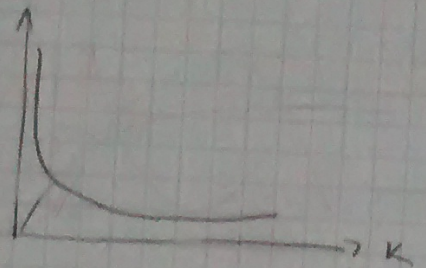
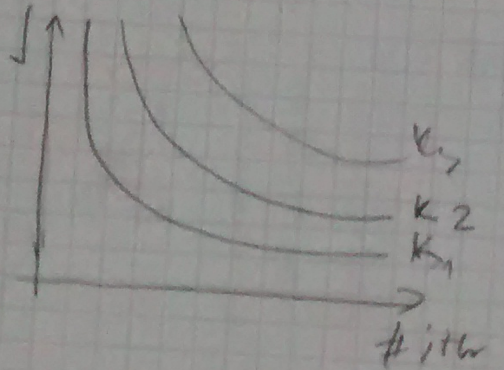
1. a)



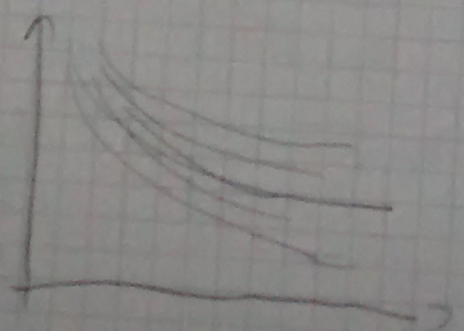
minimum je 0 $(x^{(i)} - \mu_{c^{(i)}})^2$

b)



$k_1 > k_2 > k_3$
↑
više klast.

c)



snap
odg

zgo je
izgled nzg zn
kmeans zt
kolikoje
početne
vrijdi
↓
minimum
↓
manja
greška

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

a) $K = 3$ — k-means

$\mu_1 = b = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

$\mu_2 = c = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

$\mu_3 = e = \begin{bmatrix} 3 \\ 8 \end{bmatrix}$

① odredi vrijednosti

$b_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad b_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad b_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$

$a_1 = \sqrt{2^2 + 1^2} = \sqrt{5}$

$a_2 = \sqrt{3^2 + 2^2} = 2\sqrt{[illegible]}$

$a_3 = \sqrt{2^2 + 6^2} = 3\sqrt{[illegible]}$

$b_1 = \sqrt{0} = 0$

$b_2 = \sqrt{\ }$

$b_3 = \sqrt{5^2 + 2^2}$

$c_1 = \sqrt{2^2 + 3^2}$

$c_2 = \sqrt{\ }$

$d_1 = \sqrt{[illegible]} = \sqrt{2}$

$d_2 = \sqrt{2^2 + 2^2}$

$d_3 = \sqrt{3^2 + 6^2}$

$e_1 = \sqrt{\ }$

$f_1 = \sqrt{4^2 + 5^2} =$

$f_2 = \sqrt{2^2 + 2^2} = 2\sqrt{2}$

$f_3 = \sqrt{[illegible]}$

$g_1 = \sqrt{3^2 + 4^2}$

$g_2 = \sqrt{4^2 + 0} = 4$

$g_3 = \sqrt{3^2 + 4^2} = \sqrt{[illegible]}$

② [illegible]

$\mu_1 = \dfrac{\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 6 \\ 2 \end{bmatrix}}{3} = \begin{bmatrix} [illegible] \\ [illegible] \end{bmatrix}$

$\mu_2 = \dfrac{\begin{bmatrix} [illegible] \end{bmatrix} + \begin{bmatrix} [illegible] \end{bmatrix}}{2} = \begin{bmatrix} [illegible] \\ 4 \end{bmatrix}$

$\mu_3 = \dfrac{\begin{bmatrix} [illegible] \end{bmatrix} + \begin{bmatrix} [illegible] \end{bmatrix}}{2} = \begin{bmatrix} [illegible] \\ [illegible] \end{bmatrix}$

b) k-medoida

$\mu_1 = \begin{bmatrix} 3 \\ 9 \end{bmatrix}$ $\mu_2 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$ $\mu_3 = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$

[illegible]

④ a) EM → mala grupacija

b) $p(x) = \sum_{k=1}^{K} \pi_k p(x|\theta_k) = \sum_{k=1}^{K} \pi_k N(\mu_k, \Sigma_k)$

$\ln f(\theta|D) = \ln \prod_{i=1}^{N} p(x^{(i)})$

$= \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k p(x^{(i)}|\theta_k)$

c) $p(x|z,\theta) = \prod_{k=1}^{K} p(x|\theta_k)^{z_k}$, $p(z) = \prod_{k=1}^{K} \pi_k^{z_k}$

$P(x,z|\theta) = p(z) \cdot p(x|z,\theta)$

$= \prod_{k=1}^{K} \pi_k^{z_k} \prod_{i=1}^{K} p(x,\theta_k)^{z_k} =$

$= \prod_{k=1}^{K} \pi_k^{z_k} \cdot p(x,\theta_k)^{z_k}$

$\ln L(\theta|D,z) = \ln \prod_{i=1}^{N} p(x^{(i)}, z^{(i)}|\theta) =$

$= \ln \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z^{(i)}} p(x|\theta_k)^{z^{(i)}}$

$= \sum_{i=1}^{N} \sum_{k=1}^{K} z^{(i)} (\ln \pi_k + \ln(p(x^{(i)}|\theta_k)))$

- ne znamo, jer ne znamo vrijednosti varijable $z^{(i)}$

dni onxt   broj iteracija
                 manji znatajk

d) E-korak procjene.

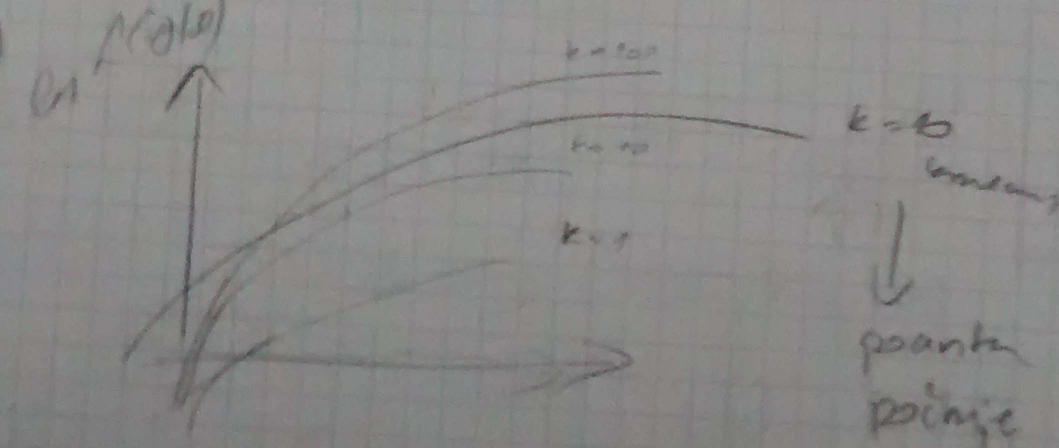
$$Q(\theta|\theta^{(t)}) = E_{z|D,\theta^{(t)}}[\ln \mathcal{L}(\theta|D,z)]$$

$$= \sum_{z} p(z|D,\theta^{(t)}) \ln p(D,z|\theta)$$

↳ Bayes

M-korak procjene

$$\theta^{(t+1)} = \underset{\theta}{\arg\max}\, Q(\theta|\theta^{(t)})$$

e) 

poanta počinje bolje

c) računalna složenost

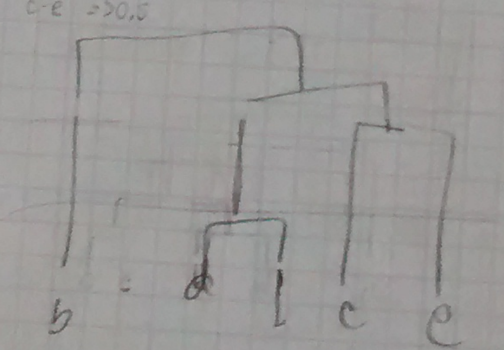
k-means $O(TnNK)$ — T: broj iteracija, n: broj značajki, N: broj primjera

k-medoida

$O(TK(N-k)^2)$ — T: broj iteracija, $(N-k)^2$: mjere sličnosti

(3)

a)



OVAKO NEKAKO IZGLEDAJU

KAKO DOĆI DO NJIH ?

b)