Strojno učenje 1 – Domaća zadaća 4

*Domaća zadaća sadrži **6 pitanja** i ukupno nosi najviše 6 bodova (skalirano na 2 boda na predmetu). Točan odgovor nosi 1 bod. Za razliku od bodovanja ispita, netočni odgovori ne nose negativne bodove. Prije nego što krenete rješavati ove zadatke preporučujemo da riješite sve zadatke iz dijela "Zadatci za učenje" za sve nastavne cjeline obuhvaćene ovom zadaćom.*

Upozorenje: *Za svaki zadatak na koji odgovorite, morate predati i ručno ispisan postupak. Ako to ne učinite čak i za samo jedan zadatak, gubite sve bodove prikupljene kroz aktivnost na domaćim zadaćama.*

Napomena: *Ova domaća zadaća je personalizirana. Svaki student dobiva jedinstvenu varijantu zadataka.*

14. Procjena parametara II

14.1 (N) U beta-Bernoullijevom modelu, apriornu vjerojatnost parametra μ modeliramo beta-distribucijom. Gustoća vjerojatnosti i mod (maksimizator) beta-distribucije su:

$$p(\mu|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\mu^{\alpha-1}(1-\mu)^{\beta-1} \qquad \mu^* = \frac{\alpha-1}{\alpha+\beta-2}$$

Neka $\alpha = \beta = 2$. Računamo MAP-procjenu za parametar μ Bernoulijeve varijable. To radimo na dva uzorka, $\mathcal{D}_1 = (N_1, m_1)$ i $\mathcal{D}_2 = (N_2, m_2)$, koji nam pristižu jedan za drugim. Pritom koristimo svojstvo konjugatnosti, na način da aposteriornu gustoću vjerojatnosti izračunatu na temelju prvog uzorka koristimo kao apriornu gustoću vjerojatnosti pri procjeni na temelju drugog uzorka. U prvom uzorku, veličine $N_1 = 80$, Bernoullijeva varijabla realizirana je s vrijednošću $y = 1$ ukupno $m_1 = 14$ puta. U drugom uzorku, veličine $N_2 = 20$, Bernoullijeva varijabla realizirana je s vrijednošću $y = 1$ ukupno $m_2 = 11$ puta. Izračunajte MAP-procjene za parametar μ na temelju ova dva uzorka. **Koliko iznosi promjena u procjeni za μ između prve i druge procjene?**

$\boxed{\text{A}}$ $+0.169$ $\boxed{\text{B}}$ -0.083 $\boxed{\text{C}}$ $+0.072$ $\boxed{\text{D}}$ -0.191

14.2 (N) Zadan je uzorak kontinuirane slučajne varijable, $\mathcal{D} = \{10, 5, 6, 0\}$. Pretpostavljamo razdiobu $\mathcal{N}(\mu, \sigma^2)$, čija je gustoća vjerojatnosti:

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}$$

Za procjenu parametra μ koristimo procjenitelj $\hat{\mu}_{\text{MLE}}$, a za procjenu parametra σ^2 procjenitelj $\hat{\sigma}^2_{\text{UB}}$, koji je nepristran procjenitelj izveden odgovarajućom korekcijom iz procjenitelja $\hat{\sigma}^2_{\text{MLE}}$. Izračunajte log-izglednost procijenjenih parametara na skupu $\mathcal{D}$, tj. $\ln\mathcal{L}(\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{UB}}|\mathcal{D})$. **Koliko iznosi ta log-izglednost?**

$\boxed{\text{A}}$ -10.832 $\boxed{\text{B}}$ -13.163 $\boxed{\text{C}}$ -12.873 $\boxed{\text{D}}$ -15.200

15. Bayesov klasifikator

15.1 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 2$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene i dijagonalne kovarijacijske matrice. Izglednost klase s oznakom $y = j$ definirana je multivarijantnom Gaussovom gustoćom vjerojatnosti:

$$p(\mathbf{x}|y=j) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_j|^{1/2}}\exp\Big\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\mathbf{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\Big\}$$

Model treniramo na skupu podataka od $N = 7$ primjera:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((-1,-2),0), ((0,0),0), ((1,2),0), ((3,-1),1), ((4,-1),1), ((4,1),1), ((5,1),1)\big\}$$

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_0(\mathbf{x})$ za primjer $\mathbf{x} = (0, 0)$?**

$\boxed{\text{A}}$ -2.75 $\boxed{\text{B}}$ -3.03 $\boxed{\text{C}}$ -3.84 $\boxed{\text{D}}$ -4.13

15.2 (P) Gaussov Bayesov klasifikator koristimo za klasfikaciju jednodimenzijskih podataka u tri klase. Procijenjene izglednosti klasa su $p(x|y=1) = \mathcal{N}(-10, 2)$, $p(x|y=2) = \mathcal{N}(2, 2)$ i $p(x|y=3) = \mathcal{N}(8, 2)$, a procijenjene apriorne vjerojatnosti klasa su $P(y=1) = P(y=2) = 2/5$ i $P(y=3) = 1/5$. Međutim, nakon što smo naučili ovaj model, zaključili smo da na ispitnom skupu postoji pomak u distribuciji podataka u odnosu na skup za učenje te da zbog toga model ne generalizira dobro. Zaključili smo da se ovo može ispraviti tako da se naučeni model malo izmijeni, i to tako da se varijanca izglednosti klase $y=1$ postavi na 1 i da se apriorne vjerojatnosti klasa ujednače, $P(y=1) = P(y=2) = P(y=3) = 1/3$. Skicirajte gustoće zajedničke vjerojatnosti naučenog i izmijenjenog modela. Neka su h_1 i h_2 MAP-hipoteze prvog i drugog modela, te neka su a i b pozitivne konstante. Razmotrite segment ulaznog prostora za koji vrijedi $-10 \leq x \leq 10$. **Na kojim se dijelovima tog segmenta ulaznog prostora MAP-hipoteze prvog i drugog modela razlikuju?**

A $[-4-a, -4] \cup [5, 5+b]$ B $[-4-a, 5+b]$ C $[-4, -4+a] \cup [5-b, 5]$ D $[-4-a, -4+b]$

16. Bayesov klasifikator II

16.1 (N) Treniramo polunaivan Bayesov klasifikator sa $n=3$ binarne varijable, x_1, x_2 i x_3. Zajednička vjerojatnost tih triju varijabli definirana je sljedećom tablicom:

	$x_3=0$		$x_3=1$	
	$x_2=0$	$x_2=1$	$x_2=0$	$x_2=1$
$x_1=0$	0.1	0.0	0.0	0.1
$x_1=1$	0.1	0.4	0.1	0.2

Prije treniranja klasifikatora, koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Odlučili smo združiti onaj par varijabli koje imaju uzajamnu informaciju veću od 0.01. Ako to vrijedi za dva para varijabli, onda ćemo sve tri varijable združiti u jedan faktor. Izračunajte uzajamne informacije između svih parova varijabli te odredite koje varijable ćemo združiti u zajedničke faktore prema gornjem pravilu. **Kako glasi faktorizacija zajedničke vjerojatnosti tog polunaivnog Bayesovog klasifikatora?**

A $P(y)P(x_1, x_3|y)P(x_2|y)$ C $P(y)P(x_1|y)P(x_2|y)P(x_3|y)$
B $P(y)P(x_1, x_2|y)P(x_3|y)$ D $P(y)P(x_1, x_2, x_3|y)$

16.2 (P) Uz određene restrikcije na parametre Gaussovog Bayesovog klasifikatora, Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n=200$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

A 40400 B 20300 C 45450 D 90600