ZAVRŠNI ISPIT, 30. 01. 2014.

Zadatak 1.

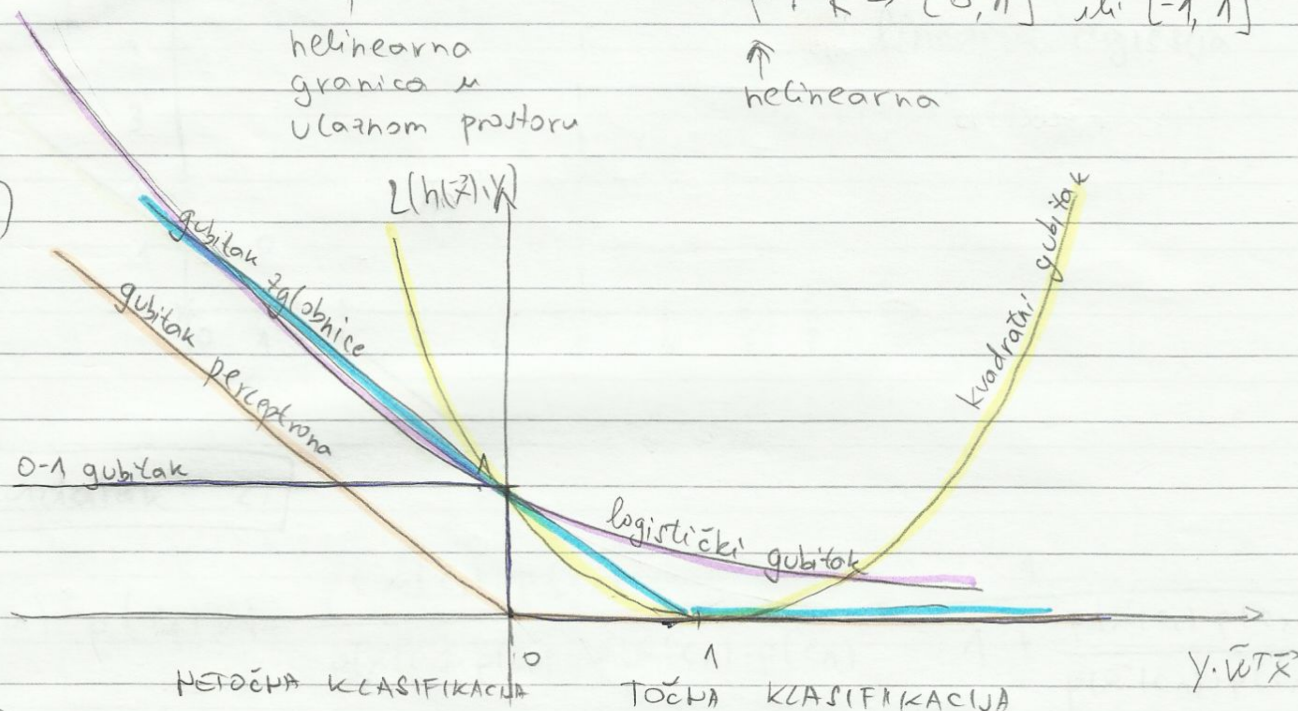
a) $h(\vec{x}) = F(\vec{\phi}(\vec{x})^T \cdot \tilde{w})$, $\vec{\phi} : \mathbb{R}^n \to \mathbb{R}^m$

$F : \mathbb{R} \to [0, 1]$ ili $[-1, 1]$

nelinearna granica u ulaznom prostoru

nelinearna

b)



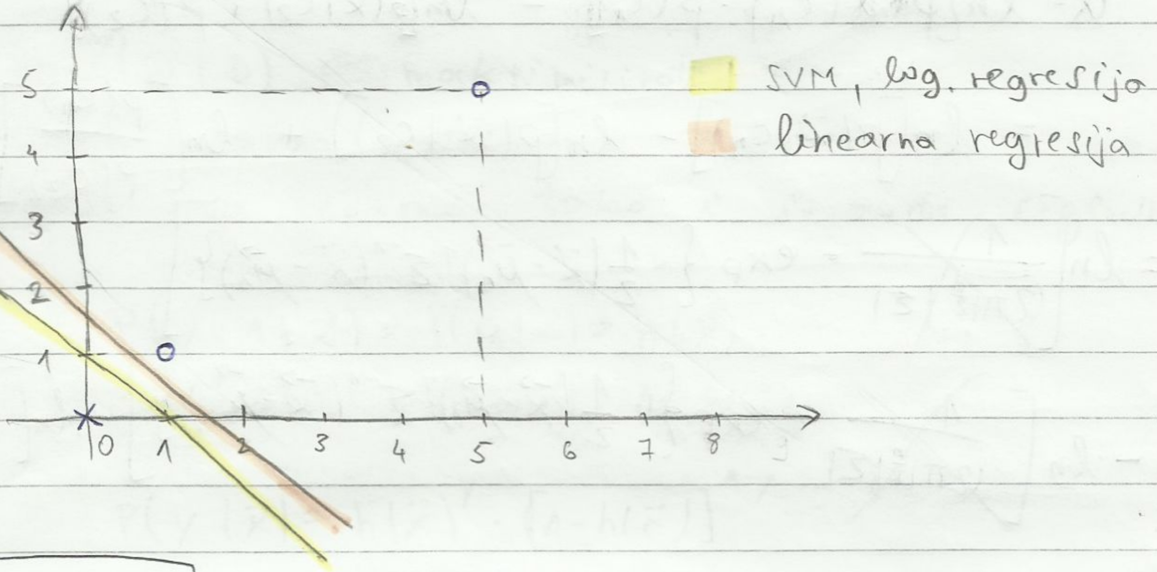
c)

Prednost logističke regresije nad linearnom regresijom je u tome što log. regresija ne kažnjava primjere koji su "previše točni" (daleko od granice na ispravnoj strani).

Nedostatak log. regresije u odnosu na SVM je taj što funkcija log. gubitka nema minimum pa (ako su primjeri linearno razdvojivi) će težine rasti u beskonačnost (gradijent nikad neće biti jednak nuli). Kod SVM-a će više težina biti pritegnuto na nulu.

d) $D = \{\vec{x}^{(i)}, y^{(i)}\} = \{((0,0),0), ((1,1),1), ((5,5),1)\}$

– regresija, logistička regresija, SVM



Zadatak 2.

a) $$p(C_1|\vec{x}) = \frac{p(\vec{x}|C_1)\cdot p(C_1)}{p(\vec{x}|C_2)\cdot p(C_2) + p(\vec{x}|C_1)\cdot p(C_1)} = \frac{1}{1 + \frac{p(\vec{x}|C_2)\cdot p(C_2)}{p(\vec{x}|C_1)\cdot p(C_1)}}$$

$$= \frac{1}{1+\exp(-\alpha)} = \sigma(\alpha)$$, gdje je

$$\exp(-\alpha) = \frac{p(\vec{x}|C_2)\cdot p(C_2)}{p(\vec{x}|C_1)\cdot p(C_1)} \quad / \ln \ /\cdot(-1)$$

$$\alpha = \ln\left[\frac{p(\vec{x}|C_1)\cdot p(C_1)}{p(\vec{x}|C_2)\cdot p(C_2)}\right]$$

$$\alpha = \vec{w}^T\vec{x} + w_0$$

Da bismo dobili linearan model uvodimo sljedeću pretpostavku: izglednosti klasa modeliramo Gaussovom razdiobom s dijeljenom kovarijacijskom matricom (kovarijancacijskom matricom.

$$\alpha = \ln[p(\vec{x}|c_1)\cdot p(c_1)] - \ln[p(\vec{x}|c_2)\cdot p(c_2)]$$

$$= \ln[p(\vec{x}|c_1)] - \ln[p(\vec{x}|c_2)] + \ln\left[\frac{p(c_1)}{p(c_2)}\right]$$

$$= \ln\left[\cancel{\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|}}\cdot \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_1)\right\}\right]$$

$$- \ln\left[\cancel{\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|}}\cdot \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_2)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_2)\right\}\right] + \ln\left[\frac{P(c_1)}{P(c_2)}\right]$$

$$= -\frac{1}{2}\left[\cancel{\vec{x}^T\Sigma^{-1}\vec{x}} - 2\vec{x}^T\Sigma^{-1}\vec{\mu}_1 + \vec{\mu}_1^T\Sigma^{-1}\vec{\mu}_1\right]$$

$$+\frac{1}{2}\left[\cancel{\vec{x}^T\Sigma^{-1}\vec{x}} - 2\vec{x}^T\Sigma^{-1}\vec{\mu}_2 + \vec{\mu}_2^T\Sigma^{-1}\vec{\mu}_2\right] + \ln\left[\frac{P(c_1)}{P(c_2)}\right]$$

$$= \vec{x}^T\underbrace{\Sigma^{-1}(\vec{\mu}_1-\vec{\mu}_2)}_{\vec{w}} \underbrace{-\frac{1}{2}\vec{\mu}_1^T\Sigma^{-1}\vec{\mu}_1 + \frac{1}{2}\vec{\mu}_2^T\Sigma^{-1}\vec{\mu}_2 + \ln\left[\frac{P(c_1)}{P(c_2)}\right]}_{w_0}$$

b)

$n = 50$, $K = 5$

Broj parametara log. regresije:

$$K\cdot(n+1) = 5\cdot(50+1) = 255$$

Broj parametara generativnog modela:

$$(5-1) + (5\cdot 50) + \frac{50\cdot(50+1)}{2} = 1529$$

$$\underbrace{(K-1)}_{\text{APRIORI}} + \underbrace{K\cdot n}_{\vec{\mu}_j} + \underbrace{\frac{n(n+1)}{2}}_{\Sigma}$$

Diskriminativni model će bolje generalizirati.

c)

$\ln \mathcal{L}(\vec{w} | D) \leftarrow$ maksimizirati

- ovdje izravno modeliramo aposteriornu vjerojatnost pa trebamo tako i izraziti izglednost

$$P(y=1|\vec{x}) = P(c_1|\vec{x}) = h(\vec{x})$$
$$P(y=0|\vec{x}) = P(c_2|\vec{x}) = 1-h(\vec{x})$$

$$P(y|\vec{x}) = h(\vec{x})^y \cdot [1-h(\vec{x})]^{1-y}$$

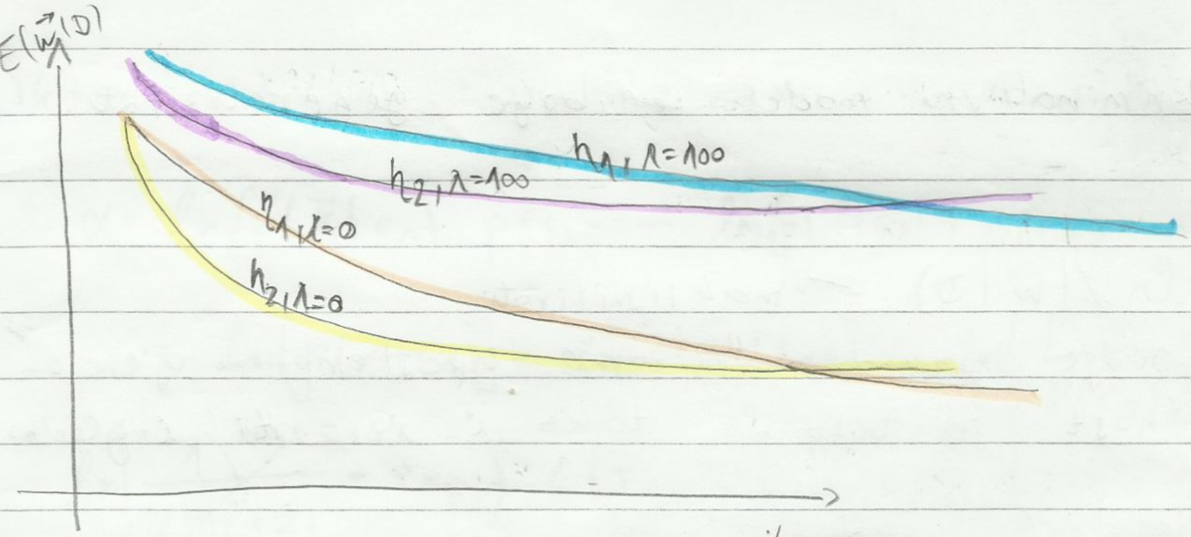
$$\ln \mathcal{L}(\vec{w}|D) = \ln p(D|\vec{w}) = \ln \prod_{i=1}^{N} p(y^{(i)}|\vec{x}^{(i)})$$
$$= \sum_{i=1}^{N} \left[y^{(i)} \cdot \ln[h(\vec{x}^{(i)})] + (1-y^{(i)}) \cdot \ln[1-h(\vec{x}^{(i)})] \right]$$

Sada je $E(\vec{w}|D) = -\ln \mathcal{L}(\vec{w}|D)$

$$= -\sum_{i=1}^{N} \left[y^{(i)} \cdot \ln[h(\vec{x}^{(i)})] + (1-y^{(i)}) \cdot \ln[1-h(\vec{x}^{(i)})] \right]$$

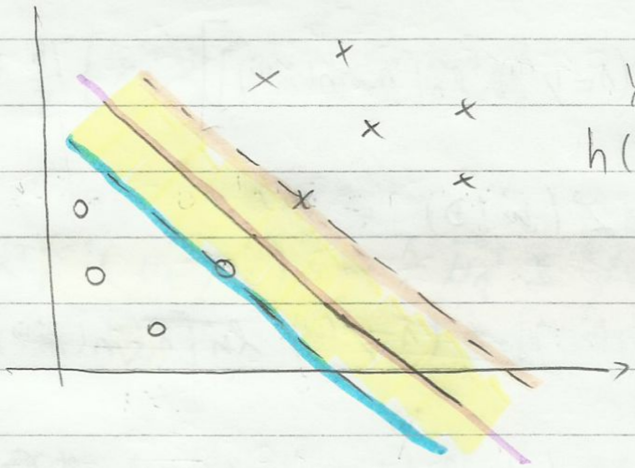
Odakle je $L(y, h(\vec{x})) = -y \cdot \ln[h(\vec{x})] - (1-y) \cdot \ln[1-h(\vec{x})]$ //

d)



Zadatak 3.

a) Problem maksimalne margine i Lagrange (primarni problem)



$$y \in \{-1, +1\}$$

$$h(\vec{x}) = \vec{\Phi}(\vec{x})^{T} \cdot \vec{w} + w_0$$

Udaljenost primjera od decizijske hiperravnine:

$$d(\vec{x}) = \frac{h(\vec{x})}{\|\vec{w}\|} \quad \text{(s predznakom).}$$

Bez predznaka (apsolutna vrijednost):

$$d(\vec{x}) = \frac{y \cdot h(\vec{x})}{\|\vec{w}\|} = \frac{1}{\|\vec{w}\|} \cdot y \cdot \left[\vec{\Phi}(\vec{x})^{T} \cdot \vec{w} + w_0 \right]$$

Margina je najmanja takva udaljenost:

$$\min_i \left[\frac{1}{\|\vec{w}\|} \cdot y^{(i)} \cdot \left[\vec{\phi}(\vec{x}^{(i)})^T \cdot \vec{w} + w_0 \right] \right]$$

To treba maksimizirati:

$$\vec{w}^*, w_0^* = \underset{\vec{w}, w_0}{\arg\max} \left\{ \frac{1}{\|\vec{w}\|} \cdot y^{(i)} \cdot \left[\vec{\phi}(\vec{x}^{(i)})^T \cdot \vec{w} + w_0 \right] \right\}$$

TEŽAK PROBLEM!

No, s obzirom da skaliranjem $\vec{w}$ i w_0 se udaljenost ne mijenja, možemo postaviti:

$$y^{(i)} \cdot \left[\vec{\phi}(\vec{x}^{(i)})^T \cdot \vec{w} + w_0 \right] \geq 1 \quad , \forall i$$

Tada imamo problem:

$$\vec{w}^*, w_0^* = \underset{\vec{w}, w_0}{\arg\max} \left\{ \frac{1}{\|\vec{w}\|} \right\} \text{, uz uvjet:}$$

$$y^{(i)} \cdot \left[\vec{\phi}(\vec{x}^{(i)})^T \cdot \vec{w} + w_0 \right] \geq 1 \text{, } \forall i$$

što je ekvivalentno:

$$\vec{w}^*, w_0^* = \underset{\vec{w}, w_0}{\arg\min} \left\{ \frac{1}{2} \|\vec{w}\|^2 \right\} \text{, uz isti uvjet.}$$

Rješavamo preko Lagrange-a:

$$\mathcal{L}(\vec{w}, w_0, \vec{\alpha}) = \frac{1}{2} \|\vec{w}\|^2 + \sum_{i=1}^{N} \alpha_i \left(1 - y^{(i)} \cdot h(\vec{x}^{(i)}) \right) \quad //$$

PRIMARNI PROBLEM.

b) Potporni vektori:

$$\vec{x}^{(1)} = \begin{bmatrix} -5 & 1 & 10 & 0 \end{bmatrix}^T \quad \vec{x}^{(2)} = \begin{bmatrix} 1 & 2 & -2 & 5 \end{bmatrix}^T$$

$$\vec{x}^{(3)} = \begin{bmatrix} 0 & -5 & -1 & 7 \end{bmatrix}^T$$

$$y^{(1)} = -1 \quad y^{(2)} = y^{(3)} = +1$$

$\alpha_1 = 0.01$, $\alpha_2 = 0.007$, $\alpha_3 = 0.003$

$w_0 = 0.45$

$$h(\vec{x}) = \vec{w}^T \cdot \vec{x} + w_0 = \underbrace{\sum_{i=1}^{N} \alpha_i y^{(i)} \cdot \vec{x}^{(i)T} \cdot \vec{x} + w_0}_{\text{DUALNA FORMULACIJA}}$$

Klasificirati $\vec{x}^{(4)} = \begin{bmatrix} 5 & 5 & -50 & 10 \end{bmatrix}^T$

$$\begin{aligned} h(\vec{x}^{(4)}) &= 0.01 \cdot [-5 \cdot 5 + 1 \cdot 5 + 10 \cdot (-50) + 0 \cdot 10] \\ &+ 0.007 \cdot [1 \cdot 5 + 2 \cdot 5 + (-2) \cdot (-50) + 5 \cdot 10] \\ &+ 0.003 \cdot [0 \cdot 5 + (-5) \cdot 5 + (-1) \cdot (-50) + 7 \cdot 10] \\ &+ 0.45 \\ &= -3.76 \end{aligned}$$

Primjer $\vec{x}^{(4)}$ je negativan i ne nalazi se unutar margine.

c) Jezgreni trik je zamjena skalarnog produkta kod treniranja i korištenja SVM-a jezgrenom funkcijom:

$$k(\vec{x}, \vec{x}') = \vec{\phi}(\vec{x})^T \cdot \vec{\phi}(\vec{x}')$$

Prednosti u odnosu na izravno preslikavanje očituju se u smanjenju računalne složenosti (izračun jezgrene funkcije često je jednostavniji nego izračun dvaju preslikavanja pa skalarnog umnoška) te što prostor značajki može biti visoko dimenzijski.

d) $k(\vec{x}, \vec{z}) = (\vec{x}^T \cdot \vec{z} + 1)^2$

Odrediti preslikavanje $\vec{\phi}(\vec{x})$ za $\vec{x} = [2 \ \ 1]^T$:

$$
\begin{aligned}
k(\vec{x}, \vec{z}) = (\vec{x}^T \cdot \vec{z} + 1)^2 &= (x_1 z_1 + x_2 z_2 + 1)^2 \\
&= \underline{x_1^2 z_1^2} + \underline{x_1 x_2 \cdot z_1 z_2} + \underline{x_1 \cdot z_1} \\
&+ \underline{x_1 x_2 z_1 z_2} + \underline{x_2^2 z_2^2} + \underline{x_2 \cdot z_2} \\
&+ \underline{x_1 z_1} + \underline{x_2 \cdot z_2} + \underline{1} \\
&= x_1^2 \cdot z_1^2 + 2 \cdot x_1 x_2 \cdot z_1 \cdot z_2 + 2 \cdot x_1 z_1 + 2 \cdot x_2 \cdot z_2 + x_2^2 z_2^2 + 1 \\
&= (x_1^2)(z_1^2) + (\sqrt{2} x_1 x_2) \cdot (\sqrt{2} z_1 \cdot z_2) + (\sqrt{2} x_1)(\sqrt{2} z_1) + (\sqrt{2} x_2)(\sqrt{2} z_2) + (x_2^2)(z_2)^2 + 1 \cdot 1
\end{aligned}
$$

Dakle:

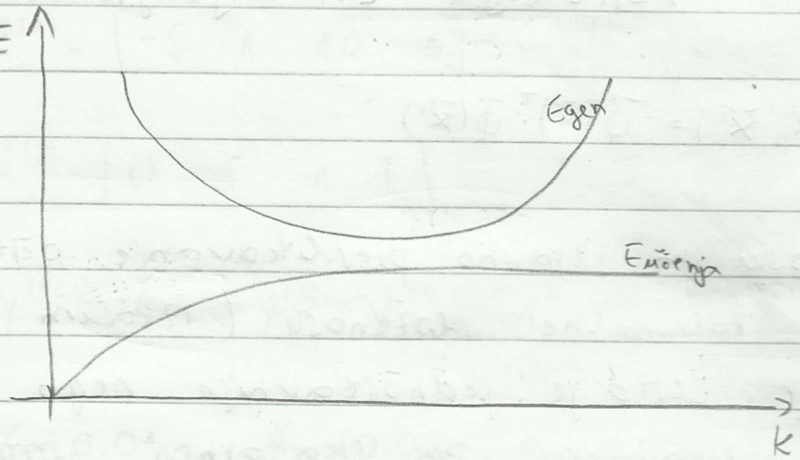
$$\vec{\phi}(\vec{x}) = [x_1^2 \quad \sqrt{2} x_1 x_2 \quad \sqrt{2} x_1 \quad \sqrt{2} x_2 \quad x_2^2 \quad 1]^T$$

$$\vec{\phi}([2 \ 1]^T) = [4 \quad 2\sqrt{2} \quad 2\sqrt{2} \quad \sqrt{2} \quad 1 \quad 1]^T$$

//

Zadatak 4.

a)



b) $\boxed{\text{ID3}}$

MJESTO	OTOK	SMJEŠTAJ	PRIJEVOZ	Y
Istra [2+, 1-]	da [2+, 1-]	privatni [2+, 1-]	auto [2+, 1-]	[3+, 4-]
Kvarner [0+, 2-]	ne [1+, 3-]	kamp [0+, 2-]	bus [0+, 2-]	
Dalmacija [1+, 1-]		hotel [1+, 1-]	avion [1+, 1-]	

Entropija na cijelom skupu:

$$E(D) = -\frac{3}{7} \cdot \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \cdot \log_2\left(\frac{4}{7}\right)$$

$$\boxed{E(D) = 0{,}9852}$$

MJESTO, SMJEŠTAJ, PRIJEVOZ (ista je raspodjela)

$$E(D, \text{Mjesto}, \text{Istra}) = E(D, \text{Smještaj}, \text{privatni}) = E(D, \text{Prijevoz}, \text{auto})$$

$$= -\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right)$$

$$= \boxed{0.9183}$$ //

$E(D, Mjesto, Kvarner) = E(D, Smještaj, kamp) = E(D, Prijevoz, bus)$

$$= \underbrace{0 \cdot \log_2 0}_{0} - \underbrace{\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right)}_{0} \boxed{= 0}$$

$E(D, Mjesto, Dalmacija) = E(D, Smještaj, hotel) = E(D, Prijevoz, avion)$

$$= -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) \boxed{= 1}$$

OTOK

$$E(D, Otok, da) = -\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right)$$

$$= \boxed{0.9183}$$

$$E(D, Otok, ne) = -\frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right)$$

$$= \boxed{0.8113}$$

INFORMACIJSKA DOBIT

$$Gain(D, Mjesto) = E(D) - \sum_v P(v) \cdot E(D, Mjesto, v)$$

$$= 0.9852 - \frac{3}{7} \cdot 0.9183 - \frac{2}{7} \cdot 0 - \frac{2}{7} \cdot 1$$

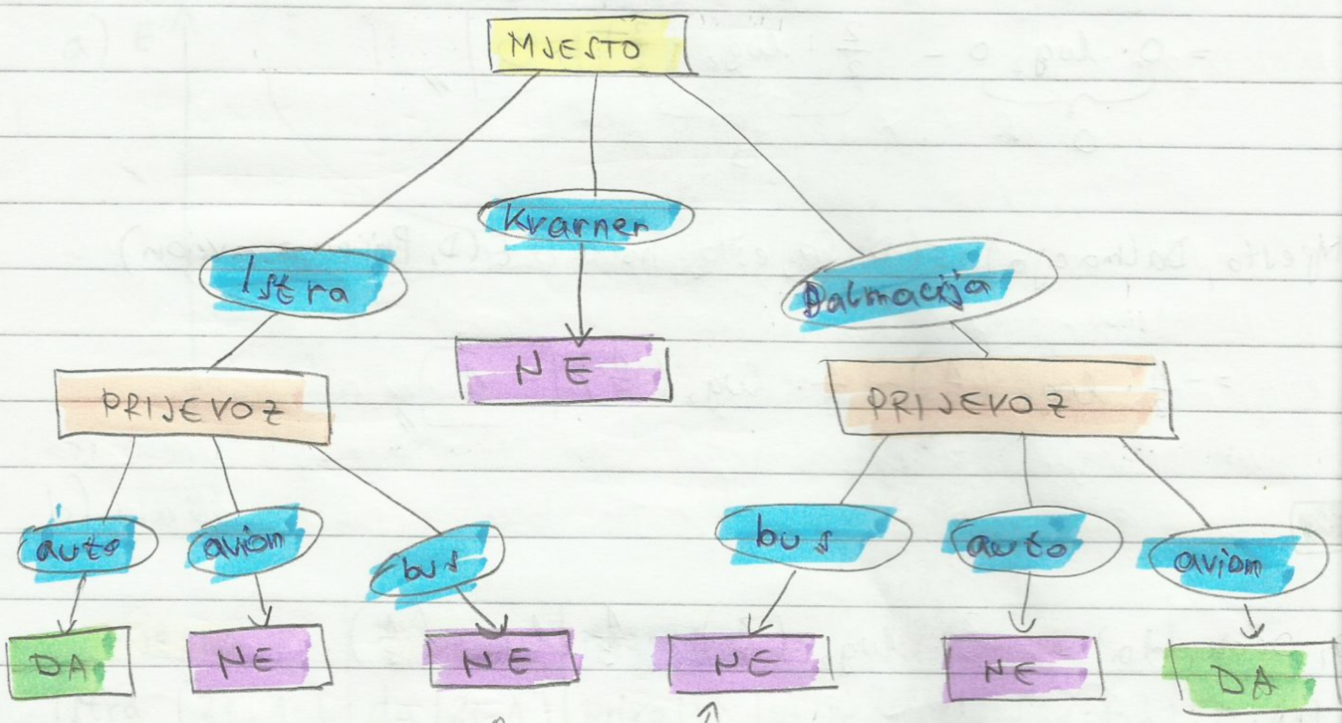
$$= 0.3059$$

$Gain(D, Otok) = 0.1280$

$Gain(D, Smještaj) = 0.3059$

$Gain(D, Prijevoz) = 0.3059$

Odabirem MJESTO.



jer se pojavljuje više puta nego DA

Zadatak 5.

	+	−
+	TP	FP
−	FN	TN

$P = \frac{TP}{TP+FP}$

$R = \frac{TP}{TP+FN}$

a)

PREDVIĐENO \ STVARNO	1	2	3
1	4	2	5
2	3	20	2
3	6	10	31

$TP_1 = 4 \quad FP_1 = 7 \quad FN_1 = 9$

$TP_2 = 20 \quad FP_2 = 5 \quad FN_2 = 12$

$TP_3 = 31 \quad FP_3 = 16 \quad FN_3 = 7$

$TP = 55 \quad FP = 28 \quad FP = 28$

Mikro

$$P = R = \frac{55}{55+28} = \frac{55}{83} \qquad F^{(mikro)} = \frac{55}{83} = \boxed{0.6627}$$

Makro

$$P_1 = \frac{4}{11} \quad R_1 = \frac{4}{13} \quad F_1 = \frac{1}{3}$$

$$P_2 = \frac{4}{5} \quad R_2 = \frac{5}{8} \quad F_2 = \frac{40}{57}$$

$$P_3 = \frac{31}{47} \quad R_3 = \frac{31}{38} \quad F_3 = \frac{62}{85}$$

$$F^{(makro)} = \frac{1}{3}\sum_{i=1}^{3} F_i = \frac{1}{3}\cdot\left[\frac{1}{3} + \frac{40}{57} + \frac{62}{85}\right] = \boxed{0.5882}$$

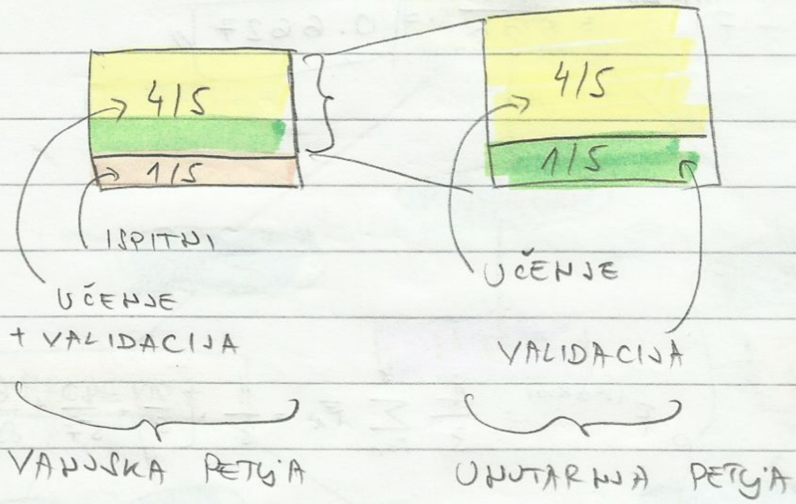
b) 1000 označenih primjera
hiperparametri C, γ
iscrpno pretraživanje po rešetci 10×10
5×5 ugniježđena unakrsna provjera

Broj učenja:

- vanjska petlja: 5 puta (skup učenje + skup validacija)
- unutarnja petlja:
  - za svaki pâr parametara
    - učenje se provodi 5 puta

Ukupno: $5 \cdot [100 \cdot 5 + 1] = \boxed{2505 \text{ puta}}$

Najmanji broj primjera s kojim ćemo učiti:

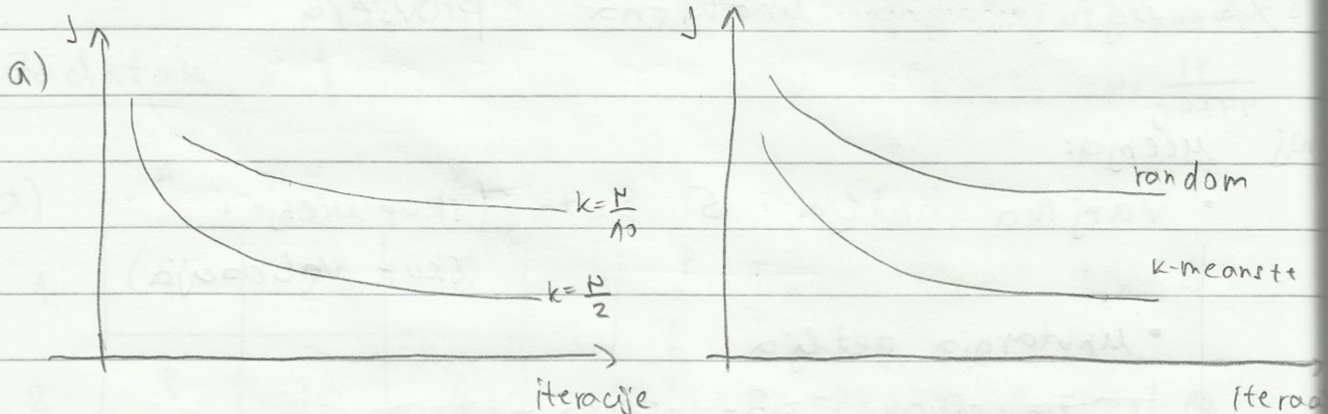


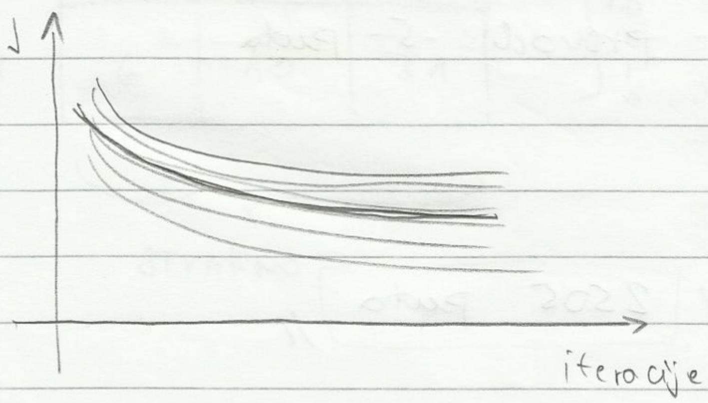
Najmanji broj primjera je:

$$1000 \cdot \frac{4}{5} \cdot \frac{4}{5} = \boxed{640}$$

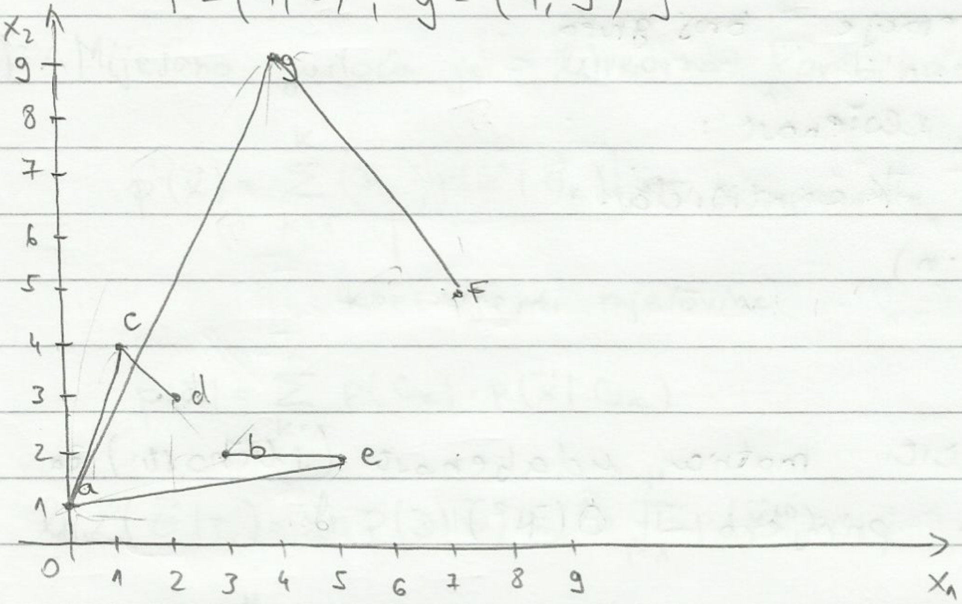
Zadatak 6.

a)

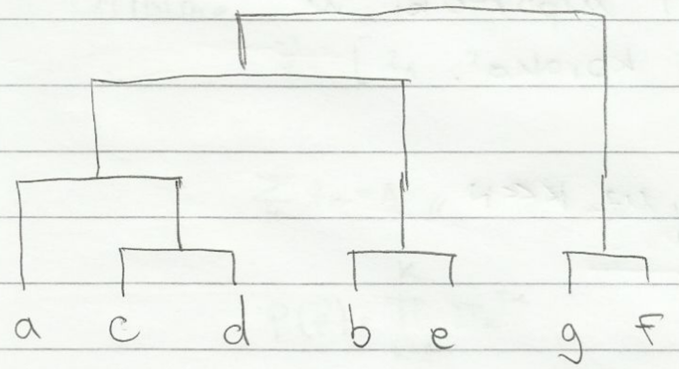




b) $D = \{ a = (0,1),\ b = (3,2),\ c = (1,4),\ d = (2,3),\ e = (5,2),$
$f = (7,5),\ g = (4,9) \}$



- aglomerativno, potpuno povezivanje



c) k-srednjih vrijednosti:
- za svaki od N primjera izračunati udaljenost od K centroida : $O(N \cdot K \cdot n)$
- izračunati K-centroida : $O(N \cdot n)$ - pribrojiti odgovarajući primjer centroidu.

broj primjera

Ukupno: $O(T \cdot N \cdot K \cdot n)$ — dimenzija

iteracije   broj grupa

Prostorna složenost:

- pamtiti K centroida:
  $O(K \cdot n)$

HAC:

Treba pamtiti matricu udaljenosti (sličnosti) za $\binom{N}{2}$ parova primjera — $O(N^2)$.

Vremenska složenost:

- nakon svakog stapanja računati udaljenost između nove grupe i svih ostalih — $O(N)$
- naći minimalno/maksimalno udaljen pār grupa — $O(N^2)$ usporedbi u svakom od $N-k$ koraka.

Ukupno: $O((N-K) \cdot N^2)$, uz $K \ll N$

$O(N^3)$ //

Zadatak 1.

a) Miješana gustoća je linearna kombinacija k gustoća:

$$p(\vec{x}) = \sum_{k=1}^{K} \underbrace{\pi_k}_{\text{koeficijenti mješavine}} \cdot \boxed{p(\vec{x} \mid \vec{\theta}_k)} \leftarrow \text{komponente mješavine}$$

$$p(\vec{x}) = \sum_{k=1}^{K} p(G_k) \cdot p(\vec{x} \mid G_k)$$

$$\ln \mathcal{L}(\vec{\theta} \mid D) = \ln p(D \mid \vec{\theta}) = \ln \prod_{i=1}^{N} p(\vec{x}^{(i)})$$

$$= \ln \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \cdot p(\vec{x}^{(i)} \mid \vec{\theta}_k)$$

$$= \sum_{i=1}^{N} \ln \Big[\underbrace{\sum_{k=1}^{K}}_{\text{PROBLEM!}} \pi_k \cdot p(\vec{x}^{(i)} \mid \vec{\theta}_k) \Big]$$

Uzmimo

$$\vec{z} = [z_1 \quad z_2 \quad \dots \quad z_k]$$ gdje $z_k = 1$ ako je $\vec{x}$ generiran iz G_k

$$\sum_k z_k = 1 \quad //$$

$$p(\vec{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

$$p(\vec{x} \mid \vec{z}, \vec{\theta}) = \prod_{k=1}^{K} p(\vec{x} \mid \vec{\theta}_k)^{z_k}$$

Zajednička vjerojatnost:

$$p(\vec{x}, \vec{z} \mid \vec{\theta}) = P(\vec{z}) \cdot p(\vec{x} \mid \vec{z}, \vec{\theta})$$

$$= \prod_{k=1}^{K} \pi_k^{z_k} \cdot \prod_{k=1}^{K} p(\vec{x} \mid \vec{\theta}_k)^{z_k} = \prod_{k=1}^{K} \pi_k^{z_k} \cdot p(\vec{x} \mid \vec{\theta}_k)^{z_k}$$

Sada je potpuna log-izglednost:

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

$$= \ln \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{z_k^{(i)}} \cdot p(\vec{x}^{(i)} \mid \vec{\theta}_k)^{z_k^{(i)}}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} z_k^{(i)} \cdot \left[\ln(\pi_k) + \ln\left[p(\vec{x}^{(i)} \mid \vec{\theta}_k) \right] \right] \quad //$$

b)

INICIJALIZIRAJ PARAMETRE $\vec{\theta} = \{\pi_k, \vec{\mu}_k, \Sigma_k\}_{k=1}^{K}$

PONAVLJAJ

$$h_k^{(i)} = \frac{p(\vec{x}^{(i)} \mid \vec{\mu}_k, \Sigma_k) \cdot \pi_k}{\sum_{j=1}^{K} p(\vec{x}^{(i)} \mid \vec{\mu}_j, \Sigma_j) \cdot \pi_j} \quad \Bigg\} \text{ E-korak}$$

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} h_k^{(i)}$$

$$\vec{\mu}_k = \frac{\sum_{i=1}^{N} h_k^{(i)} \cdot \vec{x}^{(i)}}{\sum_{i=1}^{N} h_k^{(i)}}$$

$$\Sigma_k = \frac{\sum_{i=1}^{N} h_k^{(i)} \cdot [\vec{x}^{(i)} - \vec{\mu}_k][\vec{x}^{(i)} - \vec{\mu}_k]^T}{\sum_{i=1}^{N} h_k^{(i)}}$$

(DO KONVERGENCIJE

c) $K=5$ klasa

Model Gaussove mješavine

i) model s nedijeljenom kovarijacijskom matricom

ii) model s dijeljenom izotropnom kov. matricom