MEĐUISPIT, 21.11.2013.

Zadatak 1.

a) Tri osnovne komponente algoritma nadziranog strojnog učenja:
- model (prostor hipoteza),
- funkcija gubitka,
- optimizacijski postupak.

Model uvodi induktivnu pristranost ograničavanjem (jezika), a funkcija gubitka i optimizacijski postupak uvode pristranost preferencijom (pretraživanja).

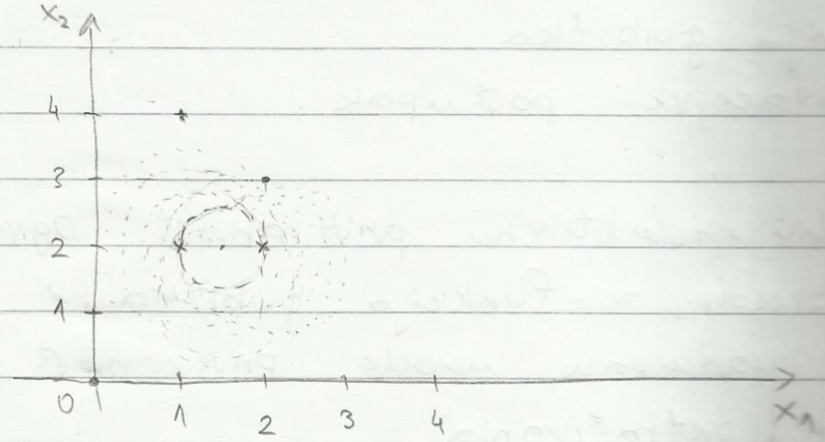
b) Kod parametarskih modela broj parametara modela je unaprijed određen. Složenost modela raste proporcionalno s brojem parametara modela.
Kod neparametarskih modela broj parametara ovisi o broju primjera za učenje. Samim time i složenost ovisi o broju primjera za učenje.

Zadatak 2.

a) Model:

$$h(\vec{x}|\vec{\theta}) = \mathbb{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leq \theta_0\}, \quad \vec{\theta} \in \mathbb{R}^3, \quad \vec{x} \in \mathbb{Z}^2$$

$$D = \{((0,0),0), ((2,3),0), ((1,2),1), ((2,2),1)\}$$



Klasifikacija primjera $\vec{x} = (1,4)$:

- model su kružnice. Jedino je moguće primjer $\vec{x}$ klasificirati u razred 0. Dakle $(\vec{x}, y) = ((1,4), 0)$.

Prostor inačica:

- sve kružnice koje obuhvaćaju primjere $(1,2)$, $(2,2)$; od najspecifičnije do najopćenitije.

b) VC-dimenzija modela $\mathcal{H}$ je najveći broj primjera koje $\mathcal{H}$ može razdijeliti.

Model $\mathcal{H}$ razdjeljuje N primjera akko:

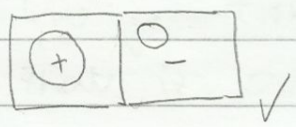
$$\exists(\vec{x}^{(1)}, \vec{x}^{(2)}, \dots, \vec{x}^{(N)}) \subseteq \mathcal{X}, \forall y, \exists h \in \mathcal{H}, \forall i \in \{1, \dots, N\}:$$

$$\left(h(\vec{x}^{(i)}) = y(\vec{x}^{(i)})\right), \quad y: \mathcal{X} \to \{0,1\}$$

Karakterizacija složenosti modela preko VC dimenzije je problematična jer daje pesimističnu ocjenu. Npr. VC-dimenzija modela pravca je 3. No u stvarnosti mi ćemo pravcem moći razdijeliti u puno veći broj primjera jer ti podaci ipak govore nekakvu priču i nisu samo bezveze razbacani po prostoru.

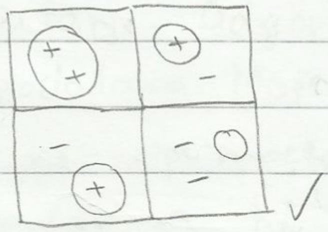
c) VC-dimenzija modela pod ⓐ.

1 točka:



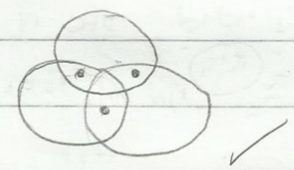
$VC(\mathcal{H}) \geq 1$

2 točke:



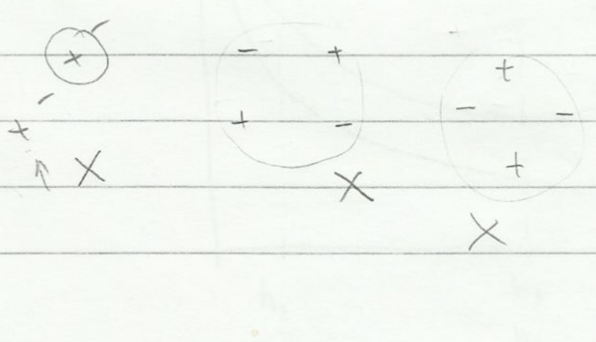
$VC(\mathcal{H}) \geq 2$

3 točke:
(1 problematičan slučaj; odvojiti svaki par točaka)



$VC(\mathcal{H}) \geq 3$

4 točke:



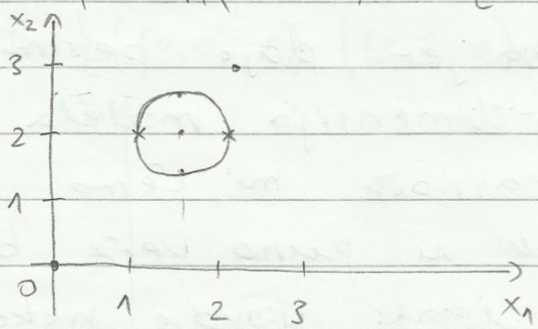
$VC(\mathcal{H}) < 4$. Dakle

$$\boxed{VC(\mathcal{H}) = 3.}$$

d) Redefinirati, tako da:

- $|VS_{\mathcal{H},D}| = 1$:

Npr. $h(\vec{x} \mid \vec{\theta}) = \mathbb{1}\{(x_1-\theta_1)^2+(x_2-\theta_2)^2 \leq 0.5^2\}, \ \vec{\theta} \in \mathbb{R}^2$

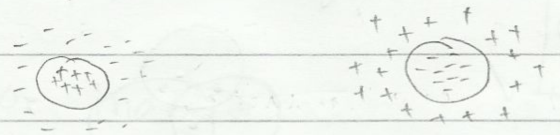


- $VC(\mathcal{H}) > 3$:

Npr. $h(\vec{x} \mid \vec{\theta}) = \begin{cases} \theta_3 & , (x_1-\theta_1)^2+(x_2-\theta_2)^2 \leq \theta_0 \\ 1-\theta_3 & , \text{inače} \end{cases}$

$\vec{\theta} \in \mathbb{R}^4, \ \theta_3 \in \{0, 1\}$

$\theta_3 = 1$ $\theta_3 = 0$



Zadatak 3.

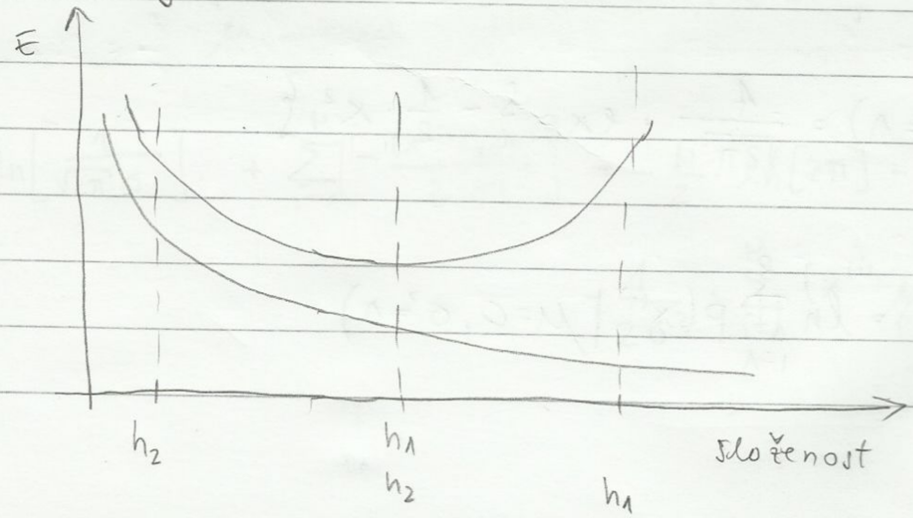
a) Model je prenaučen kada je previše složen, hipoteze su preraskošne, previše se prilagođavaju podacima pa zbog malih varijacija u podacima imamo velike oscilacije u hipotezama - model s visokom varijancom. Prenaučen model ima malu empirijsku, a visoku generalizacijsku pogrešku.

Prenaučene modele ne želimo zato što loše reagiraju na neviđenim podacima - slabo generaliziraju.

b) Dva modela: $\mathcal{H}_{\alpha_1}$, $\mathcal{H}_{\alpha_2}$, trenirana algoritmom L. α ugađa složenost modela.
Vrijedi $E(h_2|D) < E(h_1|D)$. Koji odabrati?

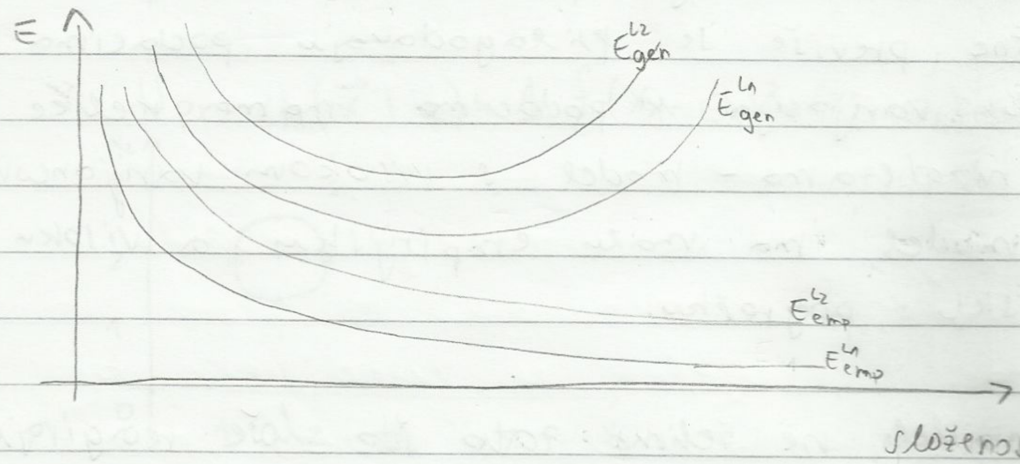
Ovo je trik pitanje. Odgovor je da ne možemo znati iz ovih podataka. Može biti da je model h_2 podnaučen, a h_1 optimalan. A može biti i da je h_2 optimalan, a h_1 prenaučen.

Ilustracija:



c) L_1 algoritam - analitički

L_2 algoritam - heuristički (suboptimalan)



Zadatak 4.

a) $\ln \mathcal{L}(\vec{\theta} \mid D) = \ln p(D \mid \vec{\theta}) = \ln p(\vec{x}^{(1)}, \vec{x}^{(2)}, \dots, \vec{x}^{(N)})$

$= \ln \prod_{i=1}^{N} p(\vec{x}^{(i)})$ uz pretpostavku i.i.d.

Gaussova razdioba:

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$D = \{0, 1, 2\}$, $\ln \mathcal{L}(\mu = 0, \sigma^2 = 1 \mid D) = ?$

$$p(x \mid \mu = 0, \sigma^2 = 1) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2} \cdot x^2\right\}$$

$$\ln \mathcal{L}(\mu = 0, \sigma^2 = 1 \mid D) = \ln \prod_{i=1}^{3} p(x^{(i)} \mid \mu = 0, \sigma^2 = 1)$$

$$= \sum_{i=1}^{3} \ln p(x^{(i)} \mid \mu = 0, \sigma^2 = 1) = \sum_{i=1}^{3} \ln \left[\frac{1}{\sqrt{2\pi}} \cdot \exp\left\{ -\frac{1}{2} (x^{(i)})^2 \right\} \right]$$

$$= 3 \cdot \left(-\frac{1}{2}\right) \cdot \ln[2\pi] - \frac{1}{2} \cdot \sum_{i=1}^{3} [x^{(i)}]^2 = -\frac{3}{2} \cdot \ln[2 \cdot \pi] - \frac{1}{2} [0 + 1 + 4]$$

$$\boxed{= -5.2568}$$

b) $\hat{\vec{\Theta}}_{ML} = \underset{\vec{\Theta}}{\arg\max} \{ \ln \mathcal{L}(\vec{\Theta} \mid D) \}$

$\hat{\vec{\Theta}}_{MAP} = \underset{\vec{\Theta}}{\arg\max} \{ p(D \mid \vec{\Theta}) \cdot p(\vec{\Theta}) \}$

Prednost MAP procjenitelja nad ML procjeniteljem je u tome što kombinira apriorno znanje o mogućim vrijednostima parametra $\vec{\Theta}$ s onim znanjem koje proizlazi iz uzorka. ML procjenitelj je sklon prenaučenosti pa MAP služi kao regularizator ML procjene.

c) $\hat{\mu}_{ML}$ za univarijatnog Gaussa

$$\hat{\mu}_{ML} = \underset{\mu}{\arg\max} \left\{ \ln \prod_{i=1}^{N} \left[\frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left\{ -\frac{(x^{(i)} - \mu)^2}{2\sigma^2} \right\} \right] \right\}$$

$$\sum_{i=1}^{N} \ln \left[\frac{1}{\sqrt{2\pi}\,\sigma} \right] + \sum_{i=1}^{N} \left[-\frac{(x^{(i)} - \mu)^2}{2\sigma^2} \right] = -\frac{N}{2} \cdot \ln[2\pi] - N \cdot \ln[\sigma]$$

$$- \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{N} (x^{(i)} - \mu)^2$$

$$\frac{\partial \ln \mathcal{L}(\mu,\sigma^2|D)}{\partial \mu} = \cancel{+}\frac{1}{\cancel{2}\sigma^2}\cdot\sum_{i=1}^{N} \cancel{2}\cdot(x^{(i)}-\mu)\cdot\cancel{(-1)} \overset{\text{postavi}}{=\!=} 0$$

$$\frac{1}{\sigma^2}\cdot\left[\sum_{i=1}^{N}(x^{(i)}-\mu)\right] = 0$$

$$\sum_{i=1}^{N} x^{(i)} - N\cdot\mu = 0 \Rightarrow \boxed{\hat{\mu}_{ML} = \frac{1}{N}\cdot\sum_{i=1}^{N} x^{(i)}}$$

d)

$N=10$
$m=2$ glava } Bernoullijeva razdioba $\quad p(x|\mu)=\mu^x\cdot(1-\mu)^{1-x}$

$$p(\mu|\alpha,\beta) = \frac{\mu^{\alpha-1}(1-\mu)^{\beta-1}}{B(\alpha,\beta)}, \quad \mu z \quad \alpha=\beta=2$$

FUNKCIJA IZGLEDNOSTI:

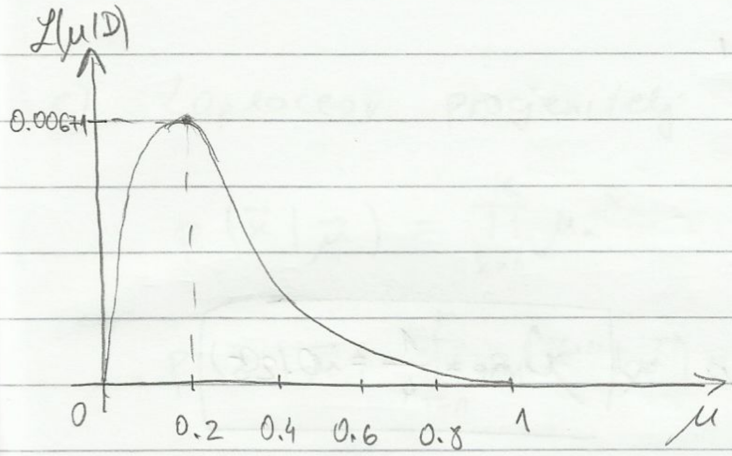
$$\mathcal{L}(\mu|D) = p(D|\mu) = \prod_{i=1}^{10} p(x^{(i)}|\mu) = \underbrace{\mu\cdot\mu}_{2}\cdot\underbrace{(1-\mu)\cdot(1-\mu)\cdots(1-\mu)}_{8}$$

$$\mathcal{L}(\mu|D) = \mu^2\cdot(1-\mu)^8$$

$$\ln \mathcal{L}(\mu|D) = 2\cdot\ln[\mu] + 8\cdot\ln[1-\mu] \quad /'$$

$$\frac{2}{\mu} - \frac{8}{1-\mu} = 0$$

$$2(1-\mu) = 8\mu \Rightarrow 2 = 10\mu \Rightarrow \boxed{\hat{\mu}_{ML} = \frac{1}{5} = 0.2}$$

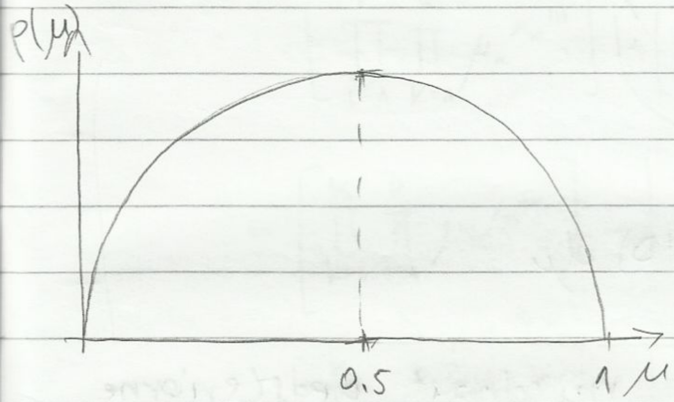


$$\hat{\mu}_{ML} = 0.2$$

$$\mathcal{L}(\hat{\mu}_{ML} \mid D) = \hat{\mu}_{ML}^{\,2}(1-\hat{\mu}_{ML})^8$$
$$= 0.2^2 \cdot 0.8^8$$
$$= 0.00671$$

FUNKCIJA GUSTOĆE:

$$p(\mu \mid \alpha=2, \beta=2) = \frac{\mu^{2-1}(1-\mu)^{2-1}}{B(2,2)} \propto \mu \cdot (1-\mu)$$



$$\frac{\alpha}{\alpha+\beta} = \frac{2}{2+2} = \frac{1}{2} = 0.5$$

APOSTERIORI:

$$p(\mu \mid D) = \frac{p(D \mid \mu) \cdot p(\mu)}{P(D)} \propto p(D \mid \mu) \cdot p(\mu)$$

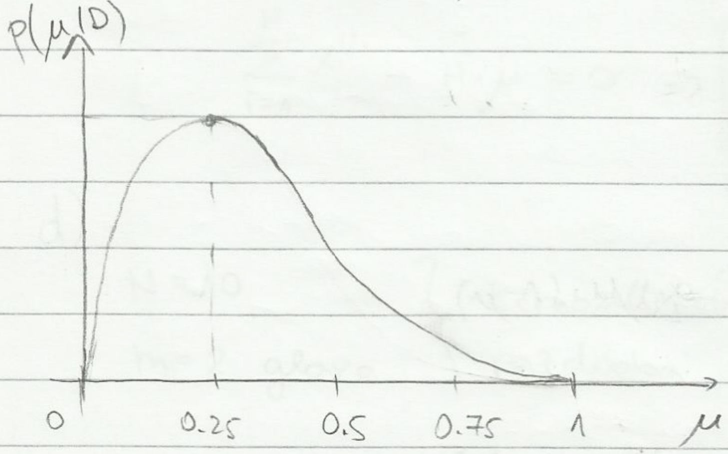
$$\hat{\mu}_{MAP} = \underset{\mu}{\arg\max} \left\{ p(D \mid \mu) \cdot p(\mu) \right\}$$

$$\left[\mu^2 \cdot (1-\mu)^8\right] \cdot \left[\mu \cdot (1-\mu)\right] = \mu^3 \cdot (1-\mu)^9 \quad / \ln$$

$$3 \cdot \ln[\mu] + 9 \cdot \ln[1-\mu] \quad \Big/ '$$

$$\frac{3}{\mu} - \frac{9}{1-\mu} = 0$$

$$3(1-\mu) = 9 \cdot \mu \quad \Rightarrow \quad 3 = 12\mu \quad \Rightarrow \quad \boxed{\hat{\mu}_{MAP} = \frac{1}{4} = 0.25}$$



$$MOD(\alpha', \beta') = \frac{\alpha'-1}{\alpha'+\beta'-2} = \frac{4-1}{4+10-2}$$

$$= 0.25 //$$

BAYES:

$$\hat{\mu}_{Bayes} = E[\mu|D] = \int \mu \cdot p(\mu|D) \, d\mu$$

Bayesov procjenitelj je srednja vrijednost aposteriorne distribucije.

$$p(\mu|D) \propto p(D|\mu) \cdot p(\mu) = \frac{\mu^3 \cdot (1-\mu)^9}{B(2,2)} = \frac{\mu^{\overset{\alpha'}{4}-1} \cdot (1-\mu)^{\overset{\beta'}{10}-1}}{B(2,2)}$$

$\alpha' = 4$, $\beta' = 10$ (općenito $\alpha' = m + \alpha$, $\beta' = N - m + \beta$)

Aposteriorna distribucija je beta distribucija uz $\alpha = 4$, $\beta = 10$.

$$\hat{\mu}_{Bayes} = \frac{\alpha'}{\alpha'+\beta'} = \frac{4}{4+10} = \frac{2}{7} = \boxed{0.2857} //$$

e) Laplaceov procjenitelj za $\vec{\mu}$.

$$p(\vec{x} \mid \vec{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

$$p(D \mid \vec{\mu}) = \prod_{i=1}^{N} p(\vec{x}^{(i)} \mid \vec{\mu}) = \prod_{i=1}^{N} \prod_{k=1}^{K} \mu_k^{x_k^{(i)}}$$

$$p(\vec{\mu} \mid \vec{\alpha}) = \mathrm{Dir}(\vec{\mu} \mid \vec{\alpha}) = \frac{1}{B(\vec{\alpha})} \cdot \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$

$$p(\vec{\mu} \mid D) \propto p(D \mid \vec{\mu}) \cdot p(\vec{\mu} \mid \vec{\alpha}) =$$

$$\left[\prod_{i=1}^{N} \prod_{k=1}^{K} \mu_k^{x_k^{(i)}}\right] \cdot \left[\frac{1}{B(\vec{\alpha})} \cdot \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}\right]$$

$$= \left[\prod_{k=1}^{K} \prod_{i=1}^{N} \mu_k^{x_k^{(i)}}\right] \cdot \left[\frac{1}{B(\vec{\alpha})} \cdot \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}\right]$$

$$= \left[\prod_{k=1}^{K} \mu_k^{\sum_{i=1}^{N} x_k^{(i)}}\right] \cdot \left[\frac{1}{B(\vec{\alpha})} \cdot \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}\right]$$

$$= \frac{1}{B(\vec{\alpha})} \cdot \prod_{k=1}^{K} \mu_k^{\underbrace{\sum_{i=1}^{N} x_k^{(i)} + \alpha_k}_{\alpha_k'} - 1}$$

$$E[\mu_k] = \frac{\alpha_k}{\sum_k \alpha_k} \qquad \hat{\mu}_{k,\mathrm{Bayes}} = E[\mu_k \mid D] = \frac{\alpha_k'}{\sum_k \alpha_k'}$$

$$\frac{\sum_{i=1}^{N} x_k^{(i)} + \alpha_k}{\sum_{k=1}^{K} \left[\sum_{i=1}^{N} x_k^{(i)} + \alpha_k\right]} = \boxed{\frac{N_k + \alpha_k}{N + \sum_{k=1}^{K} \alpha_k}}$$

Zadatak 5.

a) Naivan Bayesov klasifikator:

$$h(\vec{x}) = \arg\max_j \left\{ P(C_j) \cdot \prod_{k=1}^{n} P(x_k | C_j) \right\}$$

Broj parametara:

- $K-1$ za apriorne vjerojatnosti $P(C_j)$
- $(K_k - 1) \cdot K$ za likelihood, gdje je K_k broj mogućih vrijednosti koje značajka k može poprimiti

Ukupno: $K-1 + \sum_{k=1}^{n} (K_k - 1) \cdot K$

b) C_1 - spam
C_2 - $\neg$spam

Matrica gubitka:

$$L = \underbrace{\begin{array}{c|cc} & \text{spam} & \neg\text{spam} \\ \hline \text{spam} & 0 & 1 \\ \neg\text{spam} & 10 & 0 \end{array}}_{\text{PREDVIĐENO}} \Bigg\} \text{STVARNO}$$

$P(\vec{x} | C_1) = 0.7 \qquad P(C_1) = 0.8$

$P(\vec{x} | C_2) = 0.1 \qquad P(C_2) = 0.2$

$$R(C_j | \vec{x}) = \sum_{k=1}^{K} L_{kj} \cdot P(C_k | \vec{x})$$

$k=1: \quad R(c_1|\vec{x}) = \sum_k L_{k1} \cdot P(c_k|\vec{x}) = 10 \cdot P(c_2|\vec{x}) = \boxed{0.345}$

$k=2: \quad R(c_2|\vec{x}) = \sum_k L_{k2} \cdot P(c_k|\vec{x}) = 1 \cdot P(c_1|\vec{x}) = \boxed{0.9655}$

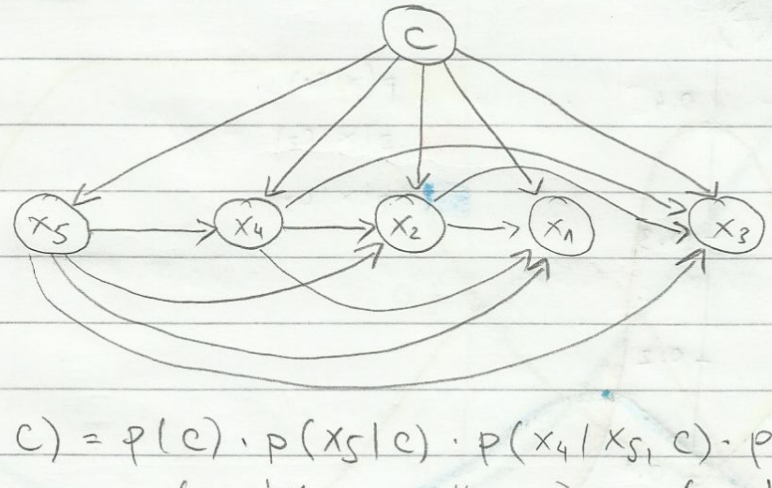
$$P(\vec{x}) = P(c_1)\cdot p(\vec{x}|c_1) + P(c_2)\cdot p(\vec{x}|c_2)$$
$$= 0.8 \cdot 0.7 + 0.2 \cdot 0.1$$
$$= 0.58$$

$$P(c_1|\vec{x}) = \frac{p(\vec{x}|c_1)\cdot p(c_1)}{p(\vec{x})} = \frac{0.7\cdot 0.8}{0.58} = 0.9655$$

$$P(c_2|\vec{x}) = \frac{P(\vec{x}|c_2)\cdot p(c_2)}{p(\vec{x})} = \frac{0.1\cdot 0.2}{0.58} = 0.0345$$

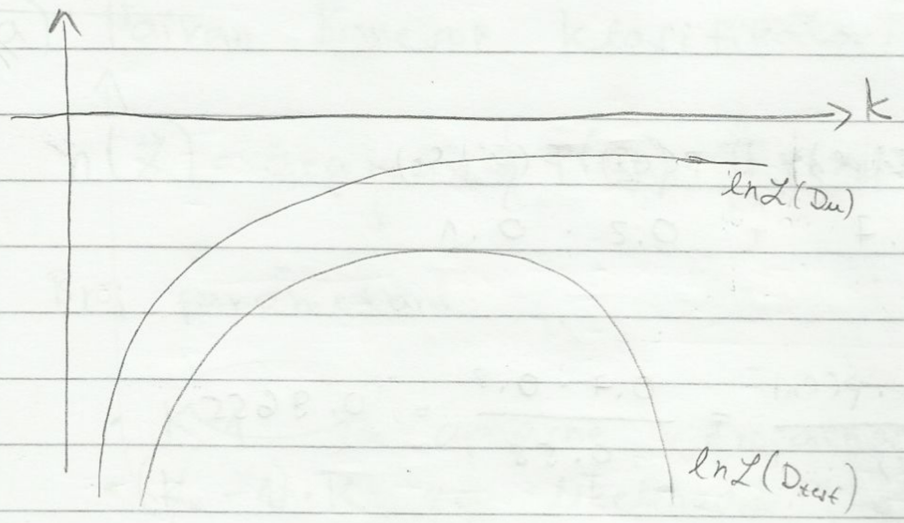
Klasificiram $\vec{x}$ u klasu c_1 – spam jer je $R(c_1|\vec{x}) < R(c_2|\vec{x})$.

c) $\mathcal{X} = \{0,1\}^5$, $K=2$, 3-DB



$$p(\vec{x}, C) = p(C)\cdot p(x_5|C)\cdot p(x_4|x_5, C)\cdot p(x_2|x_4, x_5, C)$$
$$\cdot\, p(x_1|x_2, x_4, x_5, C)\cdot p(x_3|x_2, x_4, x_5, C)$$

d) log-izglednost na skupu za učenje i provjeru ovisno o k za k-DB.



Veći k → veća složenost !

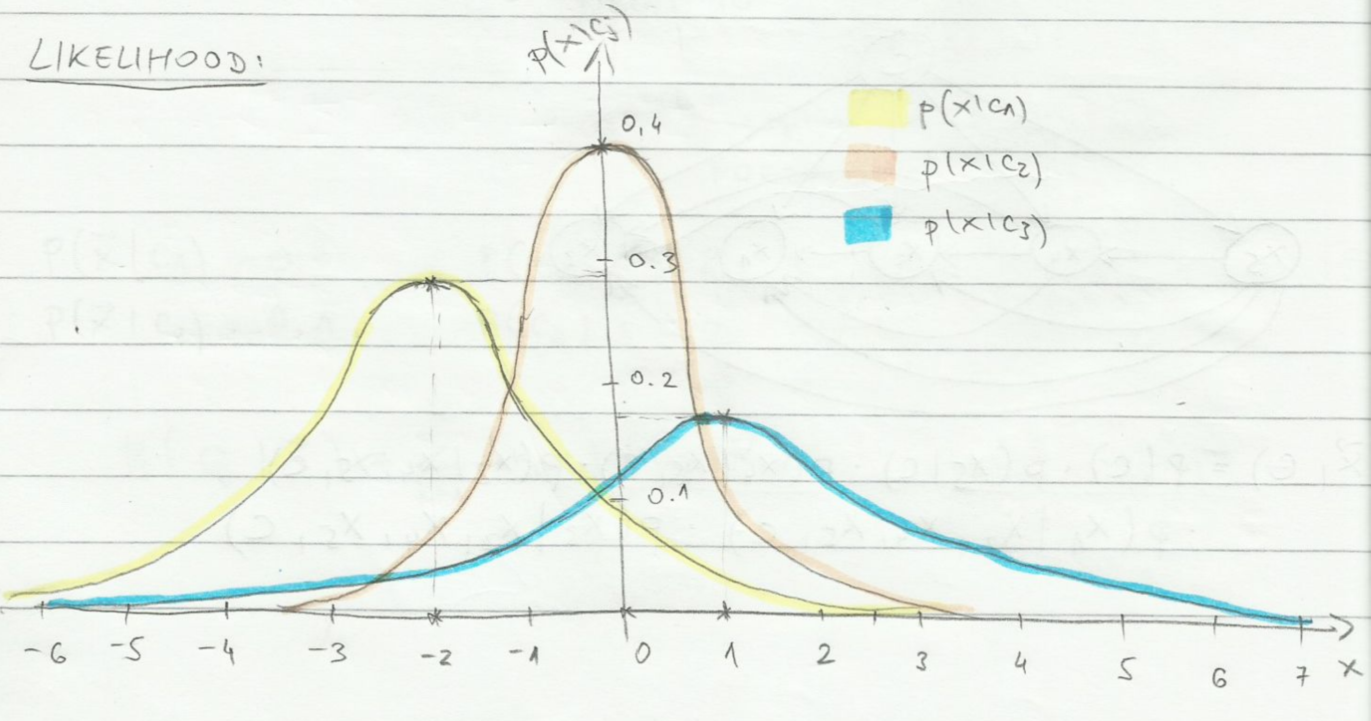
$\boxed{\text{Zadatak 6.}}$

a) $P(C_1) = 0.7$, $P(C_2) = 0.2$, $P(C_3) = 0.1$

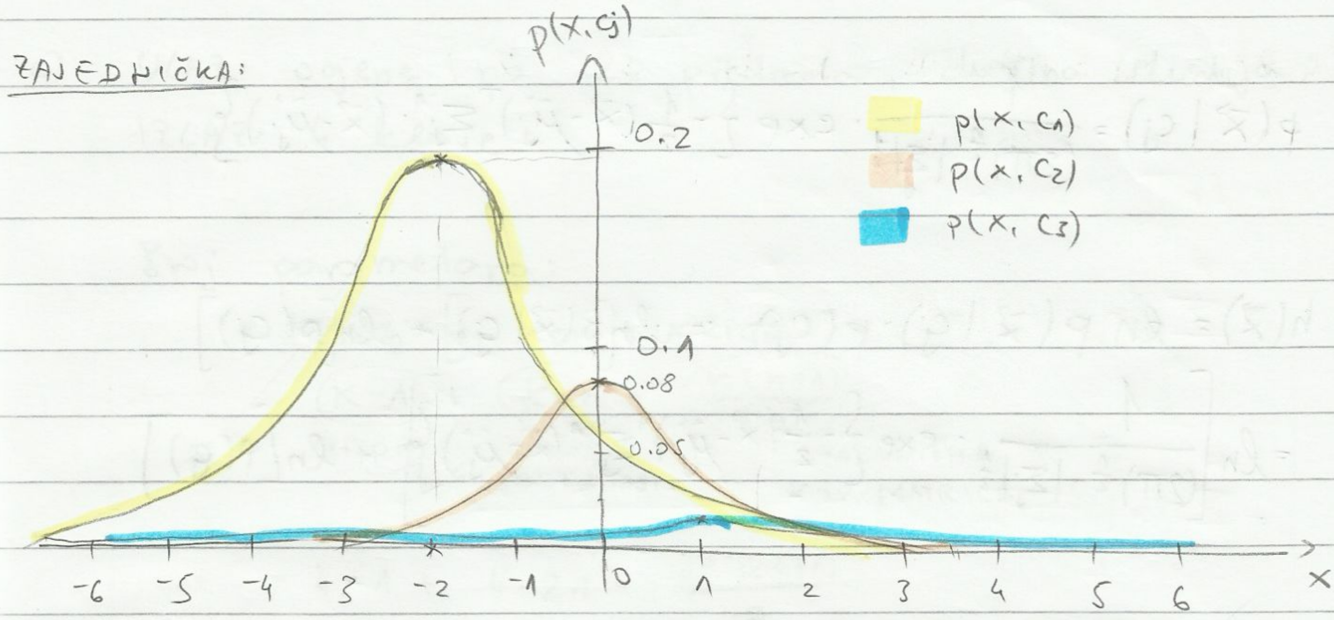
$\mu_1 = -2$, $\mu_2 = 0$, $\mu_3 = 1$

$\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $\sigma_3^2 = 5$

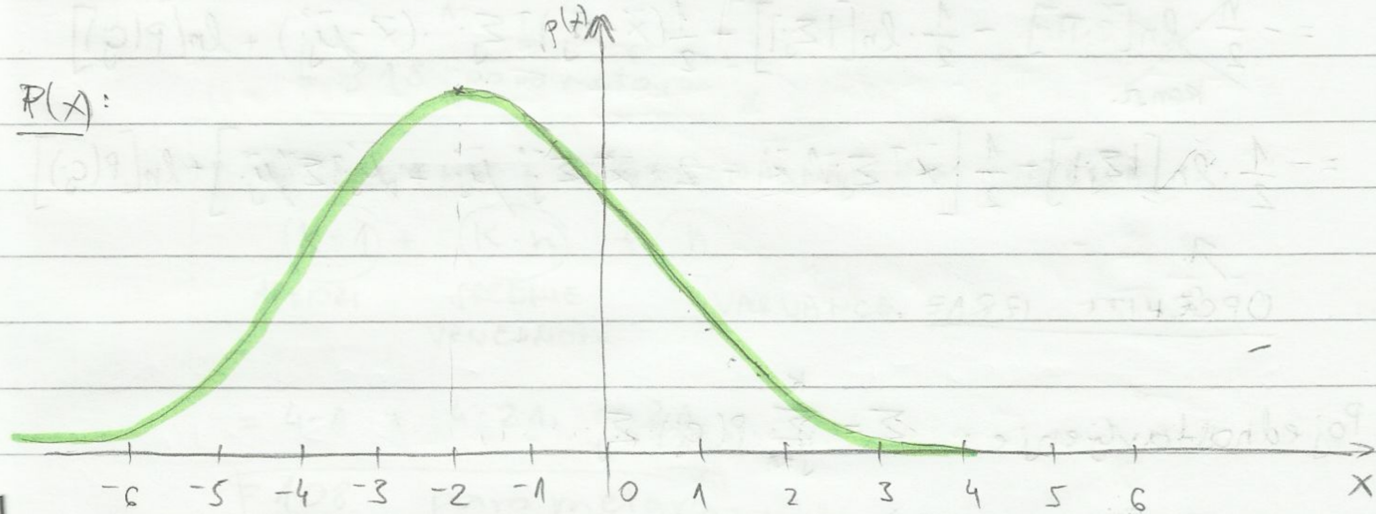
LIKELIHOOD:



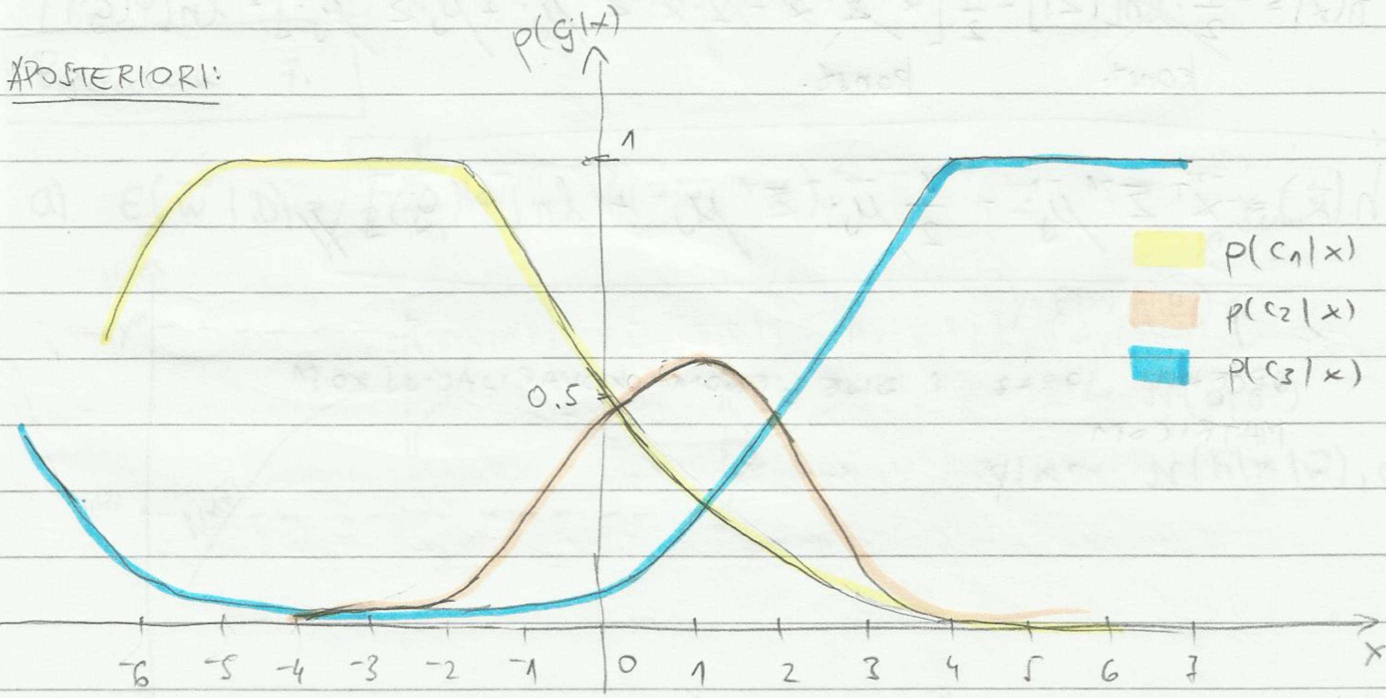
ZAJEDNIČKA:



P(X):



APOSTERIORI:



b)

$$p(\vec{x} \mid c_j) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_j|^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_j)^T \Sigma_j^{-1} \cdot (\vec{x}-\vec{\mu}_j)\right\}$$

$$h(\vec{x}) = \ln p(\vec{x} \mid c_j) \cdot p(c_j) = \ln[p(\vec{x} \mid c_j)] + \ln[p(c_j)]$$

$$= \ln\left[\frac{1}{(2\pi)^{\frac{n}{2}} \cdot |\Sigma_j|^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_j)^T \Sigma_j^{-1} (\vec{x}-\vec{\mu}_j)\right\}\right] + \ln[p(c_j)]$$

$$= \underbrace{\cancel{-\frac{n}{2} \cdot \ln[2 \cdot \pi]}}_{\text{konst.}} - \frac{1}{2} \cdot \ln[|\Sigma_j|] - \frac{1}{2}(\vec{x}-\vec{\mu}_j)^T \Sigma_j^{-1} \cdot (\vec{x}-\vec{\mu}_j) + \ln[p(c_j)]$$

$$= -\frac{1}{2} \cdot \ln[|\Sigma_j|] - \frac{1}{2}\left[\vec{x}^T \Sigma_j^{-1} \vec{x} - 2 \cdot \vec{x}^T \Sigma_j^{-1} \vec{\mu}_j + \vec{\mu}_j^T \Sigma^{-1} \vec{\mu}_j\right] + \ln[p(c_j)]$$

OPĆENITI IZRAZ.

Pojednostavljenje: $\Sigma = \sum_{j=1}^{K} p(c_j) \cdot \Sigma_j$

$$h(\vec{x}) = \underbrace{\cancel{-\frac{1}{2} \cdot \ln[|\Sigma|]}}_{\text{konst.}} - \frac{1}{2}\left[\underbrace{\cancel{\vec{x}^T \Sigma^{-1} \vec{x}}}_{\text{konst.}} - 2 \cdot \vec{x}^T \Sigma^{-1} \vec{\mu}_j + \vec{\mu}_j \cdot \Sigma^{-1} \vec{\mu}_j\right] + \ln[p(c_j)]$$

$$\boxed{h(\vec{x}) = \vec{x}^T \Sigma^{-1} \vec{\mu}_j - \frac{1}{2} \vec{\mu}_j^T \Sigma^{-1} \vec{\mu}_j + \ln[p(c_j)]}$$

OPĆENITI IZRAZ S DIJELJENOM KOVARIJACIJSKOM MATRICOM

c) ULAZ: ocjene na 20 predmeta, duljina studija

IZLAZ: 4 klase

Broj parametara:

- DIJELJENA KOV. MATRICA:

$$\underbrace{K-1}_{\text{APRIORI}} + \underbrace{K \cdot n}_{\text{SREDNJE VRIJEDNOSTI}} + \underbrace{\frac{n(n+1)}{2}}_{\text{ZAJEDNIČKA KOV. MATRICA}}$$

$$= 4-1 + 4 \cdot 21 + \frac{21 \cdot (21+1)}{2}$$

$$\boxed{= 318 \text{ parametara}}$$

- DIJELJENA I DIJAGONALNA KOV. MATRICA:

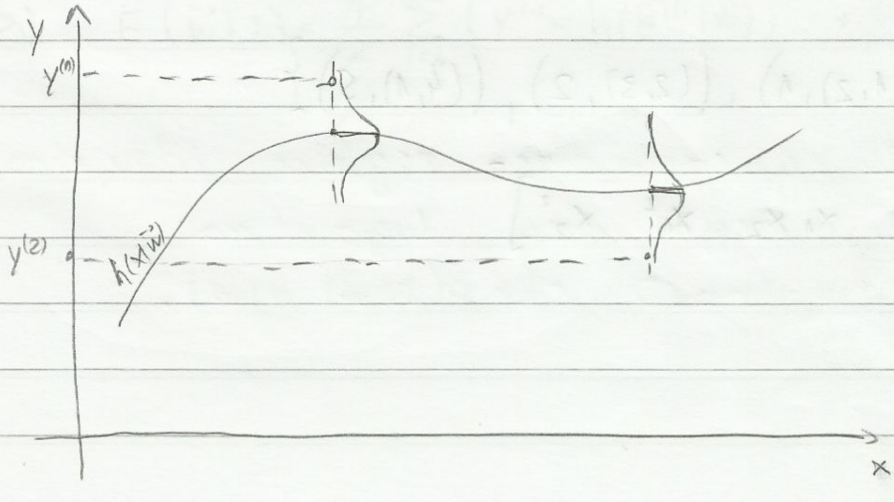
$$\underbrace{K-1}_{\text{APRIORI}} + \underbrace{K \cdot n}_{\text{SREDNJE VRIJEDNOSTI}} + \underbrace{n}_{\text{VARIJANCE}}$$

$$= 4-1 + 4 \cdot 21 + 21$$

$$\boxed{= 108 \text{ parametara}}$$

Zadatak 7.

a) $E(\vec{w} | D) = \frac{1}{2} \sum_{i=1}^{N} (y^{(i)} - \Phi(\vec{x}^{(i)})^T \cdot \vec{w})^2$



$y = f(x) + \varepsilon$ ← šum

$\varepsilon \sim \mathcal{N}(0, \sigma^2)$

$y|x \sim \mathcal{N}(h(x|\vec{w}), \sigma^2)$

$$\ln \mathcal{L}(\vec{w} | D) = \ln p(D | \vec{w}) = \ln \prod_{i=1}^{N} p(x^{(i)}, y^{(i)})$$

$$= \ln \prod_{i=1}^{N} p(y^{(i)} | x^{(i)}) \cdot p(x^{(i)})$$

$$= \sum_{i=1}^{N} \ln\left[p(y^{(i)} | x^{(i)})\right] + \underbrace{\cancel{\sum_{i=1}^{N} \ln\left[p(x^{(i)})\right]}}_{\text{ne ovisi o } \vec{w}}$$

$$= \sum_{i=1}^{N} \ln\left[\frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left\{-\frac{(y^{(i)} - h(x^{(i)} | \vec{w}))^2}{2\sigma^2}\right\}\right]$$

$$= -\frac{N}{2} \underbrace{\cancel{\ln[2\pi]}}_{\text{konst.}} - \underbrace{\cancel{N \cdot \ln[\sigma]}}_{\text{konst.}} - \frac{1}{2\cancel{\sigma^2}_{\text{konst.}}} \cdot \sum_{i=1}^{N} (y^{(i)} - h(x^{(i)} | \vec{w}))^2$$

$$= -\frac{1}{2} \sum_{i=1}^{N} \left[y^{(i)} - h(x^{(i)} | \vec{w})\right]^2$$

$$\vec{w}^{*} = \underset{\vec{w}}{\arg\min}\, E(\vec{w} | D) = \underset{\vec{w}}{\arg\max} \{\ln \mathcal{L}(\vec{w} | D)\}$$

$$= \underset{\vec{w}}{\arg\min} \{-\ln \mathcal{L}(\vec{w} | D)\}$$

b)

$$\vec{w} = (\Phi^T \Phi + \lambda \cdot I)^{-1} \cdot \Phi^T \cdot \vec{y}$$

$$D = \{((0,1),4), ((1,2),1), ((2,3),2), ((4,1),5)\}$$

$$\vec{\Phi}(\vec{x}) = \begin{bmatrix} 1 & x_1 & x_2 & x_1 x_2 & x_1^2 & x_2^2 \end{bmatrix}^T$$

$$\Phi = \begin{bmatrix} - \vec{\phi}(\vec{x}^{(1)})^T - \\ - \vec{\phi}(\vec{x}^{(2)})^T - \\ \vdots \\ - \vec{\phi}(\vec{x}^{(N)})^T - \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 2 & 2 & 1 & 4 \\ 1 & 2 & 3 & 6 & 4 & 9 \\ 1 & 4 & 1 & 4 & 16 & 1 \end{bmatrix}$$

$$\vec{w} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 \\ 1 & 2 & 3 & 1 \\ 0 & 2 & 6 & 4 \\ 0 & 1 & 4 & 16 \\ 1 & 4 & 9 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 2 & 2 & 1 & 4 \\ 1 & 2 & 3 & 6 & 4 & 9 \\ 1 & 4 & 1 & 4 & 16 & 1 \end{bmatrix} + \begin{bmatrix} \lambda & & & & & \\ & \lambda & & & 0 & \\ & & \lambda & & & \\ & & & \lambda & & \\ & 0 & & & \lambda & \\ & & & & & \lambda \end{bmatrix} \right)^{-1} \cdot$$

$$\cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 \\ 1 & 2 & 3 & 1 \\ 0 & 2 & 6 & 4 \\ 0 & 1 & 4 & 16 \\ 1 & 4 & 9 & 1 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 1 \\ 2 \\ 5 \end{bmatrix}$$

Treba invertirati matricu dimenzija $[6 \times 6]$.

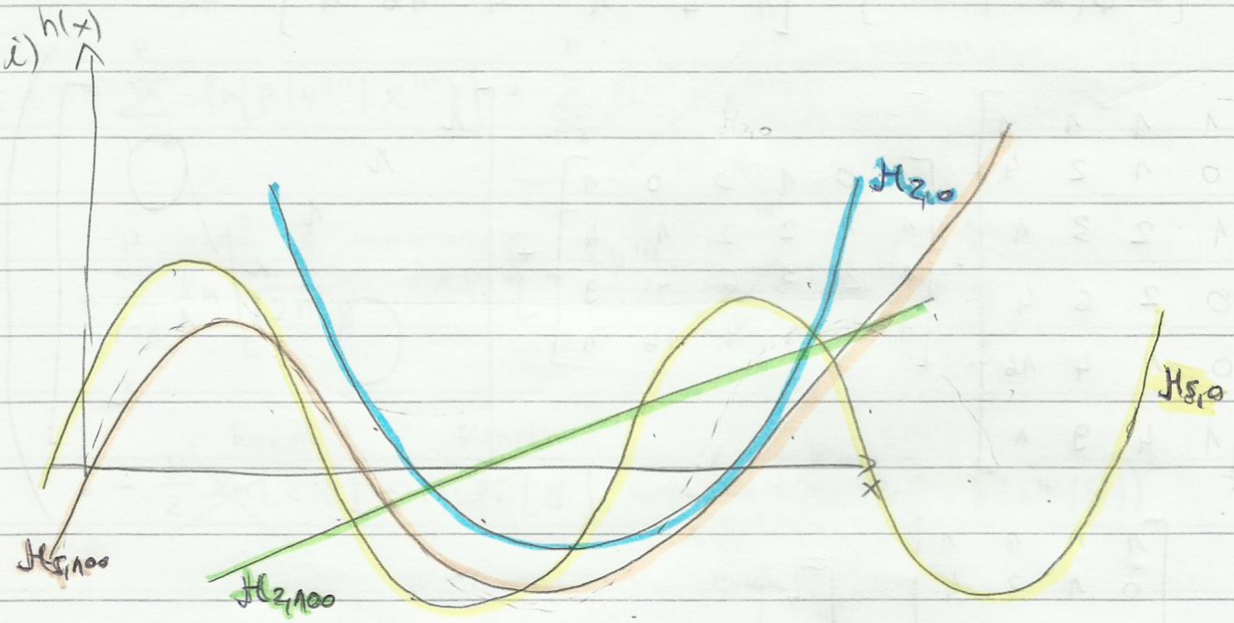
$\boxed{\text{Zadatak 8.}}$

a) $E(\vec{w} \mid D) = \frac{1}{2} \sum_{i=1}^{N} \left(y^{(i)} - h(\vec{x}^{(i)} \mid \vec{w}) \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n} |w_j|^2$

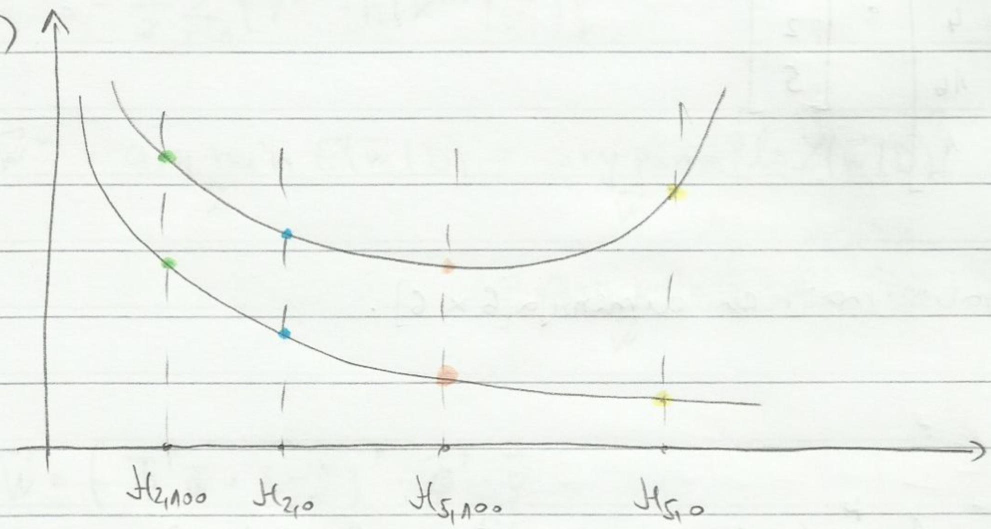
b) $\mathcal{H}_{d,\lambda}$: d - stupanj polinoma

λ - regularizacijski faktor

i)



ii)



e)

Minimizacija L2 regularizirane pogreške istovjetna je minimizaciji negativnog logaritma aposteriorne vjerojatnosti $-\ln p(D|\vec{w}) \cdot p(\vec{w})$.