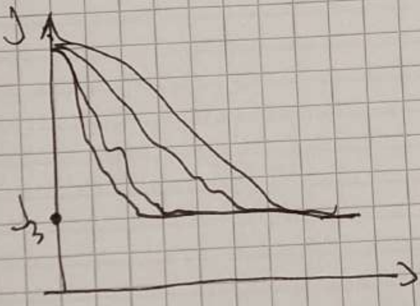

Minimalna vrijednost fje J = 0.
i to za K = broj primjera.


MARKO ĐUKIN
0036524589


b)

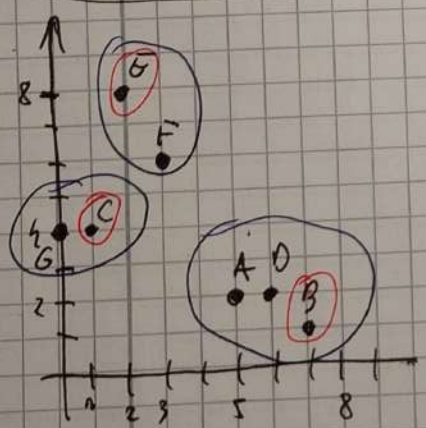



c) Uzmimo $K_3$




Za k++ najizglednija krivulja je ona koja PRVA dođe do J3. Zato jer k++ već počinje iteracije s "dobrim" centroidima.

1.2.




a, b) Oba algoritma u 1. koraku rade grupiranja kao na slici. k-means zatim u svakoj grupi računa novi centroid. A to su:

$$\vec{\mu}_k = \left[ \frac{1}{2}\left( (2,8)+(3,6) \right), \frac{1}{2}\left( (0,4)+(1,4) \right), \frac{1}{3}\left( (5,2)+(6,2)+(7,1) \right) \right]$$

k-medoids gleda koji primjer od trenutnih medoida minimizira grešku u svakoj grupi. Novi mogući medoidi su:

E ili F, C ili G, D

nejasno, ista greška

1.2.

c) k-medians: $O(Tn Nk)$ N - broj primjera

k-medoids: $O(Tk(N-k)^2)$ k - broj grupa

m - broj značajki

d) Prednost je što može raditi na podatcima koji nisu iz vektorskog prostora. Nedostatak je taj što nema efikasne implementacije pa je složenost velika

MARKO ĐUKIN, 0036524589

V20

1.1

a) EM-algoritam koristi meko grupiranje gdje imamo probabilistički izlaz. U odnosu na k-means složeniji je.

b)

$$p(x) = \sum_k^K p(x, y=k) = \sum_k^K \pi_k \, p(x|\theta_k)$$

$$\ln L(\theta|D) = \sum_i \ln \sum_k \pi_k \, p(x^{(i)}|\theta_k)$$

c) $$p(x, z|\theta) = \prod_k^K \pi_k^{z_k} \, p(x|\theta_k)^{z_k}$$

$$\ln L(\theta|D, z) = \sum_i \sum_k z_k^i (\ln \pi_k + \ln p(x^{(i)}|\theta_k))$$

Ne možemo jer su $z^{(i)}$ varijable nisu poznate.

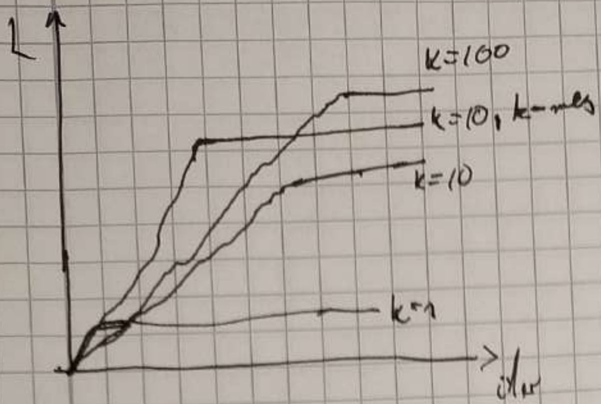
d)

E-korak: $\forall x^i, \forall k$ $$h_k^{(i)} = \frac{p(x^{(i)}|\mu_k, \Sigma_k)\pi_k}{\sum_j p(x^{(i)}|\mu_j, \Sigma_j)\pi_j}$$

M-korak: $\forall k$

$$\mu_k = \frac{\sum_i h_k^i x^i}{\sum_i h_k^i}, \quad \Sigma_k = \frac{\sum_i h_k^i (x^i-\mu_k)(x^i-\mu_k)^T}{\sum_i h_k^i}, \quad \pi_k = \frac{1}{N}\sum_i h_k^i$$

1.1. b)



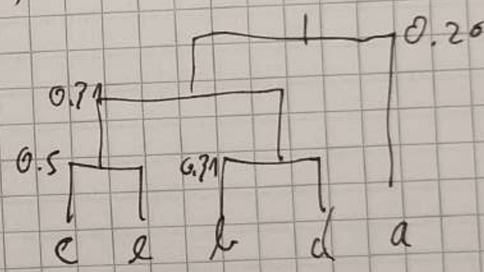

MARKO ĐUKIN, 0036524589

2.2

a) MAX, MAX

$$S = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{ccccc} a & b & c & d & e \end{array}$$
$$\left(\begin{array}{ccccc} 1 & 0.26 & 0.15 & 0.2 & 0.17 \\ & 1 & 0.24 & 0.31 & 0.31 \\ & & 1 & 0.2 & 0.5 \\ & & & 1 & 0.24 \\ & & & & 1 \end{array}\right) \rightarrow \begin{array}{c|cccc} & a & b & d & ce \\ \hline a & 1 & 0.26 & 0.2 & 0.17 \\ b & & 1 & 0.31 & 0.31 \\ d & & & 1 & 0.24 \\ ce & & & & 1 \end{array}$$

$$\rightarrow \begin{array}{c|ccc} a & 1 & 0.26 & 0.17 \\ bd & & 1 & 0.31 \\ ce & & & 1 \end{array} \rightarrow \begin{array}{c|cc} a & 1 & 0.26 \\ bdce & & 1 \end{array}$$

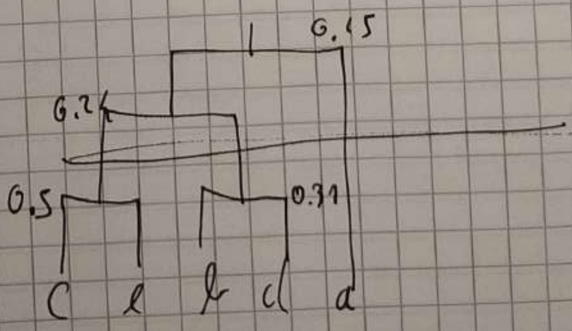



Na razini gdje bi bilo {{c,e},{b,d},{a}}

b) MIN

$$\rightarrow \begin{array}{c|cccc} & a & b & d & ce \\ \hline a & 1 & 0.26 & 0.2 & 0.15 \\ b & & 1 & 0.31 & 0.24 \\ d & & & 1 & 0.24 \\ ce & & & & 1 \end{array} \rightarrow \begin{array}{c|ccc} a & 1 & 0.2 & 0.15 \\ bd & & 1 & 0.24 \\ ce & & & 1 \end{array}$$

$$\rightarrow \begin{array}{c|cc} a & 1 & 0.15 \\ bdce & & 1 \end{array}$$




Ovdje presjeći

V21.

1.1

pred

| | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 1 | 0 | 1 |
| 1 | 1 | 3 | 2 |
| 2 | 1 | 1 | 1 |

2:

| | 2 | 01 |
|---|---|---|
| 2 | 1 | 2 |
| 01 | 3 | 5 |

0:

| | 0 | 12 |
|---|---|---|
| 0 | 1 | 1 |
| 12 | 2 | 7 |

1:

| | 1 | 02 |
|---|---|---|
| 1 | 3 | 3 |
| 02 | 1 | 4 |

a) Točnost = $\Sigma$ po dijagonali $/N = \underline{5/11} = Acc$

b)

$$P^M = \frac{1}{3}\left(\frac{1}{2} + \frac{3}{6} + \frac{1}{3}\right) = 0.444$$

$$R^M = \frac{1}{3}\left(\frac{1}{3} + \frac{3}{4} + \frac{1}{4}\right) = 0.444$$

$$F_1^M = \frac{1}{3}\left(\frac{2\cdot\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}+\frac{1}{3}} + \frac{2\cdot\frac{3}{6}\cdot\frac{3}{4}}{\frac{3}{6}+\frac{3}{4}} + \frac{2\cdot\frac{1}{3}\cdot\frac{1}{4}}{\frac{1}{3}+\frac{1}{4}}\right) = \frac{1}{3}(0.4+0.6+0.285) = 0.428$$

MIKRO TABLICA ($\Sigma_k$ tablica)

$$= \begin{pmatrix} 5 & 6 \\ 6 & 16 \end{pmatrix}$$

$$P^\mu = \frac{5}{11} \qquad R^\mu = \frac{5}{11}$$

$$F_1^\mu = \frac{2\cdot P\cdot R}{P+R} = \frac{5}{11}$$

$$Acc = F_1^\mu = P^\mu = R^\mu$$

MARKO ĐUKIN

0036524589