

## Pismeni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.) – NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a  $1/3$  boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

### Osnovni koncepti i linearna regresija (4 pitanja)

- 1** (T) Svaki model strojnog učenja ima neku induktivnu pristranost. Što je induktivna pristranost?
- ☐ A Minimalan skup pretpostavki koje, uz skup za učenje, jednoznačno određuju klasifikaciju svakog primjera
- ☐ B Kriterij koji, na temelju modela, jednoznačno određuje hipotezu sa minimalnom empirijskom pogreškom
- ☐ C Svaka pretpostavka koja jednoznačno određuje model na temelju hipoteze i skupa za učenje
- ☐ D Odstupanje procjene parametra na temelju podataka u odnosu na pravu vrijednost parametra u populaciji
- 2** (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable  $y$  o zavisnoj varijabli  $x$ . Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka  $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_i = \{(0, 0), (2, 0), (3, 2), (5, 2)\}$ . Neka je  $h$  hipoteza koju dobivamo treniranjem modela te neka je  $L^i$  gubitak hipoteze  $h$  na primjeru  $x^{(i)}$ , tj.  $L^i = L(y^{(i)}, h(x^{(i)}))$ . Što vrijedi za gubitke hipoteze na pojedinim primjerima?
- ☐ A  $L^1 = L^4 = 1, L^2 < L^3$       ☐ C  $L^1 = L^4 < L^2 = L^3$
- ☐ B  $L^1 = L^3 = 0, L^2 = L^4 < 1$       ☐ D  $L^1 = L^2 = 1 < L^3 < L^4$
- 3** (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke  $x_1$ – $x_4$ ), prosjek ocjena sva četiri razreda ( $x_5$ ) te uspjeh iz matematike ( $x_6$ ) i fizike ( $x_7$ ) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr.  $x_1x_2$ ) i interakcije trojki (npr.  $x_1x_2x_3$ ) između svih značajki koje koristimo. Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?
- ☐ A 48    ☐ B 75    ☐ C 38    ☐ D 63
- 4** (N) Raspoložemo sljedećim skupom primjera u dvodimenzijaskome ulaznom prostoru:
- $$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2, -3), 5), ((2, 3), -10), ((-3, 5), 5), ((5, 0), 20)\}$$
- Na ovom skupu gradijentnim spustom trenirali smo  $L_1$ -regularizirani model linearne regresije sa  $\lambda = 2$ . Dobili smo težine  $\mathbf{w} = (4.568, 0.746, -0.550)$ . Koliko iznosi  $L_1$ -regularizirana pogreška  $E(\mathbf{w}|\mathcal{D})$ ?
- ☐ A 230.98    ☐ B 165.89    ☐ C 31.70    ☐ D 191.95

### Linearni klasifikacijski modeli (3 pitanja)

- 5** (T) Postoji veza između logističke regresije i modela neuronske mreže. Koja je veza između ta dva modela?
- ☐ A Multinomijalna logistička regresija s aktivacijskom funkcijom softmax istovjetna je dvoslojnoj neuronskoj mreži sa više neurona u izlaznom sloju
- ☐ B Logistička regresija koja kao adaptivne bazne funkcije koristi logističku regresiju istovjetna je neuronskoj mreži sa sigmoidnom aktivacijskom funkcijom
- ☐ C Neuronska mreža optimirana algoritmom propagacije pogreške unazad istovjetna je logističkoj regresiji optimiranoj stohastičkim gradijentnim spustom
- ☐ D Logistička regresija s linearnim jezgrenim funkcijama istovjetna je neuronskoj mreži sa linearnom aktivacijskom funkcijom i kvadratnom funkcijom pogreške

- 6 (N) Raspoložemo sljedećim skupom za učenje u dvodimenzijaskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijaski prostor značajki, koja je definirana na sljedeći način:

$$\phi(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, 2, -3, 0)$$

**Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

- ☐ A 32   ☐ B 16   ☐ C 22   ☐ D 6

- 7 (P) Na skupu označenih primjera treniramo tri modela: (1) model neregularizirane logističke regresije (NR), (2) model L2-regularizirane logističke regresije (L2R) i (3) perceptron s funkcijom preslikavanja. Sva tri modela koriste iste značajke. Za sva tri algoritma promatramo iznos empirijske pogreške učenja kroz iteracije optimizacijskog postupka. Za algoritam perceptrona opažamo da empirijska pogreška učenja ne konvergira. **Kako se u ovom slučaju ponaša empirijska pogreška učenja kroz iteracije za dva spomenuta modela logističke regresije, NR i L2R?**

- ☐ A Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste
- ☐ B Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija
- ☐ C Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali pogreška modela NR doseže manju vrijednost
- ☐ D Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira

### Jezgrene i neparametarske metode (4 pitanja)

- 8 (P) Na 800 primjera sa 50 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 28 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- ☐ A Optimiramo 801 parametar, a naučeni model ima 1429 parametara
- ☐ B Optimiramo 51 parametar, a naučeni model ima 1429 parametara
- ☐ C Optimiramo 801 parametar, a naučeni model ima 1400 parametara
- ☐ D Optimiramo 51 parametar, a naučeni model ima 1400 parametara

- 9 (T) Kod algoritma SVM preporuča se napraviti skaliranje značajki. U protivnom, pri izračunu skalarnog produkta, značajke s većim rasponom (većom skalom) dominirat će nad značajkama s manjim rasponom (manjom skalom). Međutim, skaliranje značajki nije uvijek nužno. **Kada nije potrebno napraviti skaliranje značajki, i zašto?**

- ☐ A Kada se koristi linearna jezgra i značajke su centrirane oko nule, jer se onda ne računa skalarni produkt između značajki
- ☐ B Kada se koristi Gaussova jezgrena funkcija, jer ta jezgra implicitno inducira beskonačnodimenzijaski prostor značajki
- ☐ C Kada se koristi RBF jezgra sa Mahalanobisovom udaljenošću, jer ta udaljenost uzima u obzir varijancu značajki
- ☐ D Kada se koristi dualna formulacija SVM-a i algoritam SMO, jer se tada implicitno provodi L1-regularizacija

- 10 (N) Na skupu primjera za učenje iz ulaznog prostora  $n = 3$  trenirali smo SVM s polinomijalnom jezgrenom funkcijom  $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 3)^3$ . Potporni vektori i njihove oznake su sljedeći:

$$\begin{aligned}(\mathbf{x}^{(1)}, y^{(1)}) &= ((6, 12, 6), -1) \\(\mathbf{x}^{(2)}, y^{(2)}) &= ((-11, -26, -15), -1) \\(\mathbf{x}^{(3)}, y^{(3)}) &= ((-1, -7, -6), +1)\end{aligned}$$

Lagrangeovi koeficijenti su  $\alpha_1 = 3.496 \cdot 10^{-6}$ ,  $\alpha_2 = 4.136 \cdot 10^{-7}$  i  $\alpha_3 = 3.910 \cdot 10^{-6}$ . **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze  $h(\mathbf{x})$  za primjer  $\mathbf{x} = (1, 1, 25)$ .**

- ☐ A +0.947   ☐ B -0.676   ☐ C +1.434   ☐ D -2.330

- 11 (N) U ulaznome prostoru dimenzije  $n = 3$  trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri  $((2, -5, 15), -1)$ ,  $((1, 8, -305), -1)$  i  $((1, -6, 225), +1)$ , a njima odgovarajući dualni koeficijenti su  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.8$  i  $\alpha_3 = 0.9$ . Treniranje smo proveli na skaliranim značajkama: svaku smo značajku  $x_j$  standardizirali primjenom transformacije  $\frac{x_j - \mu_j}{\sigma_j}$ , gdje su  $\mu_j$  i  $\sigma_j$  srednja vrijednost odnosno varijanca značajke  $x_j$  u skupu označenih podataka  $\mathcal{D}$ . Parametri skaliranja su  $\boldsymbol{\mu} = (15, -2, 100)$  i  $\boldsymbol{\sigma} = (4, 1, 12)$ . Model SVM-a koristimo za predikciju klase primjera  $\mathbf{x} = (1, -2, 5)$ . **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera  $\mathbf{x}$ ?**

- ☐ A -373.22   ☐ B +907.43   ☐ C -739.13   ☐ D +541.53

## Procjena parametara i Bayesov klasifikator (3 pitanja)

- 12 (T) Polunaivan Bayesov klasifikator združuje u jedan faktor varijable kod kojih postoji statistička zavisnost. Za procjenu statističke zavisnosti između varijabli može se upotrijebiti Kullback-Leiblerova divergencija (KL-divergencija). **Kako se pomoću KL-divergencije može izračunati koliko su varijable zavisne?**

- ☐ A Što su varijable više zavisne, to je manja KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti  
☐ B Što su varijable više zavisne, to je veća KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti  
☐ C Što su varijable manje zavisne, to je veća KL-divergencija između faktorizirane vjerojatnosti i marginalne vjerojatnosti  
☐ D Što su varijable manje zavisne, to je manja KL-divergencija između marginalnih vjerojatnosti i uvjetne vjerojatnosti

- 13 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u  $K = 10$  klasa sa  $n = 100$  značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu  $\Sigma$  možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$  : Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane

$\mathcal{H}_2$  : Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu

$\mathcal{H}_3$  : Između značajki postoje korelacije, ali se one ne razlikuju između klasa

Neka ‘ $\supset$ ’ označava relaciju “složeniji od”, a neka ‘ $>$ ’ označava relaciju “ima više parametara od”. **Što možemo zaključiti o složenosti i broju parametara za gornja četiri modela?**

- ☐ A  $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$ ,  $\mathcal{H}_3 \supset \mathcal{H}_1$    ☐ C  $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$ ,  $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$   
☐ B  $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1$ ,  $\mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$    ☐ D  $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$ ,  $\mathcal{H}_1 \supset \mathcal{H}_2$

- 14 (N) Raspolažemo sljedećim skupom označenih primjera:

$$\mathcal{D} = \{x^{(i)}, y^{(i)}\} = \{(4, 1), (4, 1), (2, 0), (1, 0), (1, 1), (8, 1)\}$$

Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klasa  $p(x|y)$ . Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:

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

Parametre  $\mu$  i  $\sigma^2$  gustoće vjerojatnosti  $p(x|y)$  procjenjujemo MLE-om. Neka su  $\mu_1$  i  $\sigma_1^2$  parametri gustoće vjerojatnosti  $p(x|y = 1)$  dobiveni MLE-om na podskupu primjera  $\mathcal{D}_{y=1}$ . **Koliko iznosi log-izglednost  $\mathcal{L}(\mu_1, \sigma_1^2 | \mathcal{D}_{y=1})$ ?**

- ☐ A  $-11.58$    ☐ B  $-7.42$    ☐ C  $-10.47$    ☐ D  $-9.32$

### Probabilistički grafički modeli (3 pitanja)

- 15** (T) Bayesova mreža na sažet način definira zajedničku distribuciju vjerojatnosti uz određene pretpostavke uvjetne nezavisnosti. Neka je jedna takva pretpostavka  $x_1 \perp \{x_2, x_3\} | x_4$ . **Koji je efekt uvođenja ove pretpostavke na graf Bayesove mreže?**

- ☐ A Dodavanje dva brida   ☐ C Uklanjanje dva brida  
☐ B Dodavanje tri brida   ☐ D Uklanjanje tri brida

- 16** (P) Bayesova mreža ima pet varijabli, s topološkim uređajem  $v, w, x, y, z$ . Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \quad \{v, x\} \perp z | \{w, y\}$$

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

- ☐ A  $v \perp z | x$    ☐ B  $v \perp y | z$    ☐ C  $w \perp y | x$    ☐ D  $w \perp y | z$

- 17** (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji  $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$ . Sve varijable su binarne. Vrijedi  $P(w = 1) = 0.1$ ,  $P(x = 1) = 0.2$ ,  $P(z = 1|x = 0) = 0.9$  i  $P(z = 1|x = 1) = 0.7$ . Tablica uvjetnih vjerojatnosti za čvor  $y$  je sljedeća:

$w$	$x$	$p(y = 1 w, x)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar  $\mu$  uvjetne distribucije  $P(x = 0|y = 1, z = 0)$ . Uzorkovanje smo ponovili ukupno  $N = 100$  puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od  $N$ . Na temelju dobivenog uzorka parametar  $\mu$  procjenjujemo MAP procjeniteljem uz  $\alpha = \beta = 2$ . **Koliko iznosi očekivana MAP procjena parametra  $\mu$ ?**

- ☐ A 0.0786   ☐ B 0.0490   ☐ C 0.1877   ☐ D 0.1274

### Grupiranje (3 pitanja)

- 18** (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo  $N = 5$  primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.1 & 0.4 \\ 0.2 & 1.0 & 0.9 & 0.3 & 0.7 \\ 0.8 & 0.9 & 1.0 & 0.6 & 0.5 \\ 0.1 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

- ☐ A 1.8   ☐ B 2.4   ☐ C 1.9   ☐ D 2.0

- 19** (P) Algoritmom GMM grupiramo primjere u dvodimenzijaskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^3 \frac{1}{3} \mathcal{N}(\mu_j, \Sigma_j) \quad \text{gdje} \quad \mu_1 = (5, 5), \mu_2 = (5, 10), \mu_3 = (-10, -10), \Sigma_1 = \Sigma_2 = \Sigma_3 = 2\mathbf{I}$$

Skup  $\mathcal{D}$  grupiramo u  $K = 2$  grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ( $\mathcal{H}_1$ ), nedijeljena i dijagonalna kovarijacijska matrica ( $\mathcal{H}_2$ ) i nedijeljena i izotropna kovarijacijska matrica ( $\mathcal{H}_3$ ). Neka je  $\mathcal{L}_i$  izglednost parametara dobivena modelom  $\mathcal{H}_i$  nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++.

**Što su očekivani odnosi između izglednosti za ova tri modela?**

- ☐ A  $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$    ☐ B  $\mathcal{L}_2 > \mathcal{L}_1, \mathcal{L}_2 > \mathcal{L}_3$    ☐ C  $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$    ☐ D  $\mathcal{L}_1 > \mathcal{L}_3, \mathcal{L}_2 > \mathcal{L}_3$

- 20** (T) Algoritam K-sredina je iterativan algoritam za nalaženje parametara  $b_k^{(i)}$  (pripadnosti primjera grupama) i  $\mu_k$  (centroidi grupa) koji minimiziraju kriterijsku funkciju  $J$  za  $N$  primjera i  $K$  grupa. Algoritam funkciju  $J$  optimizira iterativno, jer rješenje u zatvorenoj formi ne postoji. **Zbog čega za problem minimizacije funkcije  $J$  ne postoji rješenje u zatvorenoj formi?**

- ☐ A Jer  $b_k^{(i)}$  ovisi o  $N$ , broj vektora  $\mu_k$  ovisi o  $K$ , a  $K$  je odozgo ograničen sa  $N$
- ☐ B Jer za  $b_k^{(i)}$  mora vrijediti ograničenje  $\sum_k b_k^{(i)} = 1$  i  $b_k^{(i)} \geq 0$
- ☐ C Jer  $b_k^{(i)}$  ovisi o vektorima  $\mu_k$ , a vektor  $\mu_k$  ovisi o vrijednostima  $b_k^{(i)}$
- ☐ D Jer  $J$  ovisi o  $K$  i inicijalnom odabiru za  $\mu_k$ , pa rješenje nije jedinstveno

## Vrednovanje modela (2 pitanja)

- 21** (N) Na ispitnome skupu evaluiramo klasifikator sa  $K = 3$  klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} y = 1 \quad y = 2 \quad y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} \left( \begin{array}{ccc} 3 & 15 & 11 \\ 6 & 5 & 6 \\ 4 & 2 & 36 \end{array} \right) \end{array}$$

Izračunajte mikro-F1 ( $F_1^\mu$ ) i makro-F1 ( $F_1^M$ ) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere,  $F_1^\mu - F_1^M$ ?**

- ☐ A 0.091   ☐ B 0.114   ☐ C 0.023   ☐ D 0.059

- 22** (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$\begin{array}{ll} h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) & h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8) \\ h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) & h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0) \end{array}$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ( $1 - h(\mathbf{x})$  umjesto  $h(\mathbf{x})$ ). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

- ☐ A  $h_1$  i  $h_3$    ☐ B  $h_1$  i  $h_4$    ☐ C  $h_1$  i  $h_2$    ☐ D  $h_3$  i  $h_4$



## Pismeni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

### Osnovni koncepti i linearna regresija (4 pitanja)

- 1** (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke  $x_1$ – $x_4$ ), prosjek ocjena sva četiri razreda ( $x_5$ ) te uspjeh iz matematike ( $x_6$ ) i fizike ( $x_7$ ) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr.  $x_1x_2$ ) i interakcije trojki (npr.  $x_1x_2x_3$ ) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

☐ A 75    ☐ B 48    ☐ C 63    ☐ D 38

- 2** (N) Raspoložemo sljedećim skupom primjera u dvodimenzionom ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2, -3), 5), ((2, 3), -10), ((-3, 5), 5), ((5, 0), 20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo  $L_1$ -regularizirani model linearne regresije sa  $\lambda = 2$ . Dobili smo težine  $\mathbf{w} = (4.568, 0.746, -0.550)$ . **Koliko iznosi  $L_1$ -regularizirana pogreška  $E(\mathbf{w}|\mathcal{D})$ ?**

☐ A 191.95    ☐ B 165.89    ☐ C 230.98    ☐ D 31.70

- 3** (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable  $y$  o zavisnoj varijabli  $x$ . Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka  $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_i = \{(0, 0), (2, 0), (3, 2), (5, 2)\}$ . Neka je  $h$  hipoteza koju dobivamo treniranjem modela te neka je  $L^i$  gubitak hipoteze  $h$  na primjeru  $x^{(i)}$ , tj.  $L^i = L(y^{(i)}, h(x^{(i)}))$ . **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

☐ A  $L^1 = L^4 = 1, L^2 < L^3$     ☐ C  $L^1 = L^4 < L^2 = L^3$   
☐ B  $L^1 = L^3 = 0, L^2 = L^4 < 1$     ☐ D  $L^1 = L^2 = 1 < L^3 < L^4$

- 4** (T) Svaki model strojnog učenja ima neku induktivnu pristranost. **Što je induktivna pristranost?**

- ☐ A Odstupanje procjene parametra na temelju podataka u odnosu na pravu vrijednost parametra u populaciji  
☐ B Svaka pretpostavka koja jednoznačno određuje model na temelju hipoteze i skupa za učenje  
☐ C Kriterij koji, na temelju modela, jednoznačno određuje hipotezu sa minimalnom empirijskom pogreškom  
☐ D Minimalan skup pretpostavki koje, uz skup za učenje, jednoznačno određuju klasifikaciju svakog primjera

### Linearni klasifikacijski modeli (3 pitanja)

- 5** (T) Postoji veza između logističke regresije i modela neuronske mreže. **Koja je veza između ta dva modela?**

- ☐ A Multinomijalna logistička regresija s aktivacijskom funkcijom softmax istovjetna je dvoslojnoj neuronskoj mreži sa više neurona u izlaznom sloju  
☐ B Neuronska mreža optimirana algoritmom propagacije pogreške unazad istovjetna je logističkoj regresiji optimiranoj stohastičkim gradijentnim spustom  
☐ C Logistička regresija s linearnim jezgrenim funkcijama istovjetna je neuronskoj mreži sa linearnom aktivacijskom funkcijom i kvadratnom funkcijom pogreške  
☐ D Logistička regresija koja kao adaptivne bazne funkcije koristi logističku regresiju istovjetna je neuronskoj mreži sa sigmoidnom aktivacijskom funkcijom

- 6 (N) Raspoložemo sljedećim skupom za učenje u dvodimenzijaskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijaski prostor značajki, koja je definirana na sljedeći način:

$$\phi(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, 2, 3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

- ☐ A 32   ☐ B 16   ☐ C 6   ☐ D 22

- 7 (P) Na skupu označenih primjera treniramo tri modela: (1) model neregularizirane logističke regresije (NR), (2) model L2-regularizirane logističke regresije (L2R) i (3) perceptron s funkcijom preslikavanja. Sva tri modela koriste iste značajke. Za sva tri algoritma promatramo iznos empirijske pogreške učenja kroz iteracije optimizacijskog postupka. Za algoritam perceptrona opažamo da empirijska pogreška učenja ne konvergira. **Kako se u ovom slučaju ponaša empirijska pogreška učenja kroz iteracije za dva spomenuta modela logističke regresije, NR i L2R?**

- ☐ A Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira
- ☐ B Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija
- ☐ C Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste
- ☐ D Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali pogreška modela NR doseže manju vrijednost

### Jezgrene i neparametarske metode (4 pitanja)

- 8 (P) Na 700 primjera sa 100 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 42 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- ☐ A Optimiramo 701 parametar, a naučeni model ima 4901 parametar
- ☐ B Optimiramo 701 parametar, a naučeni model ima 4200 parametara
- ☐ C Optimiramo 101 parametar, a naučeni model ima 4243 parametara
- ☐ D Optimiramo 701 parametar, a naučeni model ima 4243 parametara

- 9 (N) Na skupu primjera za učenje iz ulaznog prostora  $n = 3$  trenirali smo SVM s polinomijalnom jezgrenom funkcijom  $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 3)^3$ . Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((6, 12, 6), -1)$$

$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$

$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su  $\alpha_1 = 3.496 \cdot 10^{-6}$ ,  $\alpha_2 = 4.136 \cdot 10^{-7}$  i  $\alpha_3 = 3.910 \cdot 10^{-6}$ . **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze  $h(\mathbf{x})$  za primjer  $\mathbf{x} = (1, 1, 25)$ .**

- ☐ A -0.676   ☐ B +0.947   ☐ C +1.434   ☐ D -2.330



- 10** (T) Kod algoritma SVM preporuča se napraviti skaliranje značajki. U protivnom, pri izračunu skalarnog produkta, značajke s većim rasponom (većom skalom) dominirat će nad značajkama s manjim rasponom (manjom skalom). Međutim, skaliranje značajki nije uvijek nužno. **Kada nije potrebno napraviti skaliranje značajki, i zašto?**
- ☐ A Kada se koristi dualna formulacija SVM-a i algoritam SMO, jer se tada implicitno provodi L1-regularizacija
- ☐ B Kada se koristi Gaussova jezgrena funkcija, jer ta jezgra implicitno inducira beskonačnodimenzijski prostor značajki
- ☐ C Kada se koristi RBF jezgra sa Mahalanobisovom udaljenošću, jer ta udaljenost uzima u obzir varijancu značajki
- ☐ D Kada se koristi linearna jezgra i značajke su centrirane oko nule, jer se onda ne računa skalarni produkt između značajki
- 11** (N) U ulaznome prostoru dimenzije  $n = 3$  trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri  $((2, -5, 15), -1)$ ,  $((1, 8, -305), -1)$  i  $((1, -6, 225), +1)$ , a njima odgovarajući dualni koeficijenti su  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.8$  i  $\alpha_3 = 0.9$ . Treniranje smo proveli na skaliranim značajkama: svaku smo značajku  $x_j$  standardizirali primjenom transformacije  $\frac{x_j - \mu_j}{\sigma_j}$ , gdje su  $\mu_j$  i  $\sigma_j$  srednja vrijednost odnosno varijanca značajke  $x_j$  u skupu označenih podataka  $\mathcal{D}$ . Parametri skaliranja su  $\boldsymbol{\mu} = (15, -2, 100)$  i  $\boldsymbol{\sigma} = (4, 1, 12)$ . Model SVM-a koristimo za predikciju klase primjera  $\mathbf{x} = (1, -2, 5)$ . **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera  $\mathbf{x}$ ?**
- ☐ A  $-739.13$    ☐ B  $+541.53$    ☐ C  $-373.22$    ☐ D  $+907.43$

### Procjena parametara i Bayesov klasifikator (3 pitanja)

- 12** (N) Raspoložemo sljedećim skupom označenih primjera:

$$\mathcal{D} = \{x^{(i)}, y^{(i)}\} = \{(-4, 1), (3, 1), (-2, 0), (1, 0), (0, 1), (-8, 0)\}$$

Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klase  $p(x|y)$ . Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:

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

Parametre  $\mu$  i  $\sigma^2$  gustoće vjerojatnosti  $p(x|y)$  procjenjujemo MLE-om. Neka su  $\mu_1$  i  $\sigma_1^2$  parametri gustoće vjerojatnosti  $p(x|y = 1)$  dobiveni MLE-om na podskupu primjera  $\mathcal{D}_{y=1}$ . **Koliko iznosi log-izglednost  $\mathcal{L}(\mu_1, \sigma_1^2 | \mathcal{D}_{y=1})$ ?**

- ☐ A  $-10.47$    ☐ B  $-9.32$    ☐ C  $-11.58$    ☐ D  $-7.42$
- 13** (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u  $K = 10$  klasa sa  $n = 100$  značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu  $\Sigma$  možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:
- $\mathcal{H}_1$  : Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane
- $\mathcal{H}_2$  : Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu
- $\mathcal{H}_3$  : Između značajki postoje korelacije, ali se one ne razlikuju između klasa
- Neka ‘ $\supset$ ’ označava relaciju “složeniji od”, a neka ‘ $>$ ’ označava relaciju “ima više parametara od”. **Što možemo zaključiti o složenosti i broju parametara za gornja četiri modela?**
- ☐ A  $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$ ,  $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$    ☐ C  $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$ ,  $\mathcal{H}_3 \supset \mathcal{H}_1$
- ☐ B  $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1$ ,  $\mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$    ☐ D  $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$ ,  $\mathcal{H}_1 \supset \mathcal{H}_2$
- 14** (T) Polunaivan Bayesov klasifikator združuje u jedan faktor varijable kod kojih postoji statistička zavisnost.

Za procjenu statističke zavisnosti između varijabli može se upotrijebiti Kullback-Leiblerova divergencija (KL-divergencija). **Kako se pomoću KL-divergencije može izračunati koliko su varijable zavisne?**

- ☐ A Što su varijable manje zavisne, to je veća KL-divergencija između faktorizirane vjerojatnosti i marginalne vjerojatnosti
- ☐ B Što su varijable više zavisne, to je veća KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti
- ☐ C Što su varijable manje zavisne, to je manja KL-divergencija između marginalnih vjerojatnosti i uvjetne vjerojatnosti
- ☐ D Što su varijable više zavisne, to je manja KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti

### Probabilistički grafički modeli (3 pitanja)

- 15** (T) Bayesova mreža na sažet način definira zajedničku distribuciju vjerojatnosti uz određene pretpostavke uvjetne nezavisnosti. Neka je jedna takva pretpostavka  $x_1 \perp \{x_2, x_3\} | x_4$ . **Koji je efekt uvođenja ove pretpostavke na graf Bayesove mreže?**

- ☐ A Dodavanje tri brida
- ☐ B Dodavanje dva brida
- ☐ C Uklanjanje tri brida
- ☐ D Uklanjanje dva brida

- 16** (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji  $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$ . Sve varijable su binarne. Vrijedi  $P(w = 1) = 0.1$ ,  $P(x = 1) = 0.2$ ,  $P(z = 1|x = 0) = 0.9$  i  $P(z = 1|x = 1) = 0.7$ . Tablica uvjetnih vjerojatnosti za čvor  $y$  je sljedeća:

$w$	$x$	$p(y = 1 w, x)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar  $\mu$  uvjetne distribucije  $P(x = 0|y = 1, z = 0)$ . Uzorkovanje smo ponovili ukupno  $N = 100$  puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od  $N$ . Na temelju dobivenog uzorka parametar  $\mu$  procjenjujemo MAP procjeniteljem uz  $\alpha = \beta = 2$ . **Koliko iznosi očekivana MAP procjena parametra  $\mu$ ?**

- ☐ A 0.0490
- ☐ B 0.1877
- ☐ C 0.1274
- ☐ D 0.0786

- 17** (P) Bayesova mreža ima pet varijabli, s topološkim uređajem  $v, w, x, y, z$ . Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \quad \{v, x\} \perp z | \{w, y\}$$

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

- ☐ A  $v \perp y | x$
- ☐ B  $w \perp y | z$
- ☐ C  $x \perp z | y$
- ☐ D  $v \perp y | w$

### Grupiranje (3 pitanja)

- 18** (T) Algoritam K-sredina je iterativan algoritam za nalaženje parametara  $b_k^{(i)}$  (pripadnosti primjera grupama) i  $\mu_k$  (centroidi grupa) koji minimiziraju kriterijsku funkciju  $J$  za  $N$  primjera i  $K$  grupa. Algoritam funkciju  $J$  optimizira iterativno, jer rješenje u zatvorenoj formi ne postoji. **Zbog čega za problem minimizacije funkcije  $J$  ne postoji rješenje u zatvorenoj formi?**

- ☐ A Jer za  $b_k^{(i)}$  mora vrijediti ograničenje  $\sum_k b_k^{(i)} = 1$  i  $b_k^{(i)} \geq 0$
- ☐ B Jer  $b_k^{(i)}$  ovisi o  $N$ , broj vektora  $\mu_k$  ovisi o  $K$ , a  $K$  je odozgo ograničen sa  $N$
- ☐ C Jer  $b_k^{(i)}$  ovisi o vektorima  $\mu_k$ , a vektor  $\mu_k$  ovisi o vrijednostima  $b_k^{(i)}$
- ☐ D Jer  $J$  ovisi o  $K$  i inicijalnom odabiru za  $\mu_k$ , pa rješenje nije jedinstveno

- 19 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo  $N = 5$  primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.2 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.8 & 0.3 & 1.0 & 0.6 & 0.5 \\ 0.2 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

- ☐ A 1.8   ☐ B 2.4   ☐ C 1.9   ☐ D 2.0

- 20 (P) Algoritmom GMM grupiramo primjere u dvodimenzijaskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^3 \frac{1}{3} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \boldsymbol{\mu}_1 = (5, 5), \boldsymbol{\mu}_2 = (5, 10), \boldsymbol{\mu}_3 = (-10, -10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup  $\mathcal{D}$  grupiramo u  $K = 2$  grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ( $\mathcal{H}_1$ ), nedijeljena i dijagonalna kovarijacijska matrica ( $\mathcal{H}_2$ ) i nedijeljena i izotropna kovarijacijska matrica ( $\mathcal{H}_3$ ). Neka je  $\mathcal{L}_i$  izglednost parametara dobivena modelom  $\mathcal{H}_i$  nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

- ☐ A  $\mathcal{L}_1 > \mathcal{L}_3, \mathcal{L}_2 > \mathcal{L}_3$    ☐ B  $\mathcal{L}_2 > \mathcal{L}_1, \mathcal{L}_2 > \mathcal{L}_3$    ☐ C  $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$    ☐ D  $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$

## Vrednovanje modela (2 pitanja)

- 21 (N) Na ispitnome skupu evaluiramo klasifikator sa  $K = 3$  klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{ccc} & y = 1 & y = 2 & y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} & \begin{pmatrix} 13 & 5 & 1 \\ 6 & 5 & 6 \\ 4 & 2 & 26 \end{pmatrix} \end{array}$$

Izračunajte mikro-F1 ( $F_1^\mu$ ) i makro-F1 ( $F_1^M$ ) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere,  $F_1^\mu - F_1^M$ ?**

- ☐ A 0.059   ☐ B 0.091   ☐ C 0.023   ☐ D 0.114

- 22 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$\begin{array}{ll} h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) & h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8) \\ h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) & h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0) \end{array}$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ( $1 - h(\mathbf{x})$  umjesto  $h(\mathbf{x})$ ). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

- ☐ A  $h_1$  i  $h_3$    ☐ B  $h_3$  i  $h_4$    ☐ C  $h_2$  i  $h_3$    ☐ D  $h_1$  i  $h_4$



## Pismeni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a  $1/3$  boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

### Osnovni koncepti i linearna regresija (4 pitanja)

- 1** (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke  $x_1$ – $x_4$ ), prosjek ocjena sva četiri razreda ( $x_5$ ) te uspjeh iz matematike ( $x_6$ ) i fizike ( $x_7$ ) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr.  $x_1x_2$ ) i interakcije trojki (npr.  $x_1x_2x_3$ ) između svih značajki koje koristimo. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

☐ A 48    ☐ B 75    ☐ C 38    ☐ D 63

- 2** (T) Svaki model strojnog učenja ima neku induktivnu pristranost. **Što je induktivna pristranost?**

- ☐ A Svaka pretpostavka koja jednoznačno određuje model na temelju hipoteze i skupa za učenje  
☐ B Minimalan skup pretpostavki koje, uz skup za učenje, jednoznačno određuju klasifikaciju svakog primjera  
☐ C Kriterij koji, na temelju modela, jednoznačno određuje hipotezu sa minimalnom empirijskom pogreškom  
☐ D Odstupanje procjene parametra na temelju podataka u odnosu na pravu vrijednost parametra u populaciji

- 3** (N) Raspoložemo sljedećim skupom primjera u dvodimenzijaskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2, -3), -5), ((2, 3), 10), ((-3, 5), 5), ((5, 0), 20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo  $L_1$ -regularizirani model linearne regresije sa  $\lambda = 2$ . Dobili smo težine  $\mathbf{w} = (-0.050, 2.965, 2.482)$ . **Koliko iznosi  $L_1$ -regularizirana pogreška  $E(\mathbf{w}|\mathcal{D})$ ?**

☐ A 165.89    ☐ B 230.98    ☐ C 31.70    ☐ D 191.95

- 4** (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable  $y$  o zavisnoj varijabli  $x$ . Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka  $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_i = \{(0, 0), (2, 0), (3, 2), (5, 2)\}$ . Neka je  $h$  hipoteza koju dobivamo treniranjem modela te neka je  $L^i$  gubitak hipoteze  $h$  na primjeru  $x^{(i)}$ , tj.  $L^i = L(y^{(i)}, h(x^{(i)}))$ . **Što vrijedi za gubitke hipoteze na pojedinim primjerima?**

☐ A  $L^1 = L^4 = 1, L^2 < L^3$     ☐ C  $L^1 = L^4 < L^2 = L^3$   
☐ B  $L^1 = L^3 = 0, L^2 = L^4 < 1$     ☐ D  $L^1 = L^2 = 1 < L^3 < L^4$

### Linearni klasifikacijski modeli (3 pitanja)

- 5** (N) Raspoložemo sljedećim skupom za učenje u dvodimenzijaskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijaski prostor značajki, koja je definirana na sljedeći način:

$$\phi(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, -1, 3, 0)$$

**Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

- ☐ A 6   ☐ B 32   ☐ C 16   ☐ D 22

**6** (P) Na skupu označenih primjera treniramo tri modela: (1) model neregularizirane logističke regresije (NR), (2) model L2-regularizirane logističke regresije (L2R) i (3) perceptron s funkcijom preslikavanja. Sva tri modela koriste iste značajke. Za sva tri algoritma promatramo iznos empirijske pogreške učenja kroz iteracije optimizacijskog postupka. Za algoritam perceptrona opažamo da empirijska pogreška učenja ne konvergira. **Kako se u ovom slučaju ponaša empirijska pogreška učenja kroz iteracije za dva spomenuta modela logističke regresije, NR i L2R?**

- ☐ A Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira
- ☐ B Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija
- ☐ C Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali pogreška modela NR doseže manju vrijednost
- ☐ D Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste

**7** (T) Postoji veza između logističke regresije i modela neuronske mreže. **Koja je veza između ta dva modela?**

- ☐ A Logistička regresija s linearnim jezgrenim funkcijama istovjetna je neuronskoj mreži sa linearnom aktivacijskom funkcijom i kvadratnom funkcijom pogreške
- ☐ B Multinomijalna logistička regresija s aktivacijskom funkcijom softmax istovjetna je dvoslojnoj neuronskoj mreži sa više neurona u izlaznom sloju
- ☐ C Logistička regresija koja kao adaptivne bazne funkcije koristi logističku regresiju istovjetna je neuronskoj mreži sa sigmoidnom aktivacijskom funkcijom
- ☐ D Neuronska mreža optimirana algoritmom propagacije pogreške unazad istovjetna je logističkoj regresiji optimiranoj stohastičkim gradijentnim spustom

### Jezgrene i neparametarske metode (4 pitanja)

**8** (N) Na skupu primjera za učenje iz ulaznog prostora  $n = 3$  trenirali smo SVM s polinomijalnom jezgrenom funkcijom  $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 2)^3$ . Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((9, 30, 21), -1)$$

$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$

$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su  $\alpha_1 = 2.214 \cdot 10^{-8}$ ,  $\alpha_2 = 3.803 \cdot 10^{-8}$  i  $\alpha_3 = 6.017 \cdot 10^{-8}$ . **Upotrijebite jezgri trik da biste odredili vrijednost hipoteze  $h(\mathbf{x})$  za primjer  $\mathbf{x} = (1, 1, 25)$ .**

- ☐ A -0.676   ☐ B +1.434   ☐ C -2.330   ☐ D +0.947

**9** (P) Na 500 primjera sa 80 značajki treniramo rijetki jezgri stroj s Gausovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 38 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- ☐ A Optimiramo 501 parametara, a naučeni model ima 3541 parametar
- ☐ B Optimiramo 501 parametara, a naučeni model ima 3040 parametara
- ☐ C Optimiramo 81 parametar, a naučeni model ima 3079 parametara
- ☐ D Optimiramo 501 parametara, a naučeni model ima 3079 parametara

- 10** (T) Kod algoritma SVM preporuča se napraviti skaliranje značajki. U protivnom, pri izračunu skalarnog produkta, značajke s većim rasponom (većom skalom) dominirat će nad značajkama s manjim rasponom (manjom skalom). Međutim, skaliranje značajki nije uvijek nužno. **Kada nije potrebno napraviti skaliranje značajki, i zašto?**
- ☐ A Kada se koristi linearna jezgra i značajke su centrirane oko nule, jer se onda ne računa skalarni produkt između značajki
  - ☐ B Kada se koristi Gaussova jezgrena funkcija, jer ta jezgra implicitno inducira beskonačnodimenzijski prostor značajki
  - ☐ C Kada se koristi dualna formulacija SVM-a i algoritam SMO, jer se tada implicitno provodi L1-regularizacija
  - ☐ D Kada se koristi RBF jezgra sa Mahalanobisovom udaljenošću, jer ta udaljenost uzima u obzir varijancu značajki
- 11** (N) U ulaznome prostoru dimenzije  $n = 3$  trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri  $((2, -5, 15), -1)$ ,  $((1, 8, -305), -1)$  i  $((1, -6, 225), +1)$ , a njima odgovarajući dualni koeficijenti su  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.8$  i  $\alpha_3 = 0.9$ . Treniranje smo proveli na skaliranim značajkama: svaku smo značajku  $x_j$  standardizirali primjenom transformacije  $\frac{x_j - \mu_j}{\sigma_j}$ , gdje su  $\mu_j$  i  $\sigma_j$  srednja vrijednost odnosno varijanca značajke  $x_j$  u skupu označenih podataka  $\mathcal{D}$ . Parametri skaliranja su  $\mu = (15, -2, 100)$  i  $\sigma = (4, 1, 12)$ . Model SVM-a koristimo za predikciju klase primjera  $\mathbf{x} = (1, -2, 5)$ . **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera  $\mathbf{x}$ ?**
- ☐ A -373.22    ☐ B -739.13    ☐ C +541.53    ☐ D +907.43

### Procjena parametara i Bayesov klasifikator (3 pitanja)

- 12** (N) Raspoložemo sljedećim skupom označenih primjera:

$$\mathcal{D} = \{x^{(i)}, y^{(i)}\} = \{(4, 1), (4, 1), (2, 0), (1, 0), (1, 1), (8, 1)\}$$

Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klasa  $p(x|y)$ . Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:

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

Parametre  $\mu$  i  $\sigma^2$  gustoće vjerojatnosti  $p(x|y)$  procjenjujemo MLE-om. Neka su  $\mu_1$  i  $\sigma_1^2$  parametri gustoće vjerojatnosti  $p(x|y = 1)$  dobiveni MLE-om na podskupu primjera  $\mathcal{D}_{y=1}$ . **Koliko iznosi log-izglednost  $\mathcal{L}(\mu_1, \sigma_1^2 | \mathcal{D}_{y=1})$ ?**

- ☐ A -9.32    ☐ B -11.58    ☐ C -10.47    ☐ D -7.42
- 13** (T) Polunaivan Bayesov klasifikator združuje u jedan faktor varijable kod kojih postoji statistička zavisnost. Za procjenu statističke zavisnosti između varijabli može se upotrijebiti Kullback-Leiblerova divergencija (KL-divergencija). **Kako se pomoću KL-divergencije može izračunati koliko su varijable zavisne?**
- ☐ A Što su varijable manje zavisne, to je veća KL-divergencija između faktorizirane vjerojatnosti i marginalne vjerojatnosti
  - ☐ B Što su varijable više zavisne, to je veća KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti
  - ☐ C Što su varijable više zavisne, to je manja KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti
  - ☐ D Što su varijable manje zavisne, to je manja KL-divergencija između marginalnih vjerojatnosti i uvjetne vjerojatnosti
- 14** (P) Gausovim Bayesovim klasifikatorom rješavamo problem klasifikacije u  $K = 10$  klasa sa  $n = 100$  značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu  $\Sigma$  možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$  : Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane

$\mathcal{H}_2$  : Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu

$\mathcal{H}_3$  : Između značajki postoje korelacije, ali se one ne razlikuju između klasa

Neka ‘ $\supset$ ’ označava relaciju “složeniji od”, a neka ‘ $>$ ’ označava relaciju “ima više parametara od”. Što možemo zaključiti o složenosti i broju parametara za gornja četiri modela?

- ☐ A  $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2, \mathcal{H}_3 \supset \mathcal{H}_1$       ☐ C  $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3, \mathcal{H}_1 \supset \mathcal{H}_2$   
☐ B  $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2, \mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$       ☐ D  $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1, \mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$

### Probabilistički grafički modeli (3 pitanja)

- 15 (T) Bayesova mreža na sažet način definira zajedničku distribuciju vjerojatnosti uz određene pretpostavke uvjetne nezavisnosti. Neka je jedna takva pretpostavka  $x_1 \perp \{x_2, x_3\} | x_4$ . Koji je efekt uvođenja ove pretpostavke na graf Bayesove mreže?

- ☐ A Dodavanje dva brida      ☐ C Dodavanje tri brida  
☐ B Uklanjanje tri brida      ☐ D Uklanjanje dva brida

- 16 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem  $v, w, x, y, z$ . Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \quad \{v, x\} \perp z | \{w, y\}$$

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?

- ☐ A  $v \perp z | y$       ☐ B  $v \perp z | w$       ☐ C  $v \perp y | x$       ☐ D  $v \perp z | x$

- 17 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji  $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$ . Sve varijable su binarne. Vrijedi  $P(w = 1) = 0.1$ ,  $P(x = 1) = 0.2$ ,  $P(z = 1|x = 0) = 0.9$  i  $P(z = 1|x = 1) = 0.7$ . Tablica uvjetnih vjerojatnosti za čvor  $y$  je sljedeća:

$w$	$x$	$p(y = 1 w, x)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar  $\mu$  uvjetne distribucije  $P(x = 0|y = 1, z = 0)$ . Uzorkovanje smo ponovili ukupno  $N = 2000$  puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od  $N$ . Na temelju dobivenog uzorka parametar  $\mu$  procjenjujemo MAP procjeniteljem uz  $\alpha = \beta = 2$ . Koliko iznosi očekivana MAP procjena parametra  $\mu$ ?

- ☐ A 0.1877      ☐ B 0.1274      ☐ C 0.0490      ☐ D 0.0786

### Grupiranje (3 pitanja)

- 18 (T) Algoritam K-sredina je iterativan algoritam za nalaženje parametara  $b_k^{(i)}$  (pripadnosti primjera grupama) i  $\mu_k$  (centroidi grupa) koji minimiziraju kriterijsku funkciju  $J$  za  $N$  primjera i  $K$  grupa. Algoritam funkciju  $J$  optimizira iterativno, jer rješenje u zatvorenoj formi ne postoji. Zbog čega za problem minimizacije funkcije  $J$  ne postoji rješenje u zatvorenoj formi?

- ☐ A Jer za  $b_k^{(i)}$  mora vrijediti ograničenje  $\sum_k b_k^{(i)} = 1$  i  $b_k^{(i)} \geq 0$   
☐ B Jer  $b_k^{(i)}$  ovisi o  $N$ , broj vektora  $\mu_k$  ovisi o  $K$ , a  $K$  je odozgo ograničen sa  $N$   
☐ C Jer  $J$  ovisi o  $K$  i inicijalnom odabiru za  $\mu_k$ , pa rješenje nije jedinstveno  
☐ D Jer  $b_k^{(i)}$  ovisi o vektorima  $\mu_k$ , a vektor  $\mu_k$  ovisi o vrijednostima  $b_k^{(i)}$

- 19 (P) Algoritmom GMM grupiramo primjere u dvodimenzijaskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^3 \frac{1}{3} \mathcal{N}(\mu_j, \Sigma_j) \quad \text{gdje} \quad \mu_1 = (5, 5), \mu_2 = (5, 10), \mu_3 = (-10, -10), \Sigma_1 = \Sigma_2 = \Sigma_3 = 2\mathbf{I}$$



Skup  $\mathcal{D}$  grupiramo u  $K = 2$  grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ( $\mathcal{H}_1$ ), nedijeljena i dijagonalna kovarijacijska matrica ( $\mathcal{H}_2$ ) i nedijeljena i izotropna kovarijacijska matrica ( $\mathcal{H}_3$ ). Neka je  $\mathcal{L}_i$  izglednost parametara dobivena modelom  $\mathcal{H}_i$  nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++.

**Što su očekivani odnosi između izglednosti za ova tri modela?**

- ☐ A  $\mathcal{L}_2 > \mathcal{L}_1, \mathcal{L}_2 > \mathcal{L}_3$    ☐ B  $\mathcal{L}_1 > \mathcal{L}_3, \mathcal{L}_2 > \mathcal{L}_3$    ☐ C  $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$    ☐ D  $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$

- 20** (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo  $N = 5$  primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.6 & 0.7 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.6 & 0.3 & 1.0 & 0.9 & 0.5 \\ 0.7 & 0.3 & 0.9 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

- ☐ A 2.0   ☐ B 2.4   ☐ C 1.8   ☐ D 1.9

## Vrednovanje modela (2 pitanja)

- 21** (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$\begin{aligned} h_1 &: (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) & h_3 &: (0.1, 0.5), (0.6, 0.6), (0.7, 0.8) \\ h_2 &: (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) & h_4 &: (0.3, 0.8), (0.4, 0.9), (0.6, 1.0) \end{aligned}$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ( $1 - h(\mathbf{x})$  umjesto  $h(\mathbf{x})$ ). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

- ☐ A  $h_1$  i  $h_3$    ☐ B  $h_1$  i  $h_4$    ☐ C  $h_1$  i  $h_2$    ☐ D  $h_3$  i  $h_4$

- 22** (N) Na ispitnome skupu evaluiramo klasifikator sa  $K = 3$  klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{ccccc} & y = 1 & y = 2 & y = 3 & \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} & \begin{pmatrix} 3 & 5 & 11 \\ 6 & 5 & 6 \\ 4 & 2 & 26 \end{pmatrix} \end{array}$$

Izračunajte mikro-F1 ( $F_1^\mu$ ) i makro-F1 ( $F_1^M$ ) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere,  $F_1^\mu - F_1^M$ ?**

- ☐ A 0.114   ☐ B 0.059   ☐ C 0.023   ☐ D 0.091



## Pismeni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.) – NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

### Osnovni koncepti i linearna regresija (4 pitanja)

- 1** (T) Svaki model strojnog učenja ima neku induktivnu pristranost. Što je induktivna pristranost?
- ☐ A Svaka pretpostavka koja jednoznačno određuje model na temelju hipoteze i skupa za učenje
  - ☐ B Odstupanje procjene parametra na temelju podataka u odnosu na pravu vrijednost parametra u populaciji
  - ☐ C Kriterij koji, na temelju modela, jednoznačno određuje hipotezu sa minimalnom empirijskom pogreškom
  - ☐ D Minimalan skup pretpostavki koje, uz skup za učenje, jednoznačno određuju klasifikaciju svakog primjera
- 2** (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable  $y$  o zavisnoj varijabli  $x$ . Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka  $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_i = \{(0, 0), (2, 0), (3, 2), (5, 2)\}$ . Neka je  $h$  hipoteza koju dobivamo treniranjem modela te neka je  $L^i$  gubitak hipoteze  $h$  na primjeru  $x^{(i)}$ , tj.  $L^i = L(y^{(i)}, h(x^{(i)}))$ . Što vrijedi za gubitke hipoteze na pojedinim primjerima?
- ☐ A  $L^1 = L^3 = 0, L^2 = L^4 < 1$
  - ☐ B  $L^1 = L^4 < L^2 = L^3$
  - ☐ C  $L^1 = L^2 = 1 < L^3 < L^4$
  - ☐ D  $L^1 = L^4 = 1, L^2 < L^3$
- 3** (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke  $x_1$ – $x_4$ ), prosjek ocjena sva četiri razreda ( $x_5$ ) te uspjeh iz matematike ( $x_6$ ) i fizike ( $x_7$ ) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr.  $x_1x_2$ ) i interakcije trojki (npr.  $x_1x_2x_3$ ) između svih značajki koje koristimo. Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?
- ☐ A 48
  - ☐ B 75
  - ☐ C 63
  - ☐ D 38
- 4** (N) Raspoložemo sljedećim skupom primjera u dvodimenzijaskome ulaznom prostoru:
- $$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2, -3), -5), ((2, 3), -10), ((-3, 5), 5), ((5, 0), 20)\}$$
- Na ovom skupu gradijentnim spustom trenirali smo  $L_1$ -regularizirani model linearne regresije sa  $\lambda = 2$ . Dobili smo težine  $\mathbf{w} = (-0.156, 1.330, 0.529)$ . Koliko iznosi  $L_1$ -regularizirana pogreška  $E(\mathbf{w}|\mathcal{D})$ ?
- ☐ A 165.89
  - ☐ B 191.95
  - ☐ C 230.98
  - ☐ D 31.70

### Linearni klasifikacijski modeli (3 pitanja)

- 5** (T) Postoji veza između logističke regresije i modela neuronske mreže. Koja je veza između ta dva modela?
- ☐ A Multinomijalna logistička regresija s aktivacijskom funkcijom softmax istovjetna je dvoslojnoj neuronskoj mreži sa više neurona u izlaznom sloju
  - ☐ B Logistička regresija s linearnim jezgrenim funkcijama istovjetna je neuronskoj mreži sa linearnom aktivacijskom funkcijom i kvadratnom funkcijom pogreške
  - ☐ C Logistička regresija koja kao adaptivne bazne funkcije koristi logističku regresiju istovjetna je neuronskoj mreži sa sigmoidnom aktivacijskom funkcijom
  - ☐ D Neuronska mreža optimirana algoritmom propagacije pogreške unazad istovjetna je logističkoj regresiji optimiranoj stohastičkim gradijentnim spustom

**6** (P) Na skupu označenih primjera treniramo tri modela: (1) model neregularizirane logističke regresije (NR), (2) model L2-regularizirane logističke regresije (L2R) i (3) perceptron s funkcijom preslikavanja. Sva tri modela koriste iste značajke. Za sva tri algoritma promatramo iznos empirijske pogreške učenja kroz iteracije optimizacijskog postupka. Za algoritam perceptrona opažamo da empirijska pogreška učenja ne konvergira. **Kako se u ovom slučaju ponaša empirijska pogreška učenja kroz iteracije za dva spomenuta modela logističke regresije, NR i L2R?**

- ☐ A Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali pogreška modela NR doseže manju vrijednost
- ☐ B Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira
- ☐ C Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste
- ☐ D Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija

**7** (N) Raspoložemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, koja je definirana na sljedeći način:

$$\phi(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, 2, 3, 0)$$

**Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

- ☐ A 6   ☐ B 16   ☐ C 22   ☐ D 32

### Jezgrene i neparametarske metode (4 pitanja)

**8** (P) Na 900 primjera sa 50 značajki treniramo rijetki jezgri stroj s Gausovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 52 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- ☐ A Optimiramo 901 parametar, a naučeni model ima 2600 parametara
- ☐ B Optimiramo 51 parametar, a naučeni model ima 2600 parametara
- ☐ C Optimiramo 901 parametar, a naučeni model ima 3501 parametar
- ☐ D Optimiramo 901 parametar, a naučeni model ima 2653 parametara

**9** (N) U ulaznome prostoru dimenzije  $n = 3$  trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri  $((2, -5, 15), -1)$ ,  $((1, 8, -305), -1)$  i  $((1, -6, 225), +1)$ , a njima odgovarajući dualni koeficijenti su  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.8$  i  $\alpha_3 = 0.9$ . Treniranje smo proveli na skaliranim značajkama: svaku smo značajku  $x_j$  standardizirali primjenom transformacije  $\frac{x_j - \mu_j}{\sigma_j}$ , gdje su  $\mu_j$  i  $\sigma_j$  srednja vrijednost odnosno varijanca značajke  $x_j$  u skupu označenih podataka  $\mathcal{D}$ . Parametri skaliranja su  $\boldsymbol{\mu} = (15, -2, 100)$  i  $\boldsymbol{\sigma} = (4, 1, 12)$ . Model SVM-a koristimo za predikciju klase primjera  $\mathbf{x} = (1, -2, 5)$ . **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera  $\mathbf{x}$ ?**

- ☐ A +907.43   ☐ B +541.53   ☐ C -373.22   ☐ D -739.13

- 10** (T) Kod algoritma SVM preporuča se napraviti skaliranje značajki. U protivnom, pri izračunu skalarnog produkta, značajke s većim rasponom (većom skalom) dominirat će nad značajkama s manjim rasponom (manjom skalom). Međutim, skaliranje značajki nije uvijek nužno. **Kada nije potrebno napraviti skaliranje značajki, i zašto?**
- ☐ A Kada se koristi dualna formulacija SVM-a i algoritam SMO, jer se tada implicitno provodi L1-regularizacija
- ☐ B Kada se koristi RBF jezgra sa Mahalanobisovom udaljenošću, jer ta udaljenost uzima u obzir varijancu značajki
- ☐ C Kada se koristi linearna jezgra i značajke su centrirane oko nule, jer se onda ne računa skalarni produkt između značajki
- ☐ D Kada se koristi Gaussova jezgrena funkcija, jer ta jezgra implicitno inducira beskonačnodimenzijski prostor značajki
- 11** (N) Na skupu primjera za učenje iz ulaznog prostora  $n = 3$  trenirali smo SVM s polinomijalnom jezgrenom funkcijom  $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 3)^3$ . Potporni vektori i njihove oznake su sljedeći:

$$\begin{aligned}(\mathbf{x}^{(1)}, y^{(1)}) &= ((6, 12, 6), -1) \\(\mathbf{x}^{(2)}, y^{(2)}) &= ((-11, -26, -15), -1) \\(\mathbf{x}^{(3)}, y^{(3)}) &= ((-1, -7, -6), +1)\end{aligned}$$

Lagrangeovi koeficijenti su  $\alpha_1 = 3.496 \cdot 10^{-6}$ ,  $\alpha_2 = 4.136 \cdot 10^{-7}$  i  $\alpha_3 = 3.910 \cdot 10^{-6}$ . **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze  $h(\mathbf{x})$  za primjer  $\mathbf{x} = (1, 1, 25)$ .**

- ☐ A +0.947   ☐ B -2.330   ☐ C -0.676   ☐ D +1.434

### Procjena parametara i Bayesov klasifikator (3 pitanja)

- 12** (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u  $K = 10$  klasa sa  $n = 100$  značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu  $\Sigma$  možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$  : Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane

$\mathcal{H}_2$  : Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu

$\mathcal{H}_3$  : Između značajki postoje korelacije, ali se one ne razlikuju između klasa

Neka ‘ $\supset$ ’ označava relaciju “složeniji od”, a neka ‘ $>$ ’ označava relaciju “ima više parametara od”. **Što možemo zaključiti o složenosti i broju parametara za gornja četiri modela?**

- ☐ A  $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$ ,  $\mathcal{H}_3 \supset \mathcal{H}_1$    ☐ C  $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$ ,  $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$
- ☐ B  $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1$ ,  $\mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$    ☐ D  $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$ ,  $\mathcal{H}_1 \supset \mathcal{H}_2$

- 13** (N) Raspoložemo sljedećim skupom označenih primjera:

$$\mathcal{D} = \{x^{(i)}, y^{(i)}\} = \{(4, 1), (4, 1), (2, 0), (1, 0), (1, 1), (8, 1)\}$$

Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klasa  $p(x|y)$ . Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:

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

Parametre  $\mu$  i  $\sigma^2$  gustoće vjerojatnosti  $p(x|y)$  procjenjujemo MLE-om. Neka su  $\mu_1$  i  $\sigma_1^2$  parametri gustoće vjerojatnosti  $p(x|y = 1)$  dobiveni MLE-om na podskupu primjera  $\mathcal{D}_{y=1}$ . **Koliko iznosi log-izglednost  $\mathcal{L}(\mu_1, \sigma_1^2|\mathcal{D}_{y=1})$ ?**

- ☐ A -7.42   ☐ B -9.32   ☐ C -10.47   ☐ D -11.58

- 14 (T) Polunaivan Bayesov klasifikator združuje u jedan faktor varijable kod kojih postoji statistička zavisnost. Za procjenu statističke zavisnosti između varijabli može se upotrijebiti Kullback-Leiblerova divergencija (KL-divergencija). **Kako se pomoću KL-divergencije može izračunati koliko su varijable zavisne?**
- ☐ A Što su varijable više zavisne, to je manja KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti
- ☐ B Što su varijable više zavisne, to je veća KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti
- ☐ C Što su varijable manje zavisne, to je veća KL-divergencija između faktorizirane vjerojatnosti i marginalne vjerojatnosti
- ☐ D Što su varijable manje zavisne, to je manja KL-divergencija između marginalnih vjerojatnosti i uvjetne vjerojatnosti

### Probabilistički grafički modeli (3 pitanja)

- 15 (T) Bayesova mreža na sažet način definira zajedničku distribuciju vjerojatnosti uz određene pretpostavke uvjetne nezavisnosti. Neka je jedna takva pretpostavka  $x_1 \perp \{x_2, x_3\} | x_4$ . **Koji je efekt uvođenja ove pretpostavke na graf Bayesove mreže?**

- ☐ A Dodavanje tri brida      ☐ C Uklanjanje tri brida
- ☐ B Uklanjanje dva brida      ☐ D Dodavanje dva brida

- 16 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji  $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$ . Sve varijable su binarne. Vrijedi  $P(w = 1) = 0.1$ ,  $P(x = 1) = 0.2$ ,  $P(z = 1|x = 0) = 0.9$  i  $P(z = 1|x = 1) = 0.7$ . Tablica uvjetnih vjerojatnosti za čvor  $y$  je sljedeća:

$w$	$x$	$p(y = 1 w, x)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar  $\mu$  uvjetne distribucije  $P(x = 0|y = 1, z = 0)$ . Uzorkovanje smo ponovili ukupno  $N = 100$  puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od  $N$ . Na temelju dobivenog uzorka parametar  $\mu$  procjenjujemo MAP procjeniteljem uz  $\alpha = \beta = 2$ . **Koliko iznosi očekivana MAP procjena parametra  $\mu$ ?**

- ☐ A 0.0786    ☐ B 0.0490    ☐ C 0.1877    ☐ D 0.1274

- 17 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem  $v, w, x, y, z$ . Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \quad \{v, x\} \perp z | \{w, y\}$$

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

- ☐ A  $x \perp z | y$     ☐ B  $x \perp z | w$     ☐ C  $w \perp y | x$     ☐ D  $v \perp y | w$

### Grupiranje (3 pitanja)

- 18 (T) Algoritam K-sredina je iterativan algoritam za nalaženje parametara  $b_k^{(i)}$  (pripadnosti primjera grupama) i  $\mu_k$  (centroidi grupa) koji minimiziraju kriterijsku funkciju  $J$  za  $N$  primjera i  $K$  grupa. Algoritam funkciju  $J$  optimizira iterativno, jer rješenje u zatvorenoj formi ne postoji. **Zbog čega za problem minimizacije funkcije  $J$  ne postoji rješenje u zatvorenoj formi?**

- ☐ A Jer za  $b_k^{(i)}$  mora vrijediti ograničenje  $\sum_k b_k^{(i)} = 1$  i  $b_k^{(i)} \geq 0$
- ☐ B Jer  $b_k^{(i)}$  ovisi o  $N$ , broj vektora  $\mu_k$  ovisi o  $K$ , a  $K$  je odozgo ograničen sa  $N$
- ☐ C Jer  $b_k^{(i)}$  ovisi o vektorima  $\mu_k$ , a vektor  $\mu_k$  ovisi o vrijednostima  $b_k^{(i)}$
- ☐ D Jer  $J$  ovisi o  $K$  i inicijalnom odabiru za  $\mu_k$ , pa rješenje nije jedinstveno

- 19 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo  $N = 5$  primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.1 & 0.4 \\ 0.2 & 1.0 & 0.9 & 0.3 & 0.7 \\ 0.8 & 0.9 & 1.0 & 0.6 & 0.5 \\ 0.1 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

- ☐ A 2.4   ☐ B 2.0   ☐ C 1.9   ☐ D 1.8

- 20 (P) Algoritmom GMM grupiramo primjere u dvodimenzijaskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^3 \frac{1}{3} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \boldsymbol{\mu}_1 = (5, 5), \boldsymbol{\mu}_2 = (5, 10), \boldsymbol{\mu}_3 = (-10, -10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

Skup  $\mathcal{D}$  grupiramo u  $K = 2$  grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ( $\mathcal{H}_1$ ), nedijeljena i dijagonalna kovarijacijska matrica ( $\mathcal{H}_2$ ) i nedijeljena i izotropna kovarijacijska matrica ( $\mathcal{H}_3$ ). Neka je  $\mathcal{L}_i$  izglednost parametara dobivena modelom  $\mathcal{H}_i$  nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

- ☐ A  $\mathcal{L}_2 > \mathcal{L}_1, \mathcal{L}_2 > \mathcal{L}_3$    ☐ B  $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$    ☐ C  $\mathcal{L}_1 > \mathcal{L}_3, \mathcal{L}_2 > \mathcal{L}_3$    ☐ D  $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$

## Vrednovanje modela (2 pitanja)

- 21 (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$\begin{aligned} h_1 &: (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) & h_3 &: (0.1, 0.5), (0.6, 0.6), (0.7, 0.8) \\ h_2 &: (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) & h_4 &: (0.3, 0.8), (0.4, 0.9), (0.6, 1.0) \end{aligned}$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ( $1 - h(\mathbf{x})$  umjesto  $h(\mathbf{x})$ ). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

- ☐ A  $h_3$  i  $h_4$    ☐ B  $h_2$  i  $h_3$    ☐ C  $h_1$  i  $h_4$    ☐ D  $h_1$  i  $h_2$

- 22 (N) Na ispitnome skupu evaluiramo klasifikator sa  $K = 3$  klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{ccc} & y = 1 & y = 2 & y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} & \begin{pmatrix} 3 & 5 & 11 \\ 6 & 5 & 6 \\ 4 & 2 & 26 \end{pmatrix} \end{array}$$

Izračunajte mikro-F1 ( $F_1^\mu$ ) i makro-F1 ( $F_1^M$ ) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere,  $F_1^\mu - F_1^M$ ?**

- ☐ A 0.091   ☐ B 0.114   ☐ C 0.023   ☐ D 0.059





## Pismeni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a  $1/3$  boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

### Osnovni koncepti i linearna regresija (4 pitanja)

**1** (T) Svaki model strojnog učenja ima neku induktivnu pristranost. Što je induktivna pristranost?

- ☐ A Kriterij koji, na temelju modela, jednoznačno određuje hipotezu sa minimalnom empirijskom pogreškom
- ☐ B Svaka pretpostavka koja jednoznačno određuje model na temelju hipoteze i skupa za učenje
- ☐ C Minimalan skup pretpostavki koje, uz skup za učenje, jednoznačno određuju klasifikaciju svakog primjera
- ☐ D Odstupanje procjene parametra na temelju podataka u odnosu na pravu vrijednost parametra u populaciji

**2** (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable  $y$  o zavisnoj varijabli  $x$ . Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka  $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_i = \{(0, 0), (2, 0), (3, 2), (5, 2)\}$ . Neka je  $h$  hipoteza koju dobivamo treniranjem modela te neka je  $L^i$  gubitak hipoteze  $h$  na primjeru  $x^{(i)}$ , tj.  $L^i = L(y^{(i)}, h(x^{(i)}))$ . Što vrijedi za gubitke hipoteze na pojedinim primjerima?

- ☐ A  $L^1 = L^2 = 1 < L^3 < L^4$
- ☐ B  $L^1 = L^4 < L^2 = L^3$
- ☐ C  $L^1 = L^4 = 1, L^2 < L^3$
- ☐ D  $L^1 = L^3 = 0, L^2 = L^4 < 1$

**3** (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke  $x_1$ – $x_4$ ), prosjek ocjena sva četiri razreda ( $x_5$ ) te uspjeh iz matematike ( $x_6$ ) i fizike ( $x_7$ ) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr.  $x_1x_2$ ) i interakcije trojki (npr.  $x_1x_2x_3$ ) između svih značajki koje koristimo. Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?

- ☐ A 38
- ☐ B 48
- ☐ C 75
- ☐ D 63

**4** (N) Raspoložemo sljedećim skupom primjera u dvodimenzijaskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2, -3), -5), ((2, 3), -10), ((-3, 5), -5), ((5, 0), 20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo  $L_1$ -regularizirani model linearne regresije sa  $\lambda = 2$ . Dobili smo težine  $\mathbf{w} = (-3.977, 2.477, 0.210)$ . Koliko iznosi  $L_1$ -regularizirana pogreška  $E(\mathbf{w}|\mathcal{D})$ ?

- ☐ A 191.95
- ☐ B 165.89
- ☐ C 230.98
- ☐ D 31.70

### Linearni klasifikacijski modeli (3 pitanja)

**5** (P) Na skupu označenih primjera treniramo tri modela: (1) model neregularizirane logističke regresije (NR), (2) model  $L_2$ -regularizirane logističke regresije (L2R) i (3) perceptron s funkcijom preslikavanja. Sva tri modela koriste iste značajke. Za sva tri algoritma promatramo iznos empirijske pogreške učenja kroz iteracije optimizacijskog postupka. Za algoritam perceptrona opažamo da empirijska pogreška učenja ne konvergira. Kako se u

ovom slučaju ponaša empirijska pogreška učenja kroz iteracije za dva spomenuta modela logističke regresije, NR i L2R?

- ☐ A Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira
- ☐ B Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste
- ☐ C Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali pogreška modela NR doseže manju vrijednost
- ☐ D Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija

6 (N) Raspoložemo sljedećim skupom za učenje u dvodimenzijaskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijaski prostor značajki, koja je definirana na sljedeći način:

$$\phi(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, -1, 3, 0)$$

Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?

- ☐ A 6    ☐ B 16    ☐ C 32    ☐ D 22

7 (T) Postoji veza između logističke regresije i modela neuronske mreže. **Koja je veza između ta dva modela?**

- ☐ A Logistička regresija koja kao adaptivne bazne funkcije koristi logističku regresiju istovjetna je neuronskoj mreži sa sigmoidnom aktivacijskom funkcijom
- ☐ B Multinomijalna logistička regresija s aktivacijskom funkcijom softmax istovjetna je dvoslojnoj neuronskoj mreži sa više neurona u izlaznom sloju
- ☐ C Logistička regresija s linearnim jezgrenim funkcijama istovjetna je neuronskoj mreži sa linearnom aktivacijskom funkcijom i kvadratnom funkcijom pogreške
- ☐ D Neuronska mreža optimirana algoritmom propagacije pogreške unazad istovjetna je logističkoj regresiji optimiranoj stohastičkim gradijentnim spustom

### Jezgrene i neparametarske metode (4 pitanja)

8 (N) U ulaznome prostoru dimenzije  $n = 3$  trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri  $((2, -5, 15), -1)$ ,  $((1, 8, -305), -1)$  i  $((1, -6, 225), +1)$ , a njima odgovarajući dualni koeficijenti su  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.8$  i  $\alpha_3 = 0.9$ . Treniranje smo proveli na skaliranim značajkama: svaku smo značajku  $x_j$  standardizirali primjenom transformacije  $\frac{x_j - \mu_j}{\sigma_j}$ , gdje su  $\mu_j$  i  $\sigma_j$  srednja vrijednost odnosno varijanca značajke  $x_j$  u skupu označenih podataka  $\mathcal{D}$ . Parametri skaliranja su  $\boldsymbol{\mu} = (15, -2, 100)$  i  $\boldsymbol{\sigma} = (4, 1, 12)$ . Model SVM-a koristimo za predikciju klase primjera  $\mathbf{x} = (1, 2, -20)$ . **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera  $\mathbf{x}$ ?**

- ☐ A +907.43    ☐ B -739.13    ☐ C -373.22    ☐ D +541.53

9 (T) Kod algoritma SVM preporuča se napraviti skaliranje značajki. U protivnom, pri izračunu skalarnog produkta, značajke s većim rasponom (većom skalom) dominirat će nad značajkama s manjim rasponom (manjom skalom). Međutim, skaliranje značajki nije uvijek nužno. **Kada nije potrebno napraviti skaliranje značajki, i zašto?**

- ☐ A Kada se koristi Gaussova jezgrena funkcija, jer ta jezgra implicitno inducira beskonačnodimenzijaski prostor značajki
- ☐ B Kada se koristi dualna formulacija SVM-a i algoritam SMO, jer se tada implicitno provodi L1-regularizacija
- ☐ C Kada se koristi RBF jezgra sa Mahalanobisovom udaljenošću, jer ta udaljenost uzima u obzir varijancu značajki
- ☐ D Kada se koristi linearna jezgra i značajke su centrirane oko nule, jer se onda ne računa skalarni produkt između značajki

- 10** (N) Na skupu primjera za učenje iz ulaznog prostora  $n = 3$  trenirali smo SVM s polinomijalnom jezgrenom funkcijom  $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 2)^3$ . Potporni vektori i njihove oznake su sljedeći:

$$\begin{aligned}(\mathbf{x}^{(1)}, y^{(1)}) &= ((9, 30, 21), -1) \\(\mathbf{x}^{(2)}, y^{(2)}) &= ((-11, -26, -15), -1) \\(\mathbf{x}^{(3)}, y^{(3)}) &= ((-1, -7, -6), +1)\end{aligned}$$

Lagrangeovi koeficijenti su  $\alpha_1 = 2.214 \cdot 10^{-8}$ ,  $\alpha_2 = 3.803 \cdot 10^{-8}$  i  $\alpha_3 = 6.017 \cdot 10^{-8}$ . **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze  $h(\mathbf{x})$  za primjer  $\mathbf{x} = (1, 1, 25)$ .**

- ☐ A +0.947   ☐ B -0.676   ☐ C +1.434   ☐ D -2.330

- 11** (P) Na 700 primjera sa 100 značajki treniramo rijetki jezgreni stroj s Gausovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 42 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- ☐ A Optimiramo 101 parametar, a naučeni model ima 4243 parametara  
☐ B Optimiramo 701 parametar, a naučeni model ima 4243 parametara  
☐ C Optimiramo 701 parametar, a naučeni model ima 4901 parametar  
☐ D Optimiramo 101 parametar, a naučeni model ima 4200 parametara

### Procjena parametara i Bayesov klasifikator (3 pitanja)

- 12** (N) Raspoložemo sljedećim skupom označenih primjera:

$$\mathcal{D} = \{x^{(i)}, y^{(i)}\} = \{(4, 1), (-3, 1), (-2, 0), (1, 0), (0, 1), (-8, 1)\}$$

Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klasa  $p(x|y)$ . Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:

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

Parametre  $\mu$  i  $\sigma^2$  gustoće vjerojatnosti  $p(x|y)$  procjenjujemo MLE-om. Neka su  $\mu_1$  i  $\sigma_1^2$  parametri gustoće vjerojatnosti  $p(x|y = 1)$  dobiveni MLE-om na podskupu primjera  $\mathcal{D}_{y=1}$ . **Koliko iznosi log-izglednost  $\mathcal{L}(\mu_1, \sigma_1^2 | \mathcal{D}_{y=1})$ ?**

- ☐ A -9.32   ☐ B -7.42   ☐ C -10.47   ☐ D -11.58

- 13** (T) Polunaivan Bayesov klasifikator združuje u jedan faktor varijable kod kojih postoji statistička zavisnost. Za procjenu statističke zavisnosti između varijabli može se upotrijebiti Kullback-Leiblerova divergencija (KL-divergencija). **Kako se pomoću KL-divergencije može izračunati koliko su varijable zavisne?**

- ☐ A Što su varijable manje zavisne, to je veća KL-divergencija između faktorizirane vjerojatnosti i marginalne vjerojatnosti  
☐ B Što su varijable manje zavisne, to je manja KL-divergencija između marginalnih vjerojatnosti i uvjetne vjerojatnosti  
☐ C Što su varijable više zavisne, to je manja KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti  
☐ D Što su varijable više zavisne, to je veća KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti

- 14** (P) Gausovim Bayesovim klasifikatorom rješavamo problem klasifikacije u  $K = 10$  klasa sa  $n = 100$  značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu  $\Sigma$  možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$  : Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane

$\mathcal{H}_2$  : Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu

$\mathcal{H}_3$  : Između značajki postoje korelacije, ali se one ne razlikuju između klasa

Neka ‘ $\supset$ ’ označava relaciju “složeniji od”, a neka ‘ $>$ ’ označava relaciju “ima više parametara od”. Što možemo zaključiti o složenosti i broju parametara za gornja četiri modela?

- ☐ A  $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3, \mathcal{H}_1 \supset \mathcal{H}_2$       ☐ C  $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1, \mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$   
☐ B  $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2, \mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$       ☐ D  $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2, \mathcal{H}_3 \supset \mathcal{H}_1$

### Probabilistički grafički modeli (3 pitanja)

- 15 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji  $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$ . Sve varijable su binarne. Vrijedi  $P(w = 1) = 0.1$ ,  $P(x = 1) = 0.2$ ,  $P(z = 1|x = 0) = 0.9$  i  $P(z = 1|x = 1) = 0.7$ . Tablica uvjetnih vjerojatnosti za čvor  $y$  je sljedeća:

$w$	$x$	$p(y = 1 w, x)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar  $\mu$  uvjetne distribucije  $P(x = 0|y = 1, z = 0)$ . Uzorkovanje smo ponovili ukupno  $N = 200$  puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od  $N$ . Na temelju dobivenog uzorka parametar  $\mu$  procjenjujemo MAP procjeniteljem uz  $\alpha = \beta = 2$ . **Koliko iznosi očekivana MAP procjena parametra  $\mu$ ?**

- ☐ A 0.1274    ☐ B 0.1877    ☐ C 0.0490    ☐ D 0.0786

- 16 (T) Bayesova mreža na sažet način definira zajedničku distribuciju vjerojatnosti uz određene pretpostavke uvjetne nezavisnosti. Neka je jedna takva pretpostavka  $x_1 \perp \{x_2, x_3\} | x_4$ . **Koji je efekt uvođenja ove pretpostavke na graf Bayesove mreže?**

- ☐ A Uklanjanje dva brida    ☐ C Dodavanje dva brida  
☐ B Dodavanje tri brida    ☐ D Uklanjanje tri brida

- 17 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem  $v, w, x, y, z$ . Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \quad \{v, x\} \perp z | \{w, y\}$$

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. **Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?**

- ☐ A  $w \perp y | z$     ☐ B  $v \perp y | x$     ☐ C  $v \perp z | x$     ☐ D  $v \perp y | z$

### Grupiranje (3 pitanja)

- 18 (T) Algoritam K-sredina je iterativan algoritam za nalaženje parametara  $b_k^{(i)}$  (pripadnosti primjera grupama) i  $\mu_k$  (centroidi grupa) koji minimiziraju kriterijsku funkciju  $J$  za  $N$  primjera i  $K$  grupa. Algoritam funkciju  $J$  optimizira iterativno, jer rješenje u zatvorenoj formi ne postoji. **Zbog čega za problem minimizacije funkcije  $J$  ne postoji rješenje u zatvorenoj formi?**

- ☐ A Jer  $b_k^{(i)}$  ovisi o  $N$ , broj vektora  $\mu_k$  ovisi o  $K$ , a  $K$  je odozgo ograničen sa  $N$   
☐ B Jer  $b_k^{(i)}$  ovisi o vektorima  $\mu_k$ , a vektor  $\mu_k$  ovisi o vrijednostima  $b_k^{(i)}$   
☐ C Jer  $J$  ovisi o  $K$  i inicijalnom odabiru za  $\mu_k$ , pa rješenje nije jedinstveno  
☐ D Jer za  $b_k^{(i)}$  mora vrijediti ograničenje  $\sum_k b_k^{(i)} = 1$  i  $b_k^{(i)} \geq 0$

- 19 (P) Algoritmom GMM grupiramo primjere u dvodimenzijaskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^3 \frac{1}{3} \mathcal{N}(\mu_j, \Sigma_j) \quad \text{gdje} \quad \mu_1 = (5, 5), \mu_2 = (5, 10), \mu_3 = (-10, -10), \Sigma_1 = \Sigma_2 = \Sigma_3 = 2\mathbf{I}$$

Skup  $\mathcal{D}$  grupiramo u  $K = 2$  grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ( $\mathcal{H}_1$ ), nedijeljena i dijagonalna kovarijacijska matrica ( $\mathcal{H}_2$ ) i nedijeljena i izotropna kovarijacijska matrica ( $\mathcal{H}_3$ ). Neka je  $\mathcal{L}_i$  izglednost parametara dobivena modelom  $\mathcal{H}_i$  nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++.

**Što su očekivani odnosi između izglednosti za ova tri modela?**

- ☐ A  $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$    ☐ B  $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$    ☐ C  $\mathcal{L}_2 > \mathcal{L}_1, \mathcal{L}_2 > \mathcal{L}_3$    ☐ D  $\mathcal{L}_1 > \mathcal{L}_3, \mathcal{L}_2 > \mathcal{L}_3$

- 20** (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo  $N = 5$  primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.6 & 0.7 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.6 & 0.3 & 1.0 & 0.9 & 0.5 \\ 0.7 & 0.3 & 0.9 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. **Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?**

- ☐ A 1.9   ☐ B 2.4   ☐ C 1.8   ☐ D 2.0

## Vrednovanje modela (2 pitanja)

- 21** (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$\begin{aligned} h_1 &: (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) & h_3 &: (0.1, 0.5), (0.6, 0.6), (0.7, 0.8) \\ h_2 &: (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) & h_4 &: (0.3, 0.8), (0.4, 0.9), (0.6, 1.0) \end{aligned}$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ( $1 - h(\mathbf{x})$  umjesto  $h(\mathbf{x})$ ). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

- ☐ A  $h_2$  i  $h_3$    ☐ B  $h_3$  i  $h_4$    ☐ C  $h_1$  i  $h_2$    ☐ D  $h_1$  i  $h_4$

- 22** (N) Na ispitnome skupu evaluiramo klasifikator sa  $K = 3$  klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{ccccc} & y = 1 & y = 2 & y = 3 & \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} & \begin{pmatrix} 3 & 15 & 11 \\ 6 & 5 & 6 \\ 4 & 2 & 36 \end{pmatrix} \end{array}$$

Izračunajte mikro-F1 ( $F_1^\mu$ ) i makro-F1 ( $F_1^M$ ) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere,  $F_1^\mu - F_1^M$ ?**

- ☐ A 0.114   ☐ B 0.023   ☐ C 0.091   ☐ D 0.059



## Pismeni ispit iz Strojnog učenja 1 (ak. god. 2022./2023.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.

### Osnovni koncepti i linearna regresija (4 pitanja)

- 1 (P) Jednostavnom regresijom modeliramo ovisnost nezavisne varijable  $y$  o zavisnoj varijabli  $x$ . Model treniramo postupkom običnih najmanjih kvadrata (OLS) na skupu podataka  $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_i = \{(0, 0), (2, 0), (3, 2), (5, 2)\}$ . Neka je  $h$  hipoteza koju dobivamo treniranjem modela te neka je  $L^i$  gubitak hipoteze  $h$  na primjeru  $x^{(i)}$ , tj.  $L^i = L(y^{(i)}, h(x^{(i)}))$ . Što vrijedi za gubitke hipoteze na pojedinim primjerima?

- ☐ A  $L^1 = L^4 = 1, L^2 < L^3$       ☐ C  $L^1 = L^4 < L^2 = L^3$   
☐ B  $L^1 = L^3 = 0, L^2 = L^4 < 1$       ☐ D  $L^1 = L^2 = 1 < L^3 < L^4$

- 2 (T) Svaki model strojnog učenja ima neku induktivnu pristranost. Što je induktivna pristranost?

- ☐ A Svaka pretpostavka koja jednoznačno određuje model na temelju hipoteze i skupa za učenje  
☐ B Kriterij koji, na temelju modela, jednoznačno određuje hipotezu sa minimalnom empirijskom pogreškom  
☐ C Minimalan skup pretpostavki koje, uz skup za učenje, jednoznačno određuju klasifikaciju svakog primjera  
☐ D Odstupanje procjene parametra na temelju podataka u odnosu na pravu vrijednost parametra u populaciji

- 3 (N) Raspoložemo sljedećim skupom primjera u dvodimenzijaskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \{((2, -3), -5), ((2, 3), 10), ((-3, 5), 5), ((5, 0), 20)\}$$

Na ovom skupu gradijentnim spustom trenirali smo  $L_1$ -regularizirani model linearne regresije sa  $\lambda = 2$ . Dobili smo težine  $\mathbf{w} = (-0.050, 2.965, 2.482)$ . Koliko iznosi  $L_1$ -regularizirana pogreška  $E(\mathbf{w}|\mathcal{D})$ ?

- ☐ A 165.89    ☐ B 31.70    ☐ C 191.95    ☐ D 230.98

- 4 (P) Koristimo regresiju za predviđanje uspjeha na studiju. Kao značajke možemo koristiti ocjene u četiri razreda srednje škole (značajke  $x_1$ – $x_4$ ), prosjek ocjena sva četiri razreda ( $x_5$ ) te uspjeh iz matematike ( $x_6$ ) i fizike ( $x_7$ ) na državnoj maturi (ukupno 7 značajki). Ne moramo iskoristiti sve značajke, ali ih želimo iskoristiti što više. Za preslikavanje u prostor značajki koristimo preslikavanje s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki (npr.  $x_1x_2$ ) i interakcije trojki (npr.  $x_1x_2x_3$ ) između svih značajki koje koristimo. Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?

- ☐ A 63    ☐ B 48    ☐ C 75    ☐ D 38

### Linearni klasifikacijski modeli (3 pitanja)

- 5 (T) Postoji veza između logističke regresije i modela neuronske mreže. Koja je veza između ta dva modela?

- ☐ A Neuronska mreža optimirana algoritmom propagacije pogreške unazad istovjetna je logističkoj regresiji optimiranoj stohastičkim gradijentnim spustom  
☐ B Multinomijalna logistička regresija s aktivacijskom funkcijom softmax istovjetna je dvoslojnoj neuronskoj mreži sa više neurona u izlaznom sloju  
☐ C Logistička regresija s linearnim jezgrenim funkcijama istovjetna je neuronskoj mreži sa linearnom aktivacijskom funkcijom i kvadratnom funkcijom pogreške  
☐ D Logistička regresija koja kao adaptivne bazne funkcije koristi logističku regresiju istovjetna je neuronskoj mreži sa sigmoidnom aktivacijskom funkcijom

**6** (P) Na skupu označenih primjera treniramo tri modela: (1) model neregularizirane logističke regresije (NR), (2) model L2-regularizirane logističke regresije (L2R) i (3) perceptron s funkcijom preslikavanja. Sva tri modela koriste iste značajke. Za sva tri algoritma promatramo iznos empirijske pogreške učenja kroz iteracije optimizacijskog postupka. Za algoritam perceptrona opažamo da empirijska pogreška učenja ne konvergira. **Kako se u ovom slučaju ponaša empirijska pogreška učenja kroz iteracije za dva spomenuta modela logističke regresije, NR i L2R?**

- ☐ A Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira
- ☐ B Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali pogreška modela NR doseže manju vrijednost
- ☐ C Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija
- ☐ D Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste

**7** (N) Raspolažemo sljedećim skupom za učenje u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u peterodimenzijski prostor značajki, koja je definirana na sljedeći način:

$$\phi(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Početne težine perceptrona neka su sljedeće:

$$\mathbf{w} = (1, 0, -1, 2, 3, 0)$$

**Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

- ☐ A 22
- ☐ B 6
- ☐ C 16
- ☐ D 32

### Jezgrene i neparametarske metode (4 pitanja)

**8** (T) Kod algoritma SVM preporuča se napraviti skaliranje značajki. U protivnom, pri izračunu skalarnog produkta, značajke s većim rasponom (većom skalom) dominirat će nad značajkama s manjim rasponom (manjom skalom). Međutim, skaliranje značajki nije uvijek nužno. **Kada nije potrebno napraviti skaliranje značajki, i zašto?**

- ☐ A Kada se koristi linearna jezgra i značajke su centrirane oko nule, jer se onda ne računa skalarni produkt između značajki
- ☐ B Kada se koristi dualna formulacija SVM-a i algoritam SMO, jer se tada implicitno provodi L1-regularizacija
- ☐ C Kada se koristi RBF jezgra sa Mahalanobisovom udaljenošću, jer ta udaljenost uzima u obzir varijancu značajki
- ☐ D Kada se koristi Gaussova jezgrena funkcija, jer ta jezgra implicitno inducira beskonačnodimenzijski prostor značajki

**9** (P) Na 500 primjera sa 80 značajki treniramo rijetki jezgreni stroj s Gaussovim jezgrama. Sve Gaussove jezgre imaju istu varijancu. Nakon treniranja, dobivamo model koji ima 38 prototipa. **Koliko parametara moramo optimirati te koliko parametara ima naučeni model?**

- ☐ A Optimiramo 81 parametar, a naučeni model ima 3040 parametara
- ☐ B Optimiramo 81 parametar, a naučeni model ima 3079 parametara
- ☐ C Optimiramo 501 parametara, a naučeni model ima 3079 parametara
- ☐ D Optimiramo 501 parametara, a naučeni model ima 3541 parametar



- 10** (N) U ulaznome prostoru dimenzije  $n = 3$  trenirali smo model SVM-a s linearnom jezgrom. Potporne vektore naučenog modela čine označeni primjeri  $((2, -5, 15), -1)$ ,  $((1, 8, -305), -1)$  i  $((1, -6, 225), +1)$ , a njima odgovarajući dualni koeficijenti su  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.8$  i  $\alpha_3 = 0.9$ . Treniranje smo proveli na skaliranim značajkama: svaku smo značajku  $x_j$  standardizirali primjenom transformacije  $\frac{x_j - \mu_j}{\sigma_j}$ , gdje su  $\mu_j$  i  $\sigma_j$  srednja vrijednost odnosno varijanca značajke  $x_j$  u skupu označenih podataka  $\mathcal{D}$ . Parametri skaliranja su  $\boldsymbol{\mu} = (15, -2, 100)$  i  $\boldsymbol{\sigma} = (4, 1, 12)$ . Model SVM-a koristimo za predikciju klase primjera  $\mathbf{x} = (1, -2, 15)$ . **Koliko će se promijeniti izlaz modela ako kod predikcije propustimo skalirati značajke primjera  $\mathbf{x}$ ?**

☐ A  $-373.22$    ☐ B  $+907.43$    ☐ C  $-739.13$    ☐ D  $+541.53$

- 11** (N) Na skupu primjera za učenje iz ulaznog prostora  $n = 3$  trenirali smo SVM s polinomijalnom jezgrenom funkcijom  $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 2)^3$ . Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((9, 30, 21), -1)$$

$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-11, -26, -15), -1)$$

$$(\mathbf{x}^{(3)}, y^{(3)}) = ((-1, -7, -6), +1)$$

Lagrangeovi koeficijenti su  $\alpha_1 = 2.214 \cdot 10^{-8}$ ,  $\alpha_2 = 3.803 \cdot 10^{-8}$  i  $\alpha_3 = 6.017 \cdot 10^{-8}$ . **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze  $h(\mathbf{x})$  za primjer  $\mathbf{x} = (1, 1, 25)$ .**

☐ A  $+0.947$    ☐ B  $-2.330$    ☐ C  $-0.676$    ☐ D  $+1.434$

## Procjena parametara i Bayesov klasifikator (3 pitanja)

- 12** (N) Raspoložemo sljedećim skupom označenih primjera:

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

Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klasa  $p(x|y)$ . Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:

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

Parametre  $\mu$  i  $\sigma^2$  gustoće vjerojatnosti  $p(x|y)$  procjenjujemo MLE-om. Neka su  $\mu_1$  i  $\sigma_1^2$  parametri gustoće vjerojatnosti  $p(x|y = 1)$  dobiveni MLE-om na podskupu primjera  $\mathcal{D}_{y=1}$ . **Koliko iznosi log-izglednost  $\mathcal{L}(\mu_1, \sigma_1^2 | \mathcal{D}_{y=1})$ ?**

☐ A  $-11.58$    ☐ B  $-7.42$    ☐ C  $-9.32$    ☐ D  $-10.47$

- 13** (T) Polunaivan Bayesov klasifikator združuje u jedan faktor varijable kod kojih postoji statistička zavisnost. Za procjenu statističke zavisnosti između varijabli može se upotrijebiti Kullback-Leiblerova divergencija (KL-divergencija). **Kako se pomoću KL-divergencije može izračunati koliko su varijable zavisne?**

- ☐ A Što su varijable više zavisne, to je veća KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti
- ☐ B Što su varijable manje zavisne, to je manja KL-divergencija između marginalnih vjerojatnosti i uvjetne vjerojatnosti
- ☐ C Što su varijable više zavisne, to je manja KL-divergencija između zajedničke vjerojatnosti i faktorizirane vjerojatnosti
- ☐ D Što su varijable manje zavisne, to je veća KL-divergencija između faktorizirane vjerojatnosti i marginalne vjerojatnosti

- 14** (P) Gausovim Bayesovim klasifikatorom rješavamo problem klasifikacije u  $K = 10$  klasa sa  $n = 100$  značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu  $\Sigma$  možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$  : Značajke imaju različite varijance, ali iste za sve klase, te nisu korelirane

$\mathcal{H}_2$  : Značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu

$\mathcal{H}_3$  : Između značajki postoje korelacije, ali se one ne razlikuju između klasa

Neka ‘ $\supset$ ’ označava relaciju “složeniji od”, a neka ‘ $>$ ’ označava relaciju “ima više parametara od”. Što možemo zaključiti o složenosti i broju parametara za gornja četiri modela?

- ☐ A  $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2, \mathcal{H}_3 \supset \mathcal{H}_1$       ☐ C  $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3, \mathcal{H}_1 \supset \mathcal{H}_2$   
☐ B  $\mathcal{H}_3 > \mathcal{H}_2 > \mathcal{H}_1, \mathcal{H}_3 \supset \mathcal{H}_1 \supset \mathcal{H}_2$       ☐ D  $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2, \mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$

### Probabilistički grafički modeli (3 pitanja)

- 15 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem  $v, w, x, y, z$ . Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

$$\{v, w\} \perp y | x \quad \{v, x\} \perp z | \{w, y\}$$

Primjenom algoritma d-odvajanja ispitujemo zavisnosti između parova varijabli. Koje od sljedećih tvrdnji o nezavisnosti vrijede u ovoj Bayesovoj mreži?

- ☐ A  $v \perp y | z$     ☐ B  $w \perp y | z$     ☐ C  $v \perp z | w$     ☐ D  $v \perp y | x$

- 16 (N) Razmotrite Bayesovu mrežu koja odgovara faktorizaciji  $P(w, x, y, z) = P(w)P(x)P(y|w, x)P(z|x)$ . Sve varijable su binarne. Vrijedi  $P(w = 1) = 0.1$ ,  $P(x = 1) = 0.2$ ,  $P(z = 1|x = 0) = 0.9$  i  $P(z = 1|x = 1) = 0.7$ . Tablica uvjetnih vjerojatnosti za čvor  $y$  je sljedeća:

$w$	$x$	$p(y = 1 w, x)$
0	0	0
0	1	0.4
1	0	0.2
1	1	0.7

Postupkom uzorkovanja s odbijanjem želimo procijeniti parametar  $\mu$  uvjetne distribucije  $P(x = 0|y = 1, z = 0)$ . Uzorkovanje smo ponovili ukupno  $N = 500$  puta, od čega smo neke vektore morali odbaciti, pa je naš uzorak manji od  $N$ . Na temelju dobivenog uzorka parametar  $\mu$  procjenjujemo MAP procjeniteljem uz  $\alpha = \beta = 2$ . Koliko iznosi očekivana MAP procjena parametra  $\mu$ ?

- ☐ A 0.0786    ☐ B 0.1877    ☐ C 0.1274    ☐ D 0.0490

- 17 (T) Bayesova mreža na sažet način definira zajedničku distribuciju vjerojatnosti uz određene pretpostavke uvjetne nezavisnosti. Neka je jedna takva pretpostavka  $x_1 \perp \{x_2, x_3\} | x_4$ . Koji je efekt uvođenja ove pretpostavke na graf Bayesove mreže?

- ☐ A Dodavanje tri brida    ☐ C Uklanjanje dva brida  
☐ B Dodavanje dva brida    ☐ D Uklanjanje tri brida

### Grupiranje (3 pitanja)

- 18 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo  $N = 5$  primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom:

$$\begin{pmatrix} 1.0 & 0.2 & 0.8 & 0.1 & 0.4 \\ 0.2 & 1.0 & 0.3 & 0.3 & 0.7 \\ 0.8 & 0.3 & 1.0 & 0.6 & 0.5 \\ 0.1 & 0.3 & 0.6 & 1.0 & 0.4 \\ 0.4 & 0.7 & 0.5 & 0.4 & 1.0 \end{pmatrix}$$

Provedite grupiranje algoritmom HAC s potpunim povezivanjem. Pritom u svakoj iteraciji bilježite na kojoj razini sličnosti se odvija stapanje dviju grupa. Koliko iznosi zbroj po svim razinama sličnosti na kojima se odvija stapanje grupa?

- ☐ A 2.4    ☐ B 1.9    ☐ C 1.8    ☐ D 2.0

**19** (T) Algoritam K-sredina je iterativan algoritam za nalaženje parametara  $b_k^{(i)}$  (pripadnosti primjera grupama) i  $\mu_k$  (centroidi grupa) koji minimiziraju kriterijsku funkciju  $J$  za  $N$  primjera i  $K$  grupa. Algoritam funkciju  $J$  optimizira iterativno, jer rješenje u zatvorenoj formi ne postoji. **Zbog čega za problem minimizacije funkcije  $J$  ne postoji rješenje u zatvorenoj formi?**

- ☐ A Jer za  $b_k^{(i)}$  mora vrijediti ograničenje  $\sum_k b_k^{(i)} = 1$  i  $b_k^{(i)} \geq 0$
- ☐ B Jer  $J$  ovisi o  $K$  i inicijalnom odabiru za  $\mu_k$ , pa rješenje nije jedinstveno
- ☐ C Jer  $b_k^{(i)}$  ovisi o  $N$ , broj vektora  $\mu_k$  ovisi o  $K$ , a  $K$  je odozgo ograničen sa  $N$
- ☐ D Jer  $b_k^{(i)}$  ovisi o vektorima  $\mu_k$ , a vektor  $\mu_k$  ovisi o vrijednostima  $b_k^{(i)}$

**20** (P) Algoritmom GMM grupiramo primjere u dvodimenzijaskome ulaznom prostoru. Skup podataka u stvarnosti je uzorkovan iz zajedničke distribucije koja se može opisati sljedećim mješavinskim modelom:

$$p(\mathbf{x}) = \sum_{j=1}^3 \frac{1}{3} \mathcal{N}(\mu_j, \Sigma_j) \quad \text{gdje} \quad \mu_1 = (5, 5), \mu_2 = (5, 10), \mu_3 = (-10, -10), \Sigma_1 = \Sigma_2 = \Sigma_3 = 2\mathbf{I}$$

Skup  $\mathcal{D}$  grupiramo u  $K = 2$  grupe. Pritom isprobavamo tri modela, koji se međusobno razlikuju po pretpostavkama na kovarijacijsku matricu. Konkretno: dijeljena i puna kovarijacijska matrica ( $\mathcal{H}_1$ ), nedijeljena i dijagonalna kovarijacijska matrica ( $\mathcal{H}_2$ ) i nedijeljena i izotropna kovarijacijska matrica ( $\mathcal{H}_3$ ). Neka je  $\mathcal{L}_i$  izglednost parametara dobivena modelom  $\mathcal{H}_i$  nakon konvergencije algoritma. Za inicijalizaciju središta koristi se algoritam K-means++. **Što su očekivani odnosi između izglednosti za ova tri modela?**

- ☐ A  $\mathcal{L}_1 = \mathcal{L}_2 > \mathcal{L}_3$     ☐ B  $\mathcal{L}_2 > \mathcal{L}_1, \mathcal{L}_2 > \mathcal{L}_3$     ☐ C  $\mathcal{L}_1 > \mathcal{L}_2 > \mathcal{L}_3$     ☐ D  $\mathcal{L}_1 > \mathcal{L}_3, \mathcal{L}_2 > \mathcal{L}_3$

## Vrednovanje modela (2 pitanja)

**21** (N) Na ispitnome skupu evaluiramo klasifikator sa  $K = 3$  klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{ccc} & y = 1 & y = 2 & y = 3 \\ \begin{array}{c} y = 1 \\ y = 2 \\ y = 3 \end{array} & \begin{pmatrix} 3 & 5 & 11 \\ 6 & 5 & 6 \\ 4 & 2 & 26 \end{pmatrix} \end{array}$$

Izračunajte mikro-F1 ( $F_1^\mu$ ) i makro-F1 ( $F_1^M$ ) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere,  $F_1^\mu - F_1^M$ ?**

- ☐ A 0.114    ☐ B 0.059    ☐ C 0.091    ☐ D 0.023

**22** (P) Na istom skupu označenih primjera vrednujemo četiri binarna klasifikatora s vjerojatnosnim izlazima. Za vrednovanje koristimo krivulju ROC. Svaki smo klasifikator ispitali s tri vrijednosti klasifikacijskog praga te smo za te vrijednosti izračunali FPR (stopa lažnog alarma) i TPR (odziv). Dobiveni parovi vrijednosti (FPR, TPR) za sva četiri klasifikatora su sljedeći:

$$\begin{array}{ll} h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) & h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8) \\ h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) & h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0) \end{array}$$

Skicirajte odgovarajuće krivulje ROC, linearno interpolirajući između izmjerenih točaka, a dodajte i točke koje odgovaraju krajnjim vrijednostima klasifikacijskog praga (0 i 1). Pritom, ako je neki klasifikator lošiji od nasumičnog klasifikatora, umjesto tog klasifikatora razmatrajte njegovu popravljenu varijantu koju ćete dobiti invertiranjem izlaza ( $1 - h(\mathbf{x})$  umjesto  $h(\mathbf{x})$ ). **Koji su od ispitanih klasifikatora (eventualno nakon popravka) najbolji prema krivulji ROC?**

- ☐ A  $h_1$  i  $h_4$     ☐ B  $h_2$  i  $h_3$     ☐ C  $h_1$  i  $h_3$     ☐ D  $h_3$  i  $h_4$



Grupa	1	2	3	4	5	6	7	8	9	1	1	1	1	1	1	1	1	1	1	2	2	2
A	A	C	A	D	B	D	C	A	C	D	D	B	A	D	C	C	C	A	B	C	B	B
B	B	A	C	D	D	A	D	D	D	C	B	D	C	B	D	B	A	C	D	B	A	D
C	A	B	C	C	D	C	C	A	D	D	C	A	B	A	D	C	C	D	A	B	B	D
D	D	B	A	C	C	A	D	D	B	B	B	A	B	B	B	C	C	C	D	A	C	A
E	C	B	B	B	C	D	A	C	C	B	B	D	D	D	A	A	B	B	C	B	D	A
F	C	C	B	B	D	B	D	C	C	B	C	D	A	A	D	A	C	B	D	B	C	A