Strojno učenje – domaća zadaća 9

UNIZG FER, ak. god. 2016./2017.

Zadano: 19. 12. 2016. Rok: 4. 1. 2017.

Napomena: Zadatke možete rješavati samostalno ili u grupi. Ako zadatke rješavate u grupi, pobrinite se da svi članovi grupe pridonose rješenju i da ga naposlijetku svi razumiju. Po potrebi konzultirajte sve dostupne izvore informacija. Rješenja zadataka ponesite na iduće auditorne vježbe. Zabilježite sve nejasnoće i nedoumice, kako bismo ih prodiskutirali.

1. [***Svrha:*** ***Razumjeti model Bayesovog klasifikatora i njegove komponente. Razumjeti što su to generativni modeli, kako se razlikuju od diskriminativnih te koje su njihove prednosti i njihovi nedostatci.***]

   (a) Definirajte model Bayesovog klasifkatora i navedite sve veličine koje se pojavljuju u definiciji modela. Objasnite zašto faktoriziramo brojnik. Objasnite ulogu nazivnika i objasnite kada ga možemo zanemariti.

   (b) Je li taj model parametarski ili neparametarski? Obrazložite odgovor.

   (c) Objasnite zašto Bayesov model nazivamo generativnim i opišite generativnu priču Bayesovog klasifikatora.

   (d) Objasnite razliku između generativnih i diskriminativnih modela te navedite prednosti jednih i drugih.

2. [***Svrha:*** ***Isprobati izračun maksimalne aposteriorne hipoteze i najvjerojatnije hipoteze uz minimizaciju rizika.***] Razmotrimo problem klasifikaciji neželjene el. pošte u klase *spam* ($y = 1$), *important* ($y = 2$) i *normal* ($y = 3$). Neka su apriorne vjerojatnosti tih klasa $P(y = 1) = 0.2$, $P(y = 2) = 0.05$ i $P(y = 3) = 0.75$. Za neku poruku el. pošte $\mathbf{x}$ izglednosti iznose $p(\mathbf{x}|y = 1) = 0.8$ i $p(\mathbf{x}|y = 2) = p(\mathbf{x}|y = 3) = 0.5$. Izračunajte aposteriorne vjerojatnost za svaku od klasa te maksimalnu aposteriornu hipotezu za primjer $\mathbf{x}$.

3. [***Svrha:*** ***Razumjeti faktorizaciju zajedničke vjerojatnosti uz pretpostavku uvjetne nezavisnosti te povezanost toga s induktivnom pristranošću i, posljedično, brojem značajki modela.***]

   (a) Definirajte naivan Bayesov klasifikator i pretpostavku na kojoj se temelji.

   (b) Zašto nam treba pretpostavka o uvjetnoj nezavisnosti značajki te kojoj vrsti induktivne pristranosti ona odgovara?

   (c) Naivan Bayesov klasifikator koristimo za klasifikaciju rukom pisanih znamenki u jednu od deset klasa. Znamenke su prikazane kao vektor binarnih značajki (crno/bijeli slikovni elementi) u matrici s razlučivošću 32×32. Odredite ukupan broj parametara naivnog Bayesovog klasifikatora.

4. [***Svrha: Isprobati na konkretnom primjeru izračun parametara naivnog Bayesovog klasifikatora.***] Naivan Bayesov model želimo upotrijebiti za binarnu klasifikaciju *"Skupo ljetovanje na Jadranu"*. Skup primjera za učenje je sljedeći:

i	Mjesto	Otok	Smještaj	Prijevoz	$y^{(i)}$
1	Istra	da	privatni	auto	da
2	Kvarner	ne	kamp	bus	ne
3	Dalmacija	da	hotel	avion	da
4	Dalmacija	ne	privatni	avion	ne
5	Istra	ne	privatni	auto	da
6	Kvarner	ne	kamp	bus	ne
7	Dalmacija	da	hotel	auto	da

   (a) Izračunajte MLE procjene svih parametara modela te klasificirajte primjere (Istra, ne, kamp, bus) i (Dalmacija, da, hotel, bus).

   (b) Izračunajte Laplaceove (zaglađene) procjene za sve parametre modela te klasificajte nanovo iste primjere.

5. [***Svrha: Razumjeti definiciju uzajamne informacije i način njezina izračuna.***]

   (a) Krenuvši od definicija za entropiju i relativnu entropiju, izvedite mjeru uzajamne informacije $I(X,Y)$ kao Kullback-Leiblerovu divergenciju između zajedničke razdiobe, $P(X,Y)$, i zajedničke razdiobe uz pretpostavku nezavisnosti, $P(X)P(Y)$.

   (b) Izračunajte mjeru uzajamne informacije $I(X,Y)$ za varijable X i Y s razdiobom definiranom u zadatku (1) u zadaći 8. Biste li, temeljem vrijednosti uzajamne informacije, rekli da su varijable X i Y nezavisne?

   (c) Uzajamna informacija nije odozgo ograničena, ali je ograničena odozdo. Primjenom Jensenove nejednakosti, dokažite da vrijedi $I(X,Y) \geqslant 0$.

6. [***Svrha: Razviti intuiciju za model kontinuiranog Bayesovog klasifikatora.***]

   Izrađujemo Bayesov model za klasifikaciju primjera iz $\mathcal{X} = \mathbb{R}$ u tri klase. Učenjem na skupu primjera dobili smo sljedeće parametre modela: $P(y=1) = 0.3$, $P(y=2) = 0.2$, $\mu_1 = -5$, $\mu_2 = 0$, $\mu_3 = 5$, $\sigma_1^2 = 5$, $\sigma_2^2 = 1$, $\sigma_3^2 = 10$. Skicirajte funkcije gustoće vjerojatnosti $p(x|y)$, $p(x,y)$, $p(x)$ i $p(y|x)$.

7. [***Svrha: Razumjeti izvod modela kontinuiranog Bayesovog klasifikatora i osvježiti potrebno znanje matematike.***]

   (a) Krenuvši od izraza (4.29) iz skripte, izvedite model višedimenzijskog Bayesovog klasifikatora s kontinuiranim ulazima s dijeljenom i dijagonalnom kovarijacijskom matricom.

   (b) Napišite broj parametara ovog modela.

   (c) Objasnite zašto je izglednost faktorizirana u produkt univarijatnih razdioba, što odgovara pretpostavci o uvjetnoj nezavisnosti, premda značajke mogu biti nelinearno uvjetno zavisne.

8. [***Svrha: Razviti intuiciju za složenost modela kontinuiranog Bayesovog klasifikatora i shvatiti kako se problem u konačnici svodi na odabir optimalnog modela.***] Želimo izgraditi klasifikator za klasifikaciju brucoša u jednu od dvije klase: $y = 1 \Rightarrow$ *"Završava FER u roku"* i $y = 2 \Rightarrow$ *"Produljuje studij"*. Svaki je primjer opisan sa šest ulaznih varijabli: prosjek ocjena 1.–4. razreda (četiri varijable), bodovi državne mature iz matematike te bodovi državne mature iz fizike. Raspolažemo trima modelima: modelom $\mathcal{H}_1$ s dijeljenom kovarijacijskom matricom, modelom $\mathcal{H}_2$ s dijagonalnom (i dijeljenom) kovarijacijskom matricom i modelom $\mathcal{H}_3$ s izotropnom kovarijacijskom matricom.

   (a) Koliko svaki od ova tri modela ima parametara?

   (b) Za koji od ova tri modela očekujete da će najbolje generalizirati u ovom konkretnom slučaju (uzmite u obzir prirodu problema i očekivane odnose između značajki)? Zašto?

   (c) Nacrtajte skicu funkcije empirijske pogreške i pogreške generalizacije i naznačite na njoj točke koje označavaju navedenim trima modelima.

   (d) Kako biste u praksi odredili koji ćete model upotrijebiti?

9. [***Svrha: Razumjeti vezu između Bayesovog klasifikatora i logističke regresije, odnosno probabilističku interpretaciju logističke regresije. Razumjeti razliku u broju parametara između diskriminativnog i generativnog modela te utjecaj broja klasa i broja primjera na taj odnos.***]

   (a) Izvedite model logističke regresije krenuvši od generativne definicije za $P(y = 1|\mathbf{x})$. Izvod napravite korak po korak te se uvjerite da možete obrazložiti svaki korak u izvodu. Napišite sve pretpostavke koje ste ugradili u izvod.

   (b) Model logističke regresije koristimo za binarnu klasifikaciju primjera s $n = 100$ značajki. Odredite broj parametara modela logističke regresije te njemu odgovarajućeg generativnog modela.

   (c) Izračunajte broj parametara za isti slučaj, ali sa $K = 5$ klasa.

   (d) Pretpostavite da klasificiramo u $K = 10$ klasa. Izračunajte koliko velika mora biti dimenzija prostora značajki n, a da bi se logistička regresija isplatila jer ima manje parametara od odgovarajućega generativnog modela.

9. DOMAĆA ZADAĆA

(1.) a)

$$p(y|x) = \frac{p(x|y)\,p(y)}{\sum_{y'} p(x|y')\,p(y')}$$

$p(y|x)$ – APOSTERIORNA VJEROJATNOST; VJEROJATNOST DA PRIMJER X PRIPADA KLASI Y

$p(x|y)$ – IZGLEDNOST KLASE; VJEROJATNOST PRIMJERA U KLASI

$p(y)$ – APRIORNA VJEROJATNOST KLASE

FAKTORIZACIJA $p(x,y)$ NA $p(x|y)p(y)$ OMOGUĆUJE MODELIRANJE SLOŽENIH DISTRIBUCIJA; $p(x|y)$ I $p(y)$ MODELIRAMO ZASEBNO.

ULOGA NAZIVNIKA JE NORMALIZIRATI VRIJEDNOST NA JEDINIČNI INTERVAL ČIME DOBIVAMO VJEROJATNOSNU INTERPRETACIJU. MOŽEMO GA ODBACITI PRI PRORAČUNU MAP HIPOTEZE JER JE ON ISTI ZA SVE KLASE PA JE MAKSIMIZACIJA BROJNIKA EKVIVALENTNA MAKSIMIZACIJI CIJELOG IZRAZA.

b) PARAMETARSKI JER NE OVISI O BROJU PRIMJERA N ALI OVISI O BROJU ZNAČAJKI n (SVAKU ZNAČAJKU MODELIRAMO NEKOM VJEROJATNOSNOM DISTRIBUCIJOM ČIJI PARAMETRI NE OVISE O BROJU PRIMJERA).

c) NAZIVAMO GA GENERATIVNIM ZATO ŠTO MODELIRA ZAJEDNIČKU DISTRIBUCIJU $p(x,y)$, A NA TEMELJU NJE

RAČUNAMO $p(y|x)$ ILI NEKU DRUGU DISTRIBUCIJU OD INTERESA. EFEKTIVNO MODELIRAMO NASTAJANJE PODATAKA $\{(x^{(i)}, y^{(i)})\}_i$, ŠTO SE NAZIVA GENERATIVNA PRIČA.

GENERATIVNA PRIČA BAYESOVOG KLASIFIKATORA $P(x,y) = p(x|y)P(y)$ UKAZUJE DA PRVO ODABIREMO OZNAKU PREMA $P(y)$, A POTOM ODABIREMO PRIMJER PREMA $P(x|y)$.

d) GENERATIVNI MODELI VJEROJATNOST $P(y|x)$ MODELIRAJU POSREDNO PREKO ZAJEDNIČKE DISTRIBUCIJE $p(x,y)$, DOK DISKRIMINATIVNI MODELI VJEROJATNOST $P(y|x)$ MODELIRAJU IZRAVNO. PREDNOSTI GENERATIVNIH SU LAKA UGRADNJA STRUČNOG ZNANJA, INTERPRETABILNOST REZULTATA, POUZDANOST KLASIFIKACIJE TE NALAŽENJE STRŠEĆIH VRIJEDNOSTI, A MANE SU VELIK BROJ POTREBNIH PRIMJERA ZA UČENJE TE NEPOTREBNA SLOŽENOST MODELIRANJA.

2.

$P(y=1)=0.2 \qquad p(x|y=1)=0.8$

$P(y=2)=0.05 \qquad p(x|y=2)=0.5$

$P(y=3)=0.75 \qquad p(x|y=3)=0.5$

$h_1(x) = p(x|y=1)P(y=1) = 0.8 \cdot 0.2 = 0.16$

$h_2(x) = 0.05 \cdot 0.5 = 0.025$

$h_3(x) = 0.75 \cdot 0.5 = 0.375$

$P(x) = \sum_{y'} P(x|y')P(y') = 0.16 + 0.025 + 0.375 = 0.56$

MAP HIPOTEZA:

$h(x) = \underset{y}{\arg\max}\ p(x|y)P(y) = \underline{\underline{3}}$ (KLASA „NORMAL")

$$P(y=1 \mid x) = \frac{h_1(x)}{P(x)} = \frac{0.16}{0.56} = 0.2857$$

$$P(y=2 \mid x) = \frac{0.025}{0.56} = 0.04464$$

$$P(y=3 \mid x) = \frac{0.375}{0.56} = 0.6696$$

3. a) NAIVAN BAYESOV KLASIFIKATOR:

$$h(x_1, \ldots, x_n \mid y) = \underset{j}{\operatorname{argmax}}\ P(y) \cdot \prod_{i=1}^{n} P(x_i \mid y)$$

TEMELJI SE NA PRETPOSTAVCI DA SU SVE ZNAČAJKE MEĐUSOBNO UVJETNO NEZAVISNE TJ. DA VRIJEDI $P(x_i \mid x_j, y) = P(x_i \mid y)$

b) TREBA NAM ZATO ŠTO AKO MODELIRAMO $p(x \mid y)$ KAO KATEGORIČKU RAZDIOBU OD X IMAMO PREVIŠE PARAMETARA, A TIME I PRESLOŽEN MODEL ODNOSNO NEMA GENERALIZACIJE. TO JE PRISTRANOST JEZIKA, OGRANIČAVAMO MODEL.

c) $K = 10$, $n = 32 \cdot 32 = 1024$, $K_k = 2$, $k = 1 \ldots N$

$$\text{BROJ PARAMETARA} = \underbrace{K-1}_{\substack{\text{MODELIRANJE} \\ P(y) \text{ SVAKE} \\ \text{KLASE}}} + \underbrace{\sum_{i=1}^{n} (K_k - 1) K}_{\substack{\text{MODELIRANJE} \\ P(X \mid Y) \text{ ZA} \\ \text{SVAKU ZNAČAJKU}}} = 9 + 1024 \cdot 10 = 10249$$

4. a) $P(x_1 = \text{ISTRA}, y = \text{DA}) = 1/2$, $P(x_1 = \text{ISTRA}, y = \text{NE}) = 0$

$P(x_1 = \text{KVARNER}, y = \text{DA}) = 0$, $P(x_1 = \text{KVARNER}, y = \text{NE}) = 2/3$

$P(x_1 = \text{DALMACIJA}, y = \text{DA}) = 1/2$, $P(x_1 = \text{DALMACIJA}, y = \text{NE}) = 1/3$

→

$P(X_2=DA,\ y=DA)=3/4,\quad P(X_2=DA,\ y=NE)=0$
$P(X_2=NE,\ y=DA)=1/4,\quad P(X_2=NE,\ y=NE)=1$

$P(X_3=PRIVATNI,\ y=DA)=1/2,\quad P(X_3=PRIVATNI,\ y=NE)=1/3$
$P(X_3=KAMP,\ y=DA)=0,\quad P(X_3=KAMP,\ y=NE)=2/3$
$P(X_3=HOTEL,\ y=DA)=1/2,\quad P(X_3=HOTEL,\ y=NE)=0$

$P(X_4=AUTO,\ y=DA)=3/4,\quad P(X_4=AUTO,\ y=NE)=0$
$P(X_4=BUS,\ y=DA)=0,\quad P(X_4=BUS,\ y=NE)=2/3$
$P(X_4=AVION,\ y=DA)=1/4,\quad P(X_4=AVION,\ y=DA)=1/3$

$h(ISTRA, NE, KAMP, BUS)=\underset{j}{\operatorname{argmax}}\ P(y)\prod_{k=1}^{n}P(X_k|y)$

$y=DA:\quad P(y)\prod_{k=1}^{n}P(X_k|y)=\frac{4}{7}\cdot\left(\frac{1}{2}\cdot\frac{1}{4}\cdot 0\cdot 0\right)=0$

$y=NE:\quad P(y)\prod_{k=1}^{n}P(X_k|y)=\frac{3}{7}\cdot\left(0\cdot 1\cdot\frac{2}{3}\cdot\frac{2}{3}\right)=0$

NE MOŽEMO DONIJETI ODLUKU!

$h(DALMACIJA, DA, HOTEL, BUS)=\underset{j}{\operatorname{argmax}}\ P(y)\prod_{k=1}^{n}P(X_k|y)$

$y=DA:\quad P(y=DA)\prod_{k=1}^{n}P(X_k|y=DA)=\frac{4}{7}\cdot\left(\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{1}{2}\cdot 0\right)=0$

$y=NE\quad P(y=NE)\prod_{k=1}^{n}P(X_k|y=NE)=\frac{3}{7}\cdot\left(\frac{1}{3}\cdot 0\cdot 0\cdot\frac{2}{3}\right)=0$

NE MOŽEMO DONIJETI ODLUKU!

b) $P(ISTRA|DA)=3/7,\quad P(ISTRA|NE)=1/6$
$P(KVARNER|DA)=1/7,\quad P(KVARNER|NE)=1/2$
$P(DALMACIJA|DA)=3/7,\quad P(DALMACIJA|NE)=1/3$

$P(DA|DA)=2/3,\quad P(DA|NE)=1/5$
$P(NE|DA)=1/3,\quad P(NE|NE)=4/5$

$P(PRIVATNI \mid DA) = 3/7$, $P(PRIVATNI \mid NE) = 1/3$
$P(KAMP \mid DA) = 1/7$, $P(KAMP \mid NE) = 1/2$
$P(HOTEL \mid DA) = 3/7$, $P(HOTEL \mid NE) = 1/6$

$P(AUTO \mid DA) = 4/7$, $P(AUTO \mid NE) = 1/6$
$P(BUS \mid DA) = 1/7$, $P(BUS \mid NE) = 1/2$
$P(AVION \mid DA) = 2/7$, $P(AVION \mid NE) = 1/3$

$$h(DA \mid ISTRA, NE, KAMP, BUS) = \frac{4}{7}\left(\frac{3}{7}\cdot\frac{1}{3}\cdot\frac{1}{7}\cdot\frac{1}{7}\right) = 0.001666$$

$$h(NE \mid ISTRA, NE, KAMP, BUS) = \frac{3}{7}\left(\frac{1}{6}\cdot\frac{4}{5}\cdot\frac{1}{2}\cdot\frac{1}{2}\right) = 0.0143$$

$$\boxed{h(x) = NE}$$

$$h(DA \mid DALMACIJA, DA, HOTEL, BUS) = \frac{4}{7}\left(\frac{3}{7}\cdot\frac{2}{3}\cdot\frac{3}{7}\cdot\frac{1}{7}\right) = 0.01$$

$$h(NE \mid DALMACIJA, DA, HOTEL, BUS) = \frac{3}{7}\left(\frac{1}{3}\cdot\frac{1}{5}\cdot\frac{1}{6}\cdot\frac{1}{2}\right) = 0.00238$$

$$\boxed{h(x) = DA}$$

5.) a) ENTROPIJA:

$$H(P) = -\sum_x P(x) \ln P(x)$$

RELATIVNA ENTROPIJA:

$$H(P,Q) - H(P) = -\sum_x P(x) \ln Q(x) + \sum_x P(x) \ln P(x) =$$

$$= -\sum_x P(x) \ln \frac{P(x)}{Q(x)} = D_{KL}(P \| Q)$$ → KULLBACK-LEIBLEROVA DIVERGENCIJA

$$D_{KL}(P(x,y) \| P(x)P(y)) = -\sum_{x,y} P(x,y) \cdot \ln \frac{P(x,y)}{P(x)P(y)} = I(X,Y)$$

→ UZAJAMNA INFORMACIJA

b) Iz DZ 8:

$$P(1,1)=0.2 \qquad P(2,1)=0.05$$
$$P(1,2)=0.05 \qquad P(2,2)=0.3$$
$$P(1,3)=0.3 \qquad P(2,3)=0.1$$

$P(X=1)=0.55$, $P(X=2)=0.45$

$P(Y=1)=0.25$, $P(Y=2)=0.35$ $P(Y=3)=0.4$

$$I(X,Y)=P(1,1)\ln\frac{P(1,1)}{P(X=1)P(Y=1)}+P(1,2)\ln\frac{P(1,2)}{P(X=1)P(Y=2)}+\dots)$$

$$I(X,Y)=\{MATLAB\}=0.1946 \rightarrow I(X,Y)>0 \rightarrow$$ NISU NEZAVISNE

c) JENSENOVA NEJEDNAKOST:

ZA KONVEKSNU FUNKCIJU $f(x)$ VRIJEDI:

$$f(E[X])\leq E[f(x)]$$

$$I(X,Y)=\sum_{X,Y}P(X,Y)\ln\frac{P(X,Y)}{P(X)P(Y)}\geq 0$$

NEKA $f=\ln$, PO JENSENU ONDA VRIJEDI:

$$\ln E\left[\frac{P(X,Y)}{P(X)P(Y)}\right]\leq E\left[\ln\frac{P(X,Y)}{P(X)P(Y)}\right]$$

$$E\left[\ln\frac{P(X,Y)}{P(X)P(Y)}\right]=\sum_{X,Y}P(X,Y)\ln\frac{P(X,Y)}{P(X)P(Y)}=-\sum_{X,Y}P(X,Y)\ln\frac{P(X)P(Y)}{P(X,Y)}$$

$$\leq \ln\sum_{X,Y}P(X,Y)\cdot\frac{P(X)P(Y)}{P(X,Y)}=\ln\sum_{X,Y}P(X)P(Y)=\ln(1)=\underline{0}$$

$$-\sum_{X,Y}P(X,Y)\ln\frac{P(X)P(Y)}{P(X,Y)}\leq 0$$

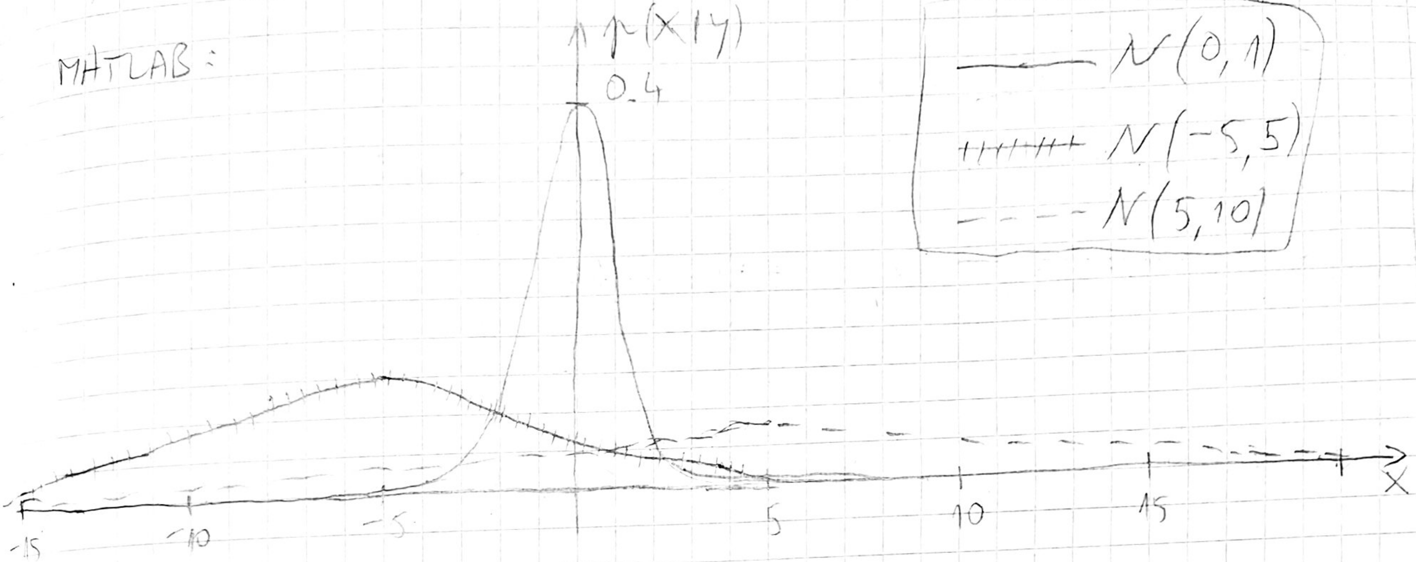
$$\boxed{I(X,Y)\geq 0}$$

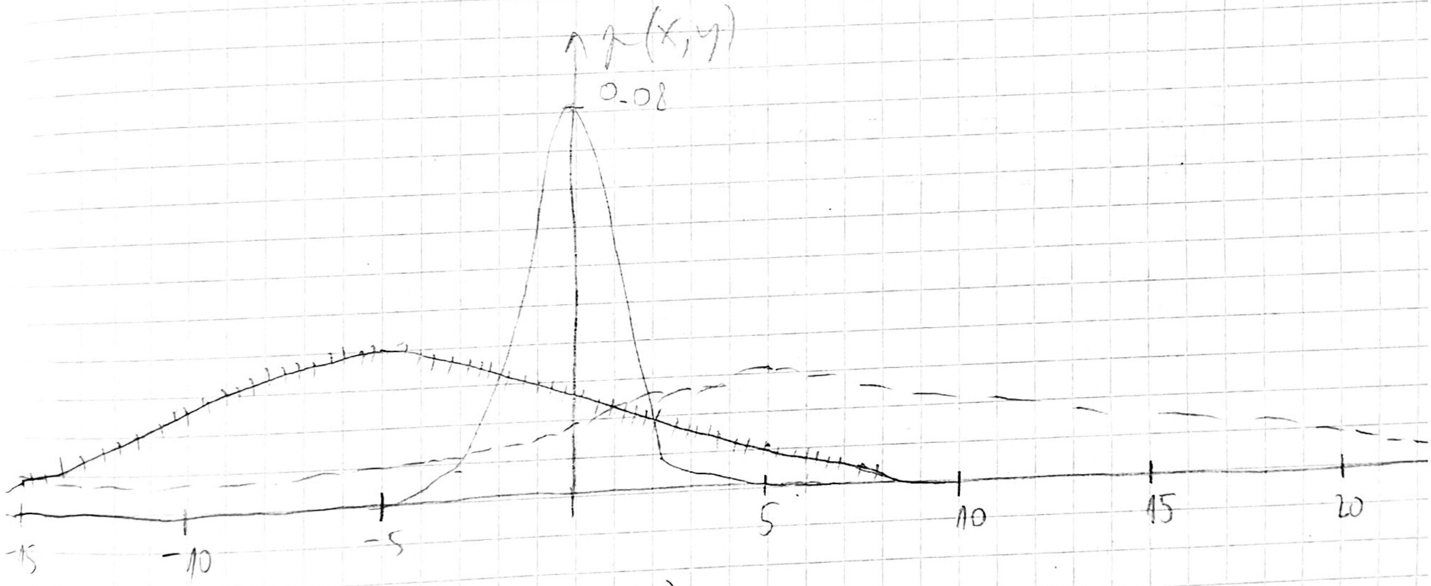
NE ULAZI U ISPIT!

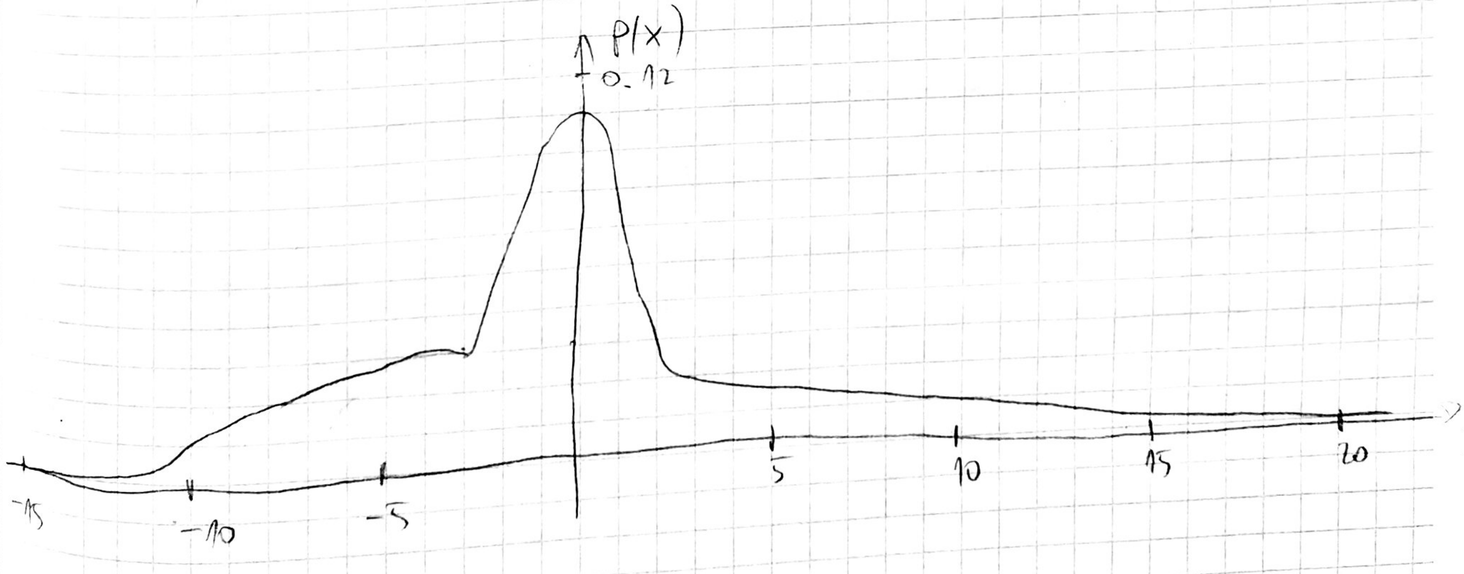
6. $x \in \mathbb{R}$ $\quad P(y=1)=0.3$, $P(y=2)=0.2$ $\rightarrow$ $P(y=3)=0.5$

$\mu_1=-5$, $\mu_2=0$, $\mu_3=5$, $\sigma_1^2=5$, $\sigma_2^2=1$, $\sigma_3^2=10$

MATLAB:







$p(y|x) \rightarrow$ MATLAB

7\. a) $h_j(x) = x^T \Sigma^{-1} \mu_j - \frac{1}{2} \mu_j^T \Sigma^{-1} \mu_j + \ln P(C_j)$ (4.29)

UZ PRETPOSTAVKU DA VARIJABLE NISU KORELIRANE KORISTIMO DIJAGONALNU KOVARIJACIJSKU MATRICU, $\Sigma = \mathrm{diag}(\sigma_i^2)$. ONDA JE $\Sigma^{-1} = \mathrm{diag}(1/\sigma_i^2)$, A $|\Sigma| = \prod_i \sigma_i$. MULTIVARIJATNA GAUSSOVA GUSTOĆA DEGENERIRA U PRODUKT UNIVARIJANTNIH GAUSSOVIH RAZDIOBI:

$$p(x|C_j) = \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} \sigma_i} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu_{ij})^2}{\sigma_i^2}\right\}$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left\{-\frac{1}{2} \frac{(x_i - \mu_{ij})^2}{\sigma_i^2}\right\}$$

$$h_j(x) = -\frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \mu_{ij}}{\sigma_i}\right)^2 + \ln P(C_j)$$

b) BROJ PARAMETARA: $\underbrace{n}_{\text{VARIJANCE}} + \underbrace{n \cdot K}_{\text{CENTROIDI}} + \underbrace{K - 1}_{P(C_j)}$

c) ZATO DA MODEL BUDE JEDNOSTAVNIJI, A TIME I MANJE SKLON PRENAUČENOSTI. UMJESTO $O(n^2)$ OVISNOSTI BROJA PARAMETARA O BROJU ZNAČAJKI DOBIVAMO $O(n)$ OVISNOST TE JE BROJ PARAMETARA MODELA OČITO ZNATNO SMANJEN.

⑧ a) H_1 – DIJELJENA KOVARIJACIJSKA MATRICA

BROJ PARAMETARA: $\frac{n}{2}(n+1) + nK + K - 1$

$= 34$ //

$\boxed{\begin{array}{c} K=2 \\ n=6 \end{array}}$

H_2 – DIJAGONALNA KOVARIJACIJSKA MATRICA

BROJ PARAMETARA: $n + nK + K - 1$

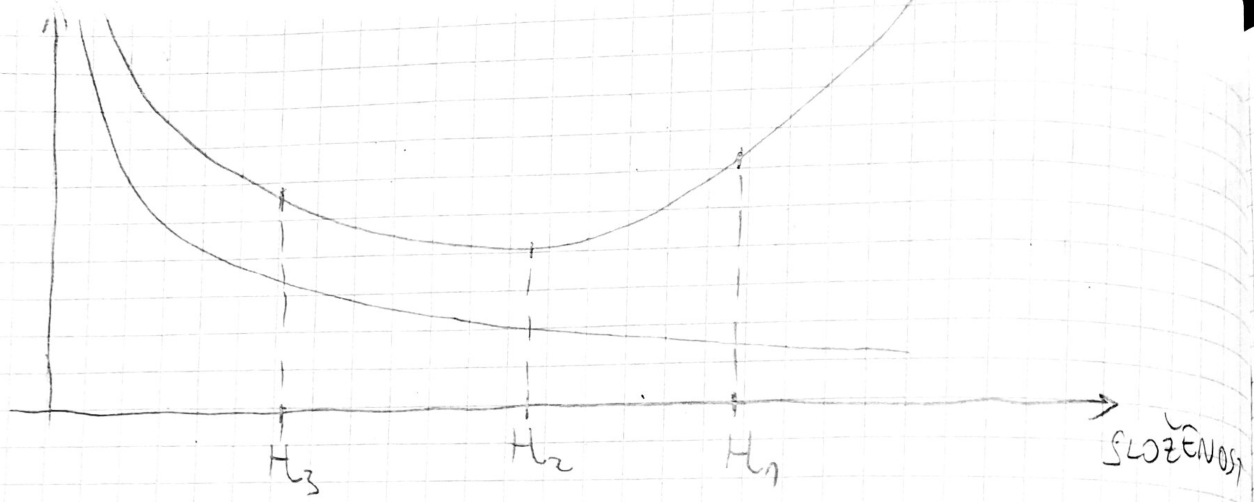
$= 19$ //

H_3 – IZOTROPNA KOVARIJACIJSKA MATRICA

BROJ PARAMETARA: $Kn + K$

$= 13$ //

b) OČEKUJEM DA ĆE MODEL S DIJAGONALNOM KOVARIJACIJSKOM MATRICOM DATI NAJBOLJE REZULTATE.
H_1 MODELIRA MEĐUKORELIRANOST, A BUDUĆI DA SU ZNAČAJKE MEĐUSOBNO KORELIRANE (NETKO TKO JE IMAO DOBRE OCJENE VJEROJATNO ĆE IMATI I PUNO BODOVA NA MATURI) MOŽE DOĆI DO LOŠE KONDICIONIRANOSTI MATRICE Σ^{-1}!
TAKOĐER IMA I NAJVIŠE PARAMETARA I SKLON JE PRENAUČENOSTI.
H_3 SVE VARIJANCE MODELIRA JEDNAKIMA ŠTO ĆE LAKO DOVESTI DO PODNAUČENOSTI; VELIKA JE RAZLIKA U SKALAMA ZA OCJENE (1–5) I BROJ BODOVA NA MATURI (0–100).
H_2 NE MODELIRA MEĐUKORELIRANOST ZBOG ČEGA GA JE TEŽE PRENAUČITI NEGO H_1, A KORISTI NORMALIZIRANE EUKLIDSKE UDALJENOSTI ZBOG ČEGA JE NEOSJETLJIV NA RAZLIKE U VARIJANCI IZMEĐU POJEDINIH DIMENZIJA, ZA RAZLIKU OD H_3.

c) 

d) UNAKRSNOM PROVJEROM.

9. a) APOSTERIORNA VJEROJATNOST ZA KONTINUIRANI BAYESOV KLASIFIKATOR:

$$P(y=1\mid x) = \frac{p(x\mid y=1)\,P(y=1)}{p(x\mid y=1)P(y=1) + p(x\mid y=2)P(y=2)} =$$

$$= \frac{1}{1+\frac{p(x\mid y=2)\,P(y=2)}{p(x\mid y=1)\,P(y=1)}} = \frac{1}{1+\exp\left(\ln\left(\frac{p(x\mid y=2)\,P(y=2)}{p(x\mid y=1)\,P(y=1)}\right)\right)}$$

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

GDJE:

$$\alpha = \ln\frac{p(x\mid y=1)\,P(y=1)}{p(x\mid y=2)\,P(y=2)} = \underbrace{\ln(p(x\mid y=1)P(y=1))}_{h_1(x)} - \underbrace{\ln(p(x\mid y=2)P(y=2))}_{h_2(x)}$$

AKO PRETPOSTAVIMO DIJELJENU KOVARIJACIJSKU MATRICU:

$$\alpha = h_1(x) - h_2(x)$$

$$\alpha = \underbrace{x^T \Sigma^{-1}(\mu_1 - \mu_2)}_{w} \underbrace{- \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2 + \ln\frac{P(y=1)}{P(y=2)}}_{w_0}$$

↳ SVE ŠTO JE UZ x^T SMO OZNAČILI (SUPSTITUIRALI) KAO VEKTOR w, A SVE SLOBODNE ČLANOVE KAO w_0

DAKLE MOŽE SE ZAPISATI:

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

UZ PRETPOSTAVKU DIJELJENE KOVARIJACIJSKE MATRICE Σ POKAZANO JE DA JE BAYESOV KLASIFIKATOR JEDNAK LOGISTIČKOJ REGRESIJI.

b) $n = 100$

LOGISTIČKA REGRESIJA: BROJ PARAMETARA $= n+1 = 101$ //

BAYESOV KLASIFIKATOR UZ DIJELJENU KOV. MAT. Σ:

BROJ PARAMETARA $= \frac{n}{2}(n+1) + 2n + 1 = 5251$ //

c) $n = 100$, $K = 5$

LOGISTIČKA REGRESIJA UZ OVR SHEMU VIŠEKLASNE KLASIFIKACIJE: BROJ PARAMETARA $= K(n+1) = 505$ //

BAYESOV KLASIFIKATOR UZ DIJELJENU Σ

BROJ PARAMETARA $= \frac{n}{2}(n+1) + nK + K - 1 = 5554$ //

d) $K = 10$

LOGREG UZ OVO SHEMU | BAYES

$$\binom{K}{2}(m+1) \quad = \quad \frac{m}{2}(m+1) + mK + K - 1$$

$$\frac{10!}{2!\,8!}(m+1) = \frac{m^2}{2} + \frac{m}{2} + 10m + 9$$

$$45m + 45 = \frac{m^2}{2} + 10.5m + 9$$

$$\frac{m^2}{2} - 34.5m - 36 = 0$$

$$m_1 = -1, \quad \boxed{m_2 = 70}$$

$\hookrightarrow$ TEK ZA 70 ZNAČAJKI SE VIŠE ISPLATI LOGISTIČKA REGRESIJA (UZ OVO SHEMU VIŠEKLASNE KLASIFIKACIJE).