Pismeni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

*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 **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Razmatramo model jednostavne regresije, $h(x; w_0, w_1) = w_0 + w_1 x$. Model linearne regresije inače koristi funkciju kvadratnog gubitka:

$$L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$$

Međutim, u našoj implementaciji greškom smo funkciju gubitka definirali ovako:

$$L(y, h(\mathbf{x})) = (y + 2h(\mathbf{x}))^2$$

S tako pogrešno definiranom funkcijom gubitka postupkom najmanjih kvadrata treniramo naš model na skupu primjera čije su oznake uzorkovane iz distribucije $\mathcal{N}(-1 + 2x, \sigma^2)$, gdje je varijanca σ^2 razmjerno malena (tj. nema mnogo šuma). **Koji vektor težina (w_0, w_1) očekujemo približno dobiti kao rezultat najmanjih kvadrata?**

A $(1, -2)$ B $(\frac{1}{2}, -1)$ C $(2, -4)$ D $(2, -\frac{1}{2})$

2 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Bez induktivne pristranosti, učenje na temelju podataka ne bi imalo smisla, odnosno algoritam bez induktivne pristranosti ne bi mogao ništa naučiti. **Zašto strojno učenje bez induktivne pristranosti nije moguće?**

A Model bi bio presložen te bi imao preveliku pogrešku na skupu za ispitivanje

B Minimizacija empirijske pogreške učenja ne bi imala rješenje u zatvorenoj formi

C Primjeri ne bi nužno bili linearno odvojivi

D Svako označavanje neviđenih primjera bilo bi jednako vjerojatno

3 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

A $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

B $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

C $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

D $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

4 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

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 31.70 B 165.89 C 230.98 D 191.95

Linearni klasifikacijski modeli *(3 pitanja)*

5 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

U jednoj iteraciji treniranja modela vektor parametara jednak je

$$\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$$

Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((-0.5, 2), 1)$?

A 4.00 B 1.28 C 2.48 D 0.70

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. Nakon određenog broja iteracija, algoritam perceptrona uspješno se zaustavio s rješenjem. **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 modelu NR za to treba više iteracija

B Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste

C Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira

D Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija

7 (T) Algoritam linearne regresije može se pokušati primijeniti na klasifikacijski problem, kao što smo pokušali na predavanjima, međutim to nije dobro funkcioniralo. Razmotrite tri komponente algoritma linearne regresije: model (M), funkciju gubitka (G) i optimizacijski postupak (O). Također, prisjetite se algoritma logističke regresije, koji je dobar klasfikacijski algoritam. Želimo preinačiti komponente algoritma linearne regresije tako da iz njega dobijemo nov algoritam koji klasifikaciju radi bolje od linearnog modela regresije, ali koji je drugačiji od logističke regresije. **Uz koje tri komponente bismo dobili takav algoritam?**

A M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y) \ln\big(1 - h(\mathbf{x})\big)$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$

B M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y) \ln\big(1 - h(\mathbf{x})\big)$ O: $\mathbf{w}^*$ izračunat gradijentnim spustom

C M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^*$ izračunat gradijentnim spustom

D M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (-3, -5, 3)$.**

A -3.553 B -4.093 C -2.330 D $+1.434$

9 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2,0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina šira od stare?**

A $\frac{1}{2}\sqrt{13}$ puta B $\frac{1}{2}\sqrt{10}$ puta C $\frac{5}{4}\sqrt{2}$ puta D $\frac{1}{2}\sqrt{17}$ puta

10 (T) Važna prednost jezgrenih strojeva je mogućnost primjene jezgrenog trika. Ta je prednost pogotovo očita kada prostor ulaznih primjera $\mathcal{X}$ nije euklidski vektorski prostor, odnosno kada primjere nije moguće prikazati kao vektore realnih brojeva. **Koja je prednost primjene jezgrenog trika u takvom slučaju?**

A Jezgrena funkcija mjeri sličnost između primjera, čime se primjeri efektivno preslikavaju u beskonačnodimenzijski prostor značajki

B Jezgrena funkcija može biti mjera sličnost između nevektoriziranih primjera, što implicitno inducira vektorski prostor značajki

C Jezgrenim trikom implicitno se ostvaruje nelinearnost prostora značajki, što povećava kapacitet modela i povećava njegovu točnost

D Umjesto vektorizacije primjera, dovoljno je defnirati nelinearnu funkciju preslikavanja iz ulaznog prostora u prostor značajki

11 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 22 & 0 & -5 \\ 1 & 6 & 28 & 8 & 2 \\ 1 & -6 & -15 & -9 & -6 \\ 1 & -11 & -34 & -16 & -9 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (-0.0880, +0.0093, +0.0493, +0.0120, +0.0013)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0045 B 0.0089 C 0.0024 D 0.0013

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$. Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

gdje je Σ_j matrica kovarijacije za klasu j. Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.1, \quad \hat{\boldsymbol{\mu}}_1 = (1,0,-2), \quad \hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{pmatrix}, \quad \hat{\mu}_2 = 0.9, \quad \hat{\boldsymbol{\mu}}_2 = (2,-1,5), \quad \hat{\Sigma}_2 = \begin{pmatrix} 5.0 & -0.222 & -0.444 \\ -0.222 & 2.778 & -0.333 \\ -0.444 & -0.333 & 0.444 \end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1, 2$ (zaokružite elemente tako dobivene kovarijacijske matrice na jednu decimalu). Zanima nas klasifikacija modela za primjer $\mathbf{x} = (0, 0, 0)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

A -5.902 B -12.788 C -8.513 D -10.913

13 (T) Umjesto naivnog Bayesov klasifikatora, ponekad je bolje koristiti polunaivan Bayesov klasifikator? **Po čemu se polunaivan Bayesov klasifikator (SNBC) razlikuje od naivnog Bayesovog klasifikatora (NBC)?**

A SNBC zajedničku vjerojatnost faktorizira u manje faktora, pa ima i manje parametara od NBC

B SNBC ima više parametara nego NBC, ali manje bridova kada ga se prikaže kao Bayesovu mrežu

C SNBC ima više parametara od NBC te modelira zavisnosti između značajki

D SNBC pretpostavlja zavisnosti između značajki i klase, ali ima manje parametara od NBC

14 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2, 1, 4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_3 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 8$ B $\alpha > 2$ C $\alpha > 7$ D $\alpha > 5$

Probabilistički grafički modeli *(3 pitanja)*

15 (N) Bayesovom mrežom s četiri varijable modeliramo konstrukte pozitivne psihologije. Koristimo binarne varijable *Ljubav* (L), *Sreća* (S), *Tjeskoba* (T), s vrijednostima 0 (nema) i 1 (ima), te ternarnu varijablu *Novac* (N), s vrijednostima 0 (nema), 1 (ima malo) i 2 (ima puno). Strukturu Bayesove mreže definirali smo tako da ona modelira sljedeće pretpostavljene kauzalne odnose: L uzrokuje S, a N uzrokuje S i T. Tako definiranu Bayesovu mrežu zatim treniramo na sljedećem skupu od $N = 7$ primjera:

L	N	S	T
1	0	1	0
1	0	1	0
0	2	0	1
1	2	1	1
1	1	1	0
0	0	0	0
0	2	1	0

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu), što je istovjetno Laplaceovom zaglađivanju MLE procjene. Na kraju nas, naravno, zanima koja je vjerojatnost života uz ljubav, sreću i malo novaca. Napravite potrebne MAP-procjene parametara. **Koliko iznosi zajednička vjerojatnost $P(L = 1, S = 1, N = 1)$?**

A 0.074   B 0.833   C 0.023   D 0.143

16 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 24   B 18   C 21   D 23

17 (T) Skriveni Markovljev model (HMM) posebna je vrsta Bayesove mreže. **Na što se odnosi pridjev "skriveni" u nazivu tog modela?**

A Opažene varijable ovise samo o trenutačnom stanju i prethodno opaženoj varijabli

B Neke varijable modela nisu opažene u podatcima, ali zavise o opaženim varijablama

C Model opisuje prijelaze između stanja, a trenutačno stanje ovisi samo o prethodnom stanju

D Stanja modela poredana su u lanac, a izlazi modela vidljivi su samo za posljednje stanje

Grupiranje *(3 pitanja)*

18 (T) Algoritam hijerarhijskog aglomerativnog grupiranja (HAC) može koristiti različite vrste povezanosti, uključivo jednostruku (*single*), potpunu (*complete*) i prosječnu (*average*). Označimo funkcije udaljenosti koje odgovaraju ovim povezanostima sa d_s, d_c odnosno d_a. Neka su G_i i G_j dvije neprazne grupe primjera. **Koji odnosi vrijede između vrijednosti ovih funkcija?**

A $d_s(G_i, G_j) \leqslant d_c(G_i, G_j) \geqslant d_a(G_i, G_j)$

B $d_s(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_c(G_i, G_j)$

C $d_s(G_i, G_j) \leqslant d_a(G_i, G_j)$, $d_c(G_i, G_j) \leqslant d_a(G_i, G_j)$

D $d_c(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_s(G_i, G_j)$

19 (N) Particijskim algoritmom grupiranja grupiramo $N = 1000$ primjera. Na temelju znanja o problemu zaključili smo da bi primjeri trebali formirati $K = 3$ grupe, pa smo s tim brojem grupa proveli grupiranje. Kako bismo evaluirali točnost grupiranja, slučajnim odabirom smo iz skupa primjera uzorkovali 10 primjera, ručno smo označili primjere iz tog uzorka, i zatim na tom uzorku računamo Randov indeks. Označavanje smo proveli tako da smo svakom primjeru iz uzorka dodijelili oznaku točne grupe. Oznake grupe dobivene algoritmom grupiranja y_{pred} i oznake točnih grupa y_{true} za svih deset primjera u uzorku su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y_{pred}^{(i)}$	0	1	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.53 B 0.49 C 0.62 D 0.58

20 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika $J^+ - J^*$?**

A 8 B 6 C 4 D 12

Vrednovanje modela *(2 pitanja)*

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{cc} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 30 & 8 & 3 \\ 11 & 25 & 5 \\ 4 & 12 & 2 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.155 B 0.185 C 0.205 D 0.085

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:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

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_1 i h_3 C h_1 i h_4 D h_3 i h_4

Pismeni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

*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 **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Razmatramo model jednostavne regresije, $h(x; w_0, w_1) = w_0 + w_1 x$. Model linearne regresije inače koristi funkciju kvadratnog gubitka:

$$L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$$

Međutim, u našoj implementaciji greškom smo funkciju gubitka definirali ovako:

$$L(y, h(\mathbf{x})) = (y + 2h(\mathbf{x}))^2$$

S tako pogrešno definiranom funkcijom gubitka postupkom najmanjih kvadrata treniramo naš model na skupu primjera čije su oznake uzorkovane iz distribucije $\mathcal{N}(-1 + 2x, \sigma^2)$, gdje je varijanca σ^2 razmjerno malena (tj. nema mnogo šuma). **Koji vektor težina (w_0, w_1) očekujemo približno dobiti kao rezultat najmanjih kvadrata?**

A $(1, -2)$ B $(\frac{1}{2}, -1)$ C $(2, -\frac{1}{2})$ D $(2, -4)$

2 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

A $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

B $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

C $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

D $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

3 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

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 31.70 B 165.89 C 191.95 D 230.98

4 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Bez induktivne pristranosti, učenje na temelju podataka ne bi imalo smisla, odnosno algoritam bez induktivne pristranosti ne bi mogao ništa naučiti. **Zašto strojno učenje bez induktivne pristranosti nije moguće?**

A Minimizacija empirijske pogreške učenja ne bi imala rješenje u zatvorenoj formi

B Oznaka niti jednog neviđenog primjera ne bi bila jednoznačno određena

C Prostor parametara bio bi neograničen, tj. postojalo bi beskonačno mnogo vektora parametara

D Svaka bi se hipoteza prilagodila šumu i ostvarila empirijsku pogrešku nula

Linearni klasifikacijski modeli *(3 pitanja)*

5 (T) Algoritam linearne regresije može se pokušati primijeniti na klasifikacijski problem, kao što smo pokušali na predavanjima, međutim to nije dobro funkcioniralo. Razmotrite tri komponente algoritma linearne regresije: model (M), funkciju gubitka (G) i optimizacijski postupak (O). Također, prisjetite se algoritma logističke regresije, koji je dobar klasfikacijski algoritam. Želimo preinačiti komponente algoritma linearne regresije tako da iz njega dobijemo nov algoritam koji klasifikaciju radi bolje od linearnog modela regresije, ali koji je drugačiji od logističke regresije. **Uz koje tri komponente bismo dobili takav algoritam?**

- A M: $h(\mathbf{x}) = \sigma(\mathbf{w}^\mathrm{T}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y) \ln\big(1 - h(\mathbf{x})\big)$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$
- B M: $h(\mathbf{x}) = \mathbf{w}^\mathrm{T}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$
- C M: $h(\mathbf{x}) = \sigma(\mathbf{w}^\mathrm{T}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- D M: $h(\mathbf{x}) = \mathbf{w}^\mathrm{T}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y) \ln\big(1 - h(\mathbf{x})\big)$ O: $\mathbf{w}^*$ izračunat gradijentnim spustom

6 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

U jednoj iteraciji treniranja modela vektor parametara jednak je

$$\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$$

Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((0.5, 2), 1)$?

A 0.70 B 2.48 C 1.28 D 4.00

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. Nakon određenog broja iteracija, algoritam perceptrona uspješno se zaustavio s rješenjem. **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 obje stagniraju nakon određenog broja iteracija, ali modelu NR za to treba više iteracija
- C Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija
- 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 (T) Važna prednost jezgrenih strojeva je mogućnost primjene jezgrenog trika. Ta je prednost pogotovo očita kada prostor ulaznih primjera $\mathcal{X}$ nije euklidski vektorski prostor, odnosno kada primjere nije moguće prikazati kao vektore realnih brojeva. **Koja je prednost primjene jezgrenog trika u takvom slučaju?**

- A Jezgrenim trikom implicitno se ostvaruje nelinearnost prostora značajki, što povećava kapacitet modela i povećava njegovu točnost
- B Jezgrena funkcija mjeri sličnost između primjera, čime se primjeri efektivno preslikavaju u beskonačnodimenzijski prostor značajki
- C Umjesto vektorizacije primjera, dovoljno je defnirati nelinearnu funkciju preslikavanja iz ulaznog prostora u prostor značajki
- D Jezgrena funkcija može biti mjera sličnost između nevektoriziranih primjera, što implicitno inducira vektorski prostor značajki

9 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2,0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina šira od stare?**

[A] $\frac{5}{4}\sqrt{2}$ puta [B] $\frac{1}{2}\sqrt{10}$ puta [C] $\frac{1}{2}\sqrt{13}$ puta [D] $\frac{1}{2}\sqrt{17}$ puta

10 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x},\mathbf{z}) = (\mathbf{x}^T\mathbf{z}+1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$\begin{aligned}(\mathbf{x}^{(1)}, y^{(1)}) &= ((11,32,21),-1)\\ (\mathbf{x}^{(2)}, y^{(2)}) &= ((5,11,6),-1)\\ (\mathbf{x}^{(3)}, y^{(3)}) &= ((6,12,6),+1)\end{aligned}$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (-3,-5,3)$.**

[A] -2.330 [B] $+1.434$ [C] -3.553 [D] -4.093

11 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 22 & 0 & -5\\ 1 & 6 & 28 & 8 & 2\\ 1 & -6 & -15 & -9 & -6\\ 1 & -11 & -34 & -16 & -9 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (-0.0880, +0.0093, +0.0493, +0.0120, +0.0013)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

[A] 0.0013 [B] 0.0089 [C] 0.0024 [D] 0.0045

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$. Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

gdje je Σ_j matrica kovarijacije za klasu j. Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.1,\quad \hat{\boldsymbol{\mu}}_1 = (1,0,-2),\quad \hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4\\ 2 & 5 & 3\\ 4 & 3 & 6\end{pmatrix},\quad \hat{\mu}_2 = 0.9,\quad \hat{\boldsymbol{\mu}}_2 = (2,-1,5),\quad \hat{\Sigma}_2 = \begin{pmatrix} 5.0 & -0.222 & -0.444\\ -0.222 & 2.778 & -0.333\\ -0.444 & -0.333 & 0.444\end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1, 2$. Zanima nas klasifikacija modela za primjer $\mathbf{x} = (1, 0, 1)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

A -5.902 B -8.513 C -12.788 D -10.913

13 (T) Umjesto naivnog Bayesov klasifikatora, ponekad je bolje koristiti polunaivan Bayesov klasifikator? **Po čemu se polunaivan Bayesov klasifikator (SNBC) razlikuje od naivnog Bayesovog klasifikatora (NBC)?**

A SNBC ima više parametara nego NBC, ali manje bridova kada ga se prikaže kao Bayesovu mrežu

B SNBC pretpostavlja zavisnosti između značajki i klase, ali ima manje parametara od NBC

C SNBC ima više parametara od NBC te modelira zavisnosti između značajki

D SNBC zajedničku vjerojatnost faktorizira u manje faktora, pa ima i manje parametara od NBC

14 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2, 1, 4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_3 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 5$ B $\alpha > 8$ C $\alpha > 2$ D $\alpha > 7$

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 18 B 24 C 23 D 21

16 (N) Bayesovom mrežom s četiri varijable modeliramo konstrukte pozitivne psihologije. Koristimo binarne varijable *Ljubav* (L), *Sreća* (S), *Tjeskoba* (T), s vrijednostima 0 (nema) i 1 (ima), te ternarnu varijablu *Novac* (N), s vrijednostima 0 (nema), 1 (ima malo) i 2 (ima puno). Strukturu Bayesove mreže definirali smo tako da ona modelira sljedeće pretpostavljene kauzalne odnose: L uzrokuje S, a N uzrokuje S i T. Tako definiranu Bayesovu mrežu zatim treniramo na sljedećem skupu od $N = 7$ primjera:

L	N	S	T
1	0	1	0
1	0	1	0
0	2	0	1
1	2	1	1
1	1	1	0
0	0	0	0
0	2	1	0

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu), što je istovjetno Laplaceovom zaglađivanju MLE procjene. Na kraju nas, naravno, zanima koja je vjerojatnost života uz ljubav, sreću i malo novaca. Napravite potrebne MAP-procjene parametara. **Koliko iznosi zajednička vjerojatnost $P(L = 1, S = 1, N = 1)$?**

A 0.833 B 0.023 C 0.074 D 0.143

17 (T) Skriveni Markovljev model (HMM) posebna je vrsta Bayesove mreže. **Na što se odnosi pridjev "skriveni" u nazivu tog modela?**

- A Stanja modela poredana su u lanac, a izlazi modela vidljivi su samo za posljednje stanje
- B Model opisuje prijelaze između stanja, a trenutačno stanje ovisi samo o prethodnom stanju
- C Opažene varijable ovise samo o trenutačnom stanju i prethodno opaženoj varijabli
- D Neke varijable modela nisu opažene u podatcima, ali zavise o opaženim varijablama

Grupiranje *(3 pitanja)*

18 (T) Algoritam hijerarhijskog aglomerativnog grupiranja (HAC) može koristiti različite vrste povezanosti, uključivo jednostruku (*single*), potpunu (*complete*) i prosječnu (*average*). Označimo funkcije udaljenosti koje odgovaraju ovim povezanostima sa d_s, d_c odnosno d_a. Neka su G_i i G_j dvije neprazne grupe primjera. **Koji odnosi vrijede između vrijednosti ovih funkcija?**

- A $d_c(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_s(G_i, G_j)$
- B $d_s(G_i, G_j) \leqslant d_a(G_i, G_j)$, $d_c(G_i, G_j) \leqslant d_a(G_i, G_j)$
- C $d_s(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_c(G_i, G_j)$
- D $d_s(G_i, G_j) \leqslant d_c(G_i, G_j) \geqslant d_a(G_i, G_j)$

19 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

A 4 B 12 C 6 D 8

20 (N) Particijskim algoritmom grupiranja grupiramo $N = 1000$ primjera. Na temelju znanja o problemu zaključili smo da bi primjeri trebali formirati $K = 3$ grupe, pa smo s tim brojem grupa proveli grupiranje. Kako bismo evaluirali točnost grupiranja, slučajnim odabirom smo iz skupa primjera uzorkovali 10 primjera, ručno smo označili primjere iz tog uzorka, i zatim na tom uzorku računamo Randov indeks. Označavanje smo proveli tako da smo svakom primjeru iz uzorka dodijelili oznaku točne grupe. Oznake grupe dobivene algoritmom grupiranja y_{pred} i oznake točnih grupa y_{true} za svih deset primjera u uzorku su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y^{(i)}_{pred}$	0	1	2	1	1	0	0	2	1	1
$y^{(i)}_{true}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.53 B 0.58 C 0.49 D 0.62

Vrednovanje modela *(2 pitanja)*

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{cc} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 30 & 8 & 3 \\ 11 & 25 & 5 \\ 4 & 12 & 2 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.205 B 0.085 C 0.155 D 0.185

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:

$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4)$ $h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$

$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6)$ $h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$

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_2 B h_2 i h_3 C h_1 i h_4 D h_3 i h_4

Pismeni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

*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 **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

- A $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$
- B $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$
- C $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$
- D $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

2 (P) Razmatramo model jednostavne regresije, $h(x; w_0, w_1) = w_0 + w_1 x$. Model linearne regresije inače koristi funkciju kvadratnog gubitka:

$$L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$$

Međutim, u našoj implementaciji greškom smo funkciju gubitka definirali ovako:

$$L(y, h(\mathbf{x})) = (y + 2h(\mathbf{x}))^2$$

S tako pogrešno definiranom funkcijom gubitka postupkom najmanjih kvadrata treniramo naš model na skupu primjera čije su oznake uzorkovane iz distribucije $\mathcal{N}(-1 + 2x, \sigma^2)$, gdje je varijanca σ^2 razmjerno malena (tj. nema mnogo šuma). **Koji vektor težina (w_0, w_1) očekujemo približno dobiti kao rezultat najmanjih kvadrata?**

A $(1, -2)$ B $(2, -4)$ C $(2, -\frac{1}{2})$ D $(\frac{1}{2}, -1)$

3 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

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 31.70 B 191.95 C 165.89 D 230.98

4 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Bez induktivne pristranosti, učenje na temelju podataka ne bi imalo smisla, odnosno algoritam bez induktivne pristranosti ne bi mogao ništa naučiti. **Zašto strojno učenje bez induktivne pristranosti nije moguće?**

- A Oznaka niti jednog neviđenog primjera ne bi bila jednoznačno određena
- B Minimizacija empirijske pogreške učenja ne bi imala rješenje u zatvorenoj formi
- C Primjeri ne bi nužno bili linearno odvojivi
- D Model bi bio presložen te bi imao preveliku pogrešku na skupu za ispitivanje

Linearni klasifikacijski modeli *(3 pitanja)*

5 (T) Algoritam linearne regresije može se pokušati primijeniti na klasifikacijski problem, kao što smo pokušali na predavanjima, međutim to nije dobro funkcioniralo. Razmotrite tri komponente algoritma linearne regresije: model (M), funkciju gubitka (G) i optimizacijski postupak (O). Također, prisjetite se algoritma logističke regresije, koji je dobar klasifikacijski algoritam. Želimo preinačiti komponente algoritma linearne regresije tako da iz njega dobijemo nov algoritam koji klasifikaciju radi bolje od linearnog modela regresije, ali koji je drugačiji od logističke regresije. **Uz koje tri komponente bismo dobili takav algoritam?**

- A M: $h(\mathbf{x}) = \mathbf{w}^\mathrm{T}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1 - h(\mathbf{x})\big)$ O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- B M: $h(\mathbf{x}) = \mathbf{w}^\mathrm{T}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$
- C M: $h(\mathbf{x}) = \sigma(\mathbf{w}^\mathrm{T}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1 - h(\mathbf{x})\big)$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$
- D M: $h(\mathbf{x}) = \sigma(\mathbf{w}^\mathrm{T}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^*$ izračunat 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. Nakon određenog broja iteracija, algoritam perceptrona uspješno se zaustavio s rješenjem. **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 dosežu nulu, ali modelu L2R za to treba više iteracija
- B Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira
- C Pogreške učenja modela NR i modela L2R obje stagniraju nakon određenog broja iteracija, ali modelu NR 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) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

U jednoj iteraciji treniranja modela vektor parametara jednak je

$$\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$$

Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((0.5, -2), 1)$?

A 4.00 B 2.48 C 1.28 D 0.70

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 22 & 0 & -5 \\ 1 & 6 & 28 & 8 & 2 \\ 1 & -6 & -15 & -9 & -6 \\ 1 & -11 & -34 & -16 & -9 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (+2.7838, +0.0225, +0.0856, +0.0315, +0.0135)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0089 B 0.0024 C 0.0013 D 0.0045

9 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (5, 15, 1)$.**

A -2.330 B -3.553 C -4.093 D $+5.454$

10 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina šira od stare?**

A $\frac{1}{2}\sqrt{17}$ puta B $\frac{1}{2}\sqrt{13}$ puta C $\frac{5}{4}\sqrt{2}$ puta D $\frac{1}{2}\sqrt{10}$ puta

11 (T) Važna prednost jezgrenih strojeva je mogućnost primjene jezgrenog trika. Ta je prednost pogotovo očita kada prostor ulaznih primjera $\mathcal{X}$ nije euklidski vektorski prostor, odnosno kada primjere nije moguće prikazati kao vektore realnih brojeva. **Koja je prednost primjene jezgrenog trika u takvom slučaju?**

A Umjesto vektorizacije primjera, dovoljno je defnirati nelinearnu funkciju preslikavanja iz ulaznog prostora u prostor značajki

B Jezgrena funkcija može biti mjera sličnost između nevektoriziranih primjera, što implicitno inducira vektorski prostor značajki

C Jezgrenim trikom implicitno se ostvaruje nelinearnost prostora značajki, što povećava kapacitet modela i povećava njegovu točnost

D Jezgrena funkcija mjeri sličnost između primjera, čime se primjeri efektivno preslikavaju u beskonačnodimenzijski prostor značajki

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$. Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

gdje je Σ_j matrica kovarijacije za klasu j. Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.1, \quad \hat{\boldsymbol{\mu}}_1 = (1, 0, -2), \quad \hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{pmatrix}, \quad \hat{\mu}_2 = 0.9, \quad \hat{\boldsymbol{\mu}}_2 = (2, -1, 5), \quad \hat{\Sigma}_2 = \begin{pmatrix} 5.0 & -0.222 & -0.444 \\ -0.222 & 2.778 & -0.333 \\ -0.444 & -0.333 & 0.444 \end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1, 2$ (zaokružite elemente tako dobivene kovarijacijske matrice na jednu decimalu). Zanima nas klasifikacija modela za primjer $\mathbf{x} = (0, 0, 0)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

A -12.788 B -8.513 C -5.902 D -10.913

13 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (1, 4, 2)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 29$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_2 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 2$ B $\alpha > 8$ C $\alpha > 5$ D $\alpha > 7$

14 (T) Umjesto naivnog Bayesov klasifikatora, ponekad je bolje koristiti polunaivan Bayesov klasifikator? **Po čemu se polunaivan Bayesov klasifikator (SNBC) razlikuje od naivnog Bayesovog klasifikatora (NBC)?**

A SNBC zajedničku vjerojatnost faktorizira u manje faktora, pa ima i manje parametara od NBC

B SNBC pretpostavlja zavisnosti između značajki i klase, ali ima manje parametara od NBC

C SNBC ima više parametara od NBC te modelira zavisnosti između značajki

D SNBC ima više parametara nego NBC, ali manje bridova kada ga se prikaže kao Bayesovu mrežu

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Skriveni Markovljev model (HMM) posebna je vrsta Bayesove mreže. **Na što se odnosi pridjev "skriveni" u nazivu tog modela?**

A Stanja modela poredana su u lanac, a izlazi modela vidljivi su samo za posljednje stanje

B Model opisuje prijelaze između stanja, a trenutačno stanje ovisi samo o prethodnom stanju

C Neke varijable modela nisu opažene u podatcima, ali zavise o opaženim varijablama

D Opažene varijable ovise samo o trenutačnom stanju i prethodno opaženoj varijabli

16 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 21 B 23 C 24 D 18

17 (N) Bayesovom mrežom s četiri varijable modeliramo konstrukte pozitivne psihologije. Koristimo binarne varijable *Ljubav* (L), *Sreća* (S), *Tjeskoba* (T), s vrijednostima 0 (nema) i 1 (ima), te ternarnu varijablu *Novac* (N), s vrijednostima 0 (nema), 1 (ima malo) i 2 (ima puno). Strukturu Bayesove mreže definirali smo tako da ona modelira sljedeće pretpostavljene kauzalne odnose: L uzrokuje S, a N uzrokuje S i T. Tako definiranu Bayesovu mrežu zatim treniramo na sljedećem skupu od $N = 7$ primjera:

L	N	S	T
1	0	1	0
1	0	1	0
0	2	0	1
1	2	1	1
1	1	1	0
0	0	0	0
0	2	1	0

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu), što je istovjetno Laplaceovom zaglađivanju MLE procjene. Na kraju nas, naravno, zanima koja je vjerojatnost života uz ljubav, sreću i malo novaca. Napravite potrebne MAP-procjene parametara. **Koliko iznosi zajednička vjerojatnost $P(L = 1, S = 1, N = 1)$?**

A 0.074 B 0.023 C 0.143 D 0.833

Grupiranje *(3 pitanja)*

18 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika $J^+ - J^*$?**

A 8 B 4 C 12 D 6

19 (N) Particijskim algoritmom grupiranja grupiramo $N = 1000$ primjera. Na temelju znanja o problemu zaključili smo da bi primjeri trebali formirati $K = 3$ grupe, pa smo s tim brojem grupa proveli grupiranje. Kako bismo evaluirali točnost grupiranja, slučajnim odabirom smo iz skupa primjera uzorkovali 10 primjera, ručno smo označili primjere iz tog uzorka, i zatim na tom uzorku računamo Randov indeks. Označavanje smo proveli tako da smo svakom primjeru iz uzorka dodijelili oznaku točne grupe. Oznake grupe dobivene algoritmom grupiranja y_{pred} i oznake točnih grupa y_{true} za svih deset primjera u uzorku su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y_{pred}^{(i)}$	0	1	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.53 B 0.58 C 0.62 D 0.49

20 (T) Algoritam hijerarhijskog aglomerativnog grupiranja (HAC) može koristiti različite vrste povezanosti, uključivo jednostruku (*single*), potpunu (*complete*) i prosječnu (*average*). Označimo funkcije udaljenosti koje odgovaraju ovim povezanostima sa d_s, d_c odnosno d_a. Neka su G_i i G_j dvije neprazne grupe primjera. **Koji odnosi vrijede između vrijednosti ovih funkcija?**

A $d_c(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_s(G_i, G_j)$

B $d_s(G_i, G_j) \leqslant d_c(G_i, G_j) \geqslant d_a(G_i, G_j)$

C $d_s(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_c(G_i, G_j)$

D $d_s(G_i, G_j) \leqslant d_a(G_i, G_j)$, $d_c(G_i, G_j) \leqslant d_a(G_i, G_j)$

Vrednovanje modela *(2 pitanja)*

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c c} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 30 & 18 & 3 \\ 11 & 25 & 2 \\ 4 & 2 & 5 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.155 B 0.085 C 0.185 D 0.205

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:

$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4)$

$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6)$

$h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$

$h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$

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_1 i h_2 [C] h_1 i h_4 [D] h_2 i h_3

Pismeni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

*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 **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

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 230.98 D 191.95

2 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0,1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

A $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

B $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

C $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

D $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

3 (P) Razmatramo model jednostavne regresije, $h(x; w_0, w_1) = w_0 + w_1 x$. Model linearne regresije inače koristi funkciju kvadratnog gubitka:

$$L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$$

Međutim, u našoj implementaciji greškom smo funkciju gubitka definirali ovako:

$$L(y, h(\mathbf{x})) = (y + 2h(\mathbf{x}))^2$$

S tako pogrešno definiranom funkcijom gubitka postupkom najmanjih kvadrata treniramo naš model na skupu primjera čije su oznake uzorkovane iz distribucije $\mathcal{N}(-1 + 2x, \sigma^2)$, gdje je varijanca σ^2 razmjerno malena (tj. nema mnogo šuma). **Koji vektor težina (w_0, w_1) očekujemo približno dobiti kao rezultat najmanjih kvadrata?**

A $(2, -\frac{1}{2})$ B $(2, -4)$ C $(\frac{1}{2}, -1)$ D $(1, -2)$

4 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Bez induktivne pristranosti, učenje na temelju podataka ne bi imalo smisla, odnosno algoritam bez induktivne pristranosti ne bi mogao ništa naučiti. **Zašto strojno učenje bez induktivne pristranosti nije moguće?**

A Model bi bio presložen te bi imao preveliku pogrešku na skupu za ispitivanje

B Oznaka niti jednog neviđenog primjera ne bi bila jednoznačno određena

C Svaka bi se hipoteza prilagodila šumu i ostvarila empirijsku pogrešku nula

D Prostor parametara bio bi neograničen, tj. postojalo bi beskonačno mnogo vektora parametara

Linearni klasifikacijski modeli *(3 pitanja)*

5 (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. Nakon određenog broja iteracija, algoritam perceptrona uspješno se zaustavio s rješenjem. **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 modelu NR za to treba više iteracija
- B Pogreške učenja modela NR i modela L2R dosežu nulu, ali modelu L2R za to treba više iteracija
- C Pogreška učenja modela NR asimptotski teži nuli, dok pogreška učenja modela L2R nakon određenog broja iteracija stagnira
- D Pogreška učenja modela NR nakon određenog broja iteracije doseže nulu, dok pogreška učenja modela L2R najprije pada pa raste

6 (T) Algoritam linearne regresije može se pokušati primijeniti na klasifikacijski problem, kao što smo pokušali na predavanjima, međutim to nije dobro funkcioniralo. Razmotrite tri komponente algoritma linearne regresije: model (M), funkciju gubitka (G) i optimizacijski postupak (O). Također, prisjetite se algoritma logističke regresije, koji je dobar klasfikacijski algoritam. Želimo preinačiti komponente algoritma linearne regresije tako da iz njega dobijemo nov algoritam koji klasifikaciju radi bolje od linearnog modela regresije, ali koji je drugačiji od logističke regresije. **Uz koje tri komponente bismo dobili takav algoritam?**

- A M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1-h(\mathbf{x})\big)$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$
- B M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- C M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1-h(\mathbf{x})\big)$ O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- D M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$

7 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

U jednoj iteraciji treniranja modela vektor parametara jednak je

$$\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$$

Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((0.5, 2), 1)$**?**

A 0.70 B 2.48 C 1.28 D 4.00

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (+0.1370, -0.0290, +0.0194, -0.0461, -0.0388)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0089   B 0.0013   C 0.0045   D 0.0024

9 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A $+1.434$ B -0.676 C -2.330 D -3.553

10 (T) Važna prednost jezgrenih strojeva je mogućnost primjene jezgrenog trika. Ta je prednost pogotovo očita kada prostor ulaznih primjera $\mathcal{X}$ nije euklidski vektorski prostor, odnosno kada primjere nije moguće prikazati kao vektore realnih brojeva. **Koja je prednost primjene jezgrenog trika u takvom slučaju?**

A Jezgrena funkcija može biti mjera sličnost između nevektoriziranih primjera, što implicitno inducira vektorski prostor značajki

B Jezgrenim trikom implicitno se ostvaruje nelinearnost prostora značajki, što povećava kapacitet modela i povećava njegovu točnost

C Jezgrena funkcija mjeri sličnost između primjera, čime se primjeri efektivno preslikavaju u beskonačnodimenzijski prostor značajki

D Umjesto vektorizacije primjera, dovoljno je defnirati nelinearnu funkciju preslikavanja iz ulaznog prostora u prostor značajki

11 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina šira od stare?**

A $\frac{1}{2}\sqrt{13}$ puta   B $\frac{5}{4}\sqrt{2}$ puta   C $\frac{1}{2}\sqrt{10}$ puta   D $\frac{1}{2}\sqrt{17}$ puta

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) Umjesto naivnog Bayesov klasifikatora, ponekad je bolje koristiti polunaivan Bayesov klasifikator? **Po čemu se polunaivan Bayesov klasifikator (SNBC) razlikuje od naivnog Bayesovog klasifikatora (NBC)?**

A SNBC ima više parametara nego NBC, ali manje bridova kada ga se prikaže kao Bayesovu mrežu

B SNBC zajedničku vjerojatnost faktorizira u manje faktora, pa ima i manje parametara od NBC

C SNBC pretpostavlja zavisnosti između značajki i klase, ali ima manje parametara od NBC

D SNBC ima više parametara od NBC te modelira zavisnosti između značajki

13 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$. Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

gdje je Σ_j matrica kovarijacije za klasu j. Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.1,\quad \hat{\boldsymbol{\mu}}_1 = (1,0,-2),\quad \hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{pmatrix},\quad \hat{\mu}_2 = 0.9,\quad \hat{\boldsymbol{\mu}}_2 = (2,-1,5),\quad \hat{\Sigma}_2 = \begin{pmatrix} 5.0 & -0.222 & -0.444 \\ -0.222 & 2.778 & -0.333 \\ -0.444 & -0.333 & 0.444 \end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1,2$. Zanima nas klasifikacija modela za primjer $\mathbf{x} = (1,0,1)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

[A] -12.788 [B] -10.913 [C] -8.513 [D] -5.902

14 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2,1,4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_3 bila najvjerojatnija vrijednost kategoričke varijable $\mathbf{x}$ prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

[A] $\alpha > 5$ [B] $\alpha > 8$ [C] $\alpha > 2$ [D] $\alpha > 7$

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Skriveni Markovljev model (HMM) posebna je vrsta Bayesove mreže. **Na što se odnosi pridjev "skriveni" u nazivu tog modela?**

[A] Opažene varijable ovise samo o trenutačnom stanju i prethodno opaženoj varijabli

[B] Model opisuje prijelaze između stanja, a trenutačno stanje ovisi samo o prethodnom stanju

[C] Stanja modela poredana su u lanac, a izlazi modela vidljivi su samo za posljednje stanje

[D] Neke varijable modela nisu opažene u podatcima, ali zavise o opaženim varijablama

16 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i y, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

[A] 23 [B] 24 [C] 21 [D] 18

17 (N) Bayesovom mrežom s četiri varijable modeliramo konstrukte pozitivne psihologije. Koristimo binarne varijable *Ljubav* (L), *Sreća* (S), *Tjeskoba* (T), s vrijednostima 0 (nema) i 1 (ima), te ternarnu varijablu *Novac* (N), s vrijednostima 0 (nema), 1 (ima malo) i 2 (ima puno). Strukturu Bayesove mreže definirali smo tako da ona modelira sljedeće pretpostavljene kauzalne odnose: L uzrokuje S, a N uzrokuje S i T. Tako definiranu Bayesovu mrežu zatim treniramo na sljedećem skupu od $N = 7$ primjera:

L	N	S	T
1	0	1	0
1	0	1	0
0	2	0	1
1	2	1	1
1	1	1	0
0	0	0	0
0	2	1	0

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu), što je istovjetno Laplaceovom zaglađivanju MLE procjene. Na kraju nas, naravno, zanima koja je vjerojatnost života uz ljubav, sreću i malo novaca. Napravite potrebne MAP-procjene parametara. **Koliko iznosi zajednička vjerojatnost $P(L = 1, S = 1, N = 1)$?**

[A] 0.833 [B] 0.074 [C] 0.143 [D] 0.023

Grupiranje *(3 pitanja)*

18 (N) Particijskim algoritmom grupiranja grupiramo $N = 1000$ primjera. Na temelju znanja o problemu zaključili smo da bi primjeri trebali formirati $K = 3$ grupe, pa smo s tim brojem grupa proveli grupiranje. Kako bismo evaluirali točnost grupiranja, slučajnim odabirom smo iz skupa primjera uzorkovali 10 primjera, ručno smo označili primjere iz tog uzorka, i zatim na tom uzorku računamo Randov indeks. Označavanje smo proveli tako da smo svakom primjeru iz uzorka dodijelili oznaku točne grupe. Oznake grupe dobivene algoritmom grupiranja y_{pred} i oznake točnih grupa y_{true} za svih deset primjera u uzorku su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y^{(i)}_{pred}$	0	1	2	1	1	0	0	2	1	1
$y^{(i)}_{true}$	1	0	0	2	1	0	2	1	0	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.53 B 0.62 C 0.58 D 0.49

19 (T) Algoritam hijerarhijskog aglomerativnog grupiranja (HAC) može koristiti različite vrste povezanosti, uključivo jednostruku (*single*), potpunu (*complete*) i prosječnu (*average*). Označimo funkcije udaljenosti koje odgovaraju ovim povezanostima sa d_s, d_c odnosno d_a. Neka su G_i i G_j dvije neprazne grupe primjera. **Koji odnosi vrijede između vrijednosti ovih funkcija?**

A $d_c(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_s(G_i, G_j)$

B $d_s(G_i, G_j) \leqslant d_a(G_i, G_j)$, $d_c(G_i, G_j) \leqslant d_a(G_i, G_j)$

C $d_s(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_c(G_i, G_j)$

D $d_s(G_i, G_j) \leqslant d_c(G_i, G_j) \geqslant d_a(G_i, G_j)$

20 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika $J^+ - J^*$?**

A 4 B 8 C 6 D 12

Vrednovanje modela *(2 pitanja)*

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 35 & 8 & 3 \\ 10 & 25 & 5 \\ 0 & 12 & 2 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.205 B 0.155 C 0.185 D 0.085

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 $(\mathrm{FPR}, \mathrm{TPR})$ za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

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_2 D h_1 i h_4

Pismeni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

*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 **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

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 191.95 B 31.70 C 165.89 D 230.98

2 (P) Razmatramo model jednostavne regresije, $h(x; w_0, w_1) = w_0 + w_1 x$. Model linearne regresije inače koristi funkciju kvadratnog gubitka:

$$L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$$

Međutim, u našoj implementaciji greškom smo funkciju gubitka definirali ovako:

$$L(y, h(\mathbf{x})) = (y + 2h(\mathbf{x}))^2$$

S tako pogrešno definiranom funkcijom gubitka postupkom najmanjih kvadrata treniramo naš model na skupu primjera čije su oznake uzorkovane iz distribucije $\mathcal{N}(-1 + 2x, \sigma^2)$, gdje je varijanca σ^2 razmjerno malena (tj. nema mnogo šuma). **Koji vektor težina (w_0, w_1) očekujemo približno dobiti kao rezultat najmanjih kvadrata?**

A $(1, -2)$ B $(2, -\frac{1}{2})$ C $(2, -4)$ D $(\frac{1}{2}, -1)$

3 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Bez induktivne pristranosti, učenje na temelju podataka ne bi imalo smisla, odnosno algoritam bez induktivne pristranosti ne bi mogao ništa naučiti. **Zašto strojno učenje bez induktivne pristranosti nije moguće?**

A Prostor parametara bio bi neograničen, tj. postojalo bi beskonačno mnogo vektora parametara

B Svako označavanje neviđenih primjera bilo bi jednako vjerojatno

C Svaka bi se hipoteza prilagodila šumu i ostvarila empirijsku pogrešku nula

D Model bi bio prejednostavan te ne bi postojala hipoteza s empirijskom pogreškom nula

4 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

A $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

B $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

C $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

D $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

Linearni klasifikacijski modeli *(3 pitanja)*

5 (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. Nakon određenog broja iteracija, algoritam perceptrona uspješno se zaustavio s rješenjem. **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 modelu NR za to treba više iteracija
- 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

6 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

U jednoj iteraciji treniranja modela vektor parametara jednak je

$$\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$$

Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((-0.5, 2), 1)$?

A 0.70 B 2.48 C 4.00 D 1.28

7 (T) Algoritam linearne regresije može se pokušati primijeniti na klasifikacijski problem, kao što smo pokušali na predavanjima, međutim to nije dobro funkcioniralo. Razmotrite tri komponente algoritma linearne regresije: model (M), funkciju gubitka (G) i optimizacijski postupak (O). Također, prisjetite se algoritma logističke regresije, koji je dobar klasfikacijski algoritam. Želimo preinačiti komponente algoritma linearne regresije tako da iz njega dobijemo nov algoritam koji klasifikaciju radi bolje od linearnog modela regresije, ali koji je drugačiji od logističke regresije. **Uz koje tri komponente bismo dobili takav algoritam?**

- A M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1 - h(\mathbf{x})\big)$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$
- B M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- C M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1 - h(\mathbf{x})\big)$ O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- D M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$

Jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Važna prednost jezgrenih strojeva je mogućnost primjene jezgrenog trika. Ta je prednost pogotovo očita kada prostor ulaznih primjera $\mathcal{X}$ nije euklidski vektorski prostor, odnosno kada primjere nije moguće prikazati kao vektore realnih brojeva. **Koja je prednost primjene jezgrenog trika u takvom slučaju?**

- A Jezgrena funkcija može biti mjera sličnost između nevektoriziranih primjera, što implicitno inducira vektorski prostor značajki
- B Umjesto vektorizacije primjera, dovoljno je defnirati nelinearnu funkciju preslikavanja iz ulaznog prostora u prostor značajki
- C Jezgrenim trikom implicitno se ostvaruje nelinearnost prostora značajki, što povećava kapacitet modela i povećava njegovu točnost
- D Jezgrena funkcija mjeri sličnost između primjera, čime se primjeri efektivno preslikavaju u beskonačnodimenzijski prostor značajki

9 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (-1.3214, +0.0714, +0.1071, 0, -0.0357)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0024 B 0.0089 C 0.0013 D 0.0045

10 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (-3, -5, 3)$.**

A -3.553 B -4.093 C $+1.434$ D -2.330

11 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina šira od stare?**

A $\frac{5}{4}\sqrt{2}$ puta B $\frac{1}{2}\sqrt{13}$ puta C $\frac{1}{2}\sqrt{10}$ puta D $\frac{1}{2}\sqrt{17}$ puta

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) Umjesto naivnog Bayesov klasifikatora, ponekad je bolje koristiti polunaivan Bayesov klasifikator? **Po čemu se polunaivan Bayesov klasifikator (SNBC) razlikuje od naivnog Bayesovog klasifikatora (NBC)?**

A SNBC zajedničku vjerojatnost faktorizira u manje faktora, pa ima i manje parametara od NBC

B SNBC pretpostavlja zavisnosti između značajki i klase, ali ima manje parametara od NBC

C SNBC ima više parametara nego NBC, ali manje bridova kada ga se prikaže kao Bayesovu mrežu

D SNBC ima više parametara od NBC te modelira zavisnosti između značajki

13 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2, 1, 4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_3 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 2$ B $\alpha > 8$ C $\alpha > 5$ D $\alpha > 7$

14 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$. Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

gdje je Σ_j matrica kovarijacije za klasu j. Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.1, \quad \hat{\boldsymbol{\mu}}_1 = (1, 0, -2), \quad \hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{pmatrix}, \quad \hat{\mu}_2 = 0.9, \quad \hat{\boldsymbol{\mu}}_2 = (2, -1, 5), \quad \hat{\Sigma}_2 = \begin{pmatrix} 5.0 & -0.222 & -0.444 \\ -0.222 & 2.778 & -0.333 \\ -0.444 & -0.333 & 0.444 \end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1, 2$. Zanima nas klasifikacija modela za primjer $\mathbf{x} = (1, 0, 1)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

A -5.902 B -10.913 C -12.788 D -8.513

Probabilistički grafički modeli *(3 pitanja)*

15 (N) Bayesovom mrežom s četiri varijable modeliramo konstrukte pozitivne psihologije. Koristimo binarne varijable *Ljubav* (L), *Sreća* (S), *Tjeskoba* (T), s vrijednostima 0 (nema) i 1 (ima), te ternarnu varijablu *Novac* (N), s vrijednostima 0 (nema), 1 (ima malo) i 2 (ima puno). Strukturu Bayesove mreže definirali smo tako da ona modelira sljedeće pretpostavljene kauzalne odnose: L uzrokuje S, a N uzrokuje S i T. Tako definiranu Bayesovu mrežu zatim treniramo na sljedećem skupu od $N = 7$ primjera:

L	N	S	T
1	0	1	0
1	0	1	0
0	2	0	1
1	2	1	1
1	1	1	0
0	0	0	0
0	2	1	0

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu), što je istovjetno Laplaceovom zaglađivanju MLE procjene. Na kraju nas, naravno, zanima koja je vjerojatnost života uz ljubav, sreću i malo novaca. Napravite potrebne MAP-procjene parametara. **Koliko iznosi zajednička vjerojatnost $P(L = 1, S = 1, N = 1)$?**

A 0.833 B 0.074 C 0.143 D 0.023

16 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 23 B 21 C 18 D 24

17 (T) Skriveni Markovljev model (HMM) posebna je vrsta Bayesove mreže. **Na što se odnosi pridjev "skriveni" u nazivu tog modela?**

- A Model opisuje prijelaze između stanja, a trenutačno stanje ovisi samo o prethodnom stanju
- B Neke varijable modela nisu opažene u podatcima, ali zavise o opaženim varijablama
- C Stanja modela poredana su u lanac, a izlazi modela vidljivi su samo za posljednje stanje
- D Opažene varijable ovise samo o trenutačnom stanju i prethodno opaženoj varijabli

Grupiranje *(3 pitanja)*

18 (T) Algoritam hijerarhijskog aglomerativnog grupiranja (HAC) može koristiti različite vrste povezanosti, uključivo jednostruku (*single*), potpunu (*complete*) i prosječnu (*average*). Označimo funkcije udaljenosti koje odgovaraju ovim povezanostima sa d_s, d_c odnosno d_a. Neka su G_i i G_j dvije neprazne grupe primjera. **Koji odnosi vrijede između vrijednosti ovih funkcija?**

- A $d_s(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_c(G_i, G_j)$
- B $d_s(G_i, G_j) \leqslant d_c(G_i, G_j) \geqslant d_a(G_i, G_j)$
- C $d_c(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_s(G_i, G_j)$
- D $d_s(G_i, G_j) \leqslant d_a(G_i, G_j)$, $d_c(G_i, G_j) \leqslant d_a(G_i, G_j)$

19 (N) Particijskim algoritmom grupiranja grupiramo $N = 1000$ primjera. Na temelju znanja o problemu zaključili smo da bi primjeri trebali formirati $K = 3$ grupe, pa smo s tim brojem grupa proveli grupiranje. Kako bismo evaluirali točnost grupiranja, slučajnim odabirom smo iz skupa primjera uzorkovali 10 primjera, ručno smo označili primjere iz tog uzorka, i zatim na tom uzorku računamo Randov indeks. Označavanje smo proveli tako da smo svakom primjeru iz uzorka dodijelili oznaku točne grupe. Oznake grupe dobivene algoritmom grupiranja y_{pred} i oznake točnih grupa y_{true} za svih deset primjera u uzorku su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y_{pred}^{(i)}$	0	0	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	1	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.62 B 0.58 C 0.49 D 0.53

20 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0), (0,4), (2,0), (2,4), (4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

A 8 B 12 C 4 D 6

Vrednovanje modela *(2 pitanja)*

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 20 & 18 & 3 \\ 21 & 25 & 2 \\ 4 & 2 & 5 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.185 B 0.205 C 0.155 D 0.085

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 $(\mathrm{FPR}, \mathrm{TPR})$ za sva četiri klasifikatora su sljedeći:

$$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4) \qquad h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$$
$$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6) \qquad h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$$

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_1 i h_4 C h_3 i h_4 D h_1 i h_2

Pismeni ispit iz Strojnog učenja 1 (ak. god. 2021./2022.)

*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 **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Bez induktivne pristranosti, učenje na temelju podataka ne bi imalo smisla, odnosno algoritam bez induktivne pristranosti ne bi mogao ništa naučiti. **Zašto strojno učenje bez induktivne pristranosti nije moguće?**

A Model bi bio prejednostavan te ne bi postojala hipoteza s empirijskom pogreškom nula

B Prostor parametara bio bi neograničen, tj. postojalo bi beskonačno mnogo vektora parametara

C Model bi bio presložen te bi imao preveliku pogrešku na skupu za ispitivanje

D Svako označavanje neviđenih primjera bilo bi jednako vjerojatno

2 (N) Raspolažemo sljedećim skupom primjera u dvodimenzijskome ulaznom prostoru:

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

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 165.89 B 191.95 C 31.70 D 230.98

3 (P) Razmatramo model jednostavne regresije, $h(x; w_0, w_1) = w_0 + w_1 x$. Model linearne regresije inače koristi funkciju kvadratnog gubitka:

$$L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$$

Međutim, u našoj implementaciji greškom smo funkciju gubitka definirali ovako:

$$L(y, h(\mathbf{x})) = (y + 2h(\mathbf{x}))^2$$

S tako pogrešno definiranom funkcijom gubitka postupkom najmanjih kvadrata treniramo naš model na skupu primjera čije su oznake uzorkovane iz distribucije $\mathcal{N}(-1 + 2x, \sigma^2)$, gdje je varijanca σ^2 razmjerno malena (tj. nema mnogo šuma). **Koji vektor težina (w_0, w_1) očekujemo približno dobiti kao rezultat najmanjih kvadrata?**

A $(2, -4)$ B $(\frac{1}{2}, -1)$ C $(2, -\frac{1}{2})$ D $(1, -2)$

4 (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije neželjenih elektroničkih poruka (*spam filtering*). Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava da je poruka neželjena (spam), a $y = 0$ da ona to nije. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

A $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) = 0$

B $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

C $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

D $L(0,0) = L(1,1) < L(0,1) < L(1,0)$

Linearni klasifikacijski modeli *(3 pitanja)*

5 (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. Nakon određenog broja iteracija, algoritam perceptrona uspješno se zaustavio s rješenjem. **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 modelu NR za to treba više iteracija
- 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

6 (T) Algoritam linearne regresije može se pokušati primijeniti na klasifikacijski problem, kao što smo pokušali na predavanjima, međutim to nije dobro funkcioniralo. Razmotrite tri komponente algoritma linearne regresije: model (M), funkciju gubitka (G) i optimizacijski postupak (O). Također, prisjetite se algoritma logističke regresije, koji je dobar klasfikacijski algoritam. Želimo preinačiti komponente algoritma linearne regresije tako da iz njega dobijemo nov algoritam koji klasifikaciju radi bolje od linearnog modela regresije, ali koji je drugačiji od logističke regresije. **Uz koje tri komponente bismo dobili takav algoritam?**

- A M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1-h(\mathbf{x})\big)$ O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- B M: $h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$
- C M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = (y - h(\mathbf{x}))^2$, O: $\mathbf{w}^*$ izračunat gradijentnim spustom
- D M: $h(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$; G: $L(y, h(\mathbf{x})) = -y \ln h(\mathbf{x}) - (1-y)\ln\big(1-h(\mathbf{x})\big)$, O: $\mathbf{w}^* = \mathbf{X}^+\mathbf{y}$

7 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom

$$\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$$

U jednoj iteraciji treniranja modela vektor parametara jednak je

$$\mathbf{w} = (0.2, 0.5, -1.1, 2.7)$$

Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((-0.5, 2), 1)$?

A 2.48 B 1.28 C 0.70 D 4.00

Jezgrene i neparametarske metode *(4 pitanja)*

8 (P) Raspolažemo sljedećim skupom označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a s tvrdom marginom. Međutim, naknadno smo utvrdili da je primjer $(2, 0)$ imao pogrešnu oznaku, pa smo to ispravili te ponovno trenirali SVM. Na ispravljenom skupu primjera dobili smo granicu između klasa sa znatno širom marginom nego na početnom skupu primjera. **Koliko je nova margina šira od stare?**

A $\frac{1}{2}\sqrt{13}$ puta B $\frac{5}{4}\sqrt{2}$ puta C $\frac{1}{2}\sqrt{10}$ puta D $\frac{1}{2}\sqrt{17}$ puta

9 (T) Važna prednost jezgrenih strojeva je mogućnost primjene jezgrenog trika. Ta je prednost pogotovo očita kada prostor ulaznih primjera $\mathcal{X}$ nije euklidski vektorski prostor, odnosno kada primjere nije moguće prikazati kao vektore realnih brojeva. **Koja je prednost primjene jezgrenog trika u takvom slučaju?**

- A Jezgrenim trikom implicitno se ostvaruje nelinearnost prostora značajki, što povećava kapacitet modela i povećava njegovu točnost
- B Umjesto vektorizacije primjera, dovoljno je defnirati nelinearnu funkciju preslikavanja iz ulaznog prostora u prostor značajki
- C Jezgrena funkcija može biti mjera sličnost između nevektoriziranih primjera, što implicitno inducira vektorski prostor značajki
- D Jezgrena funkcija mjeri sličnost između primjera, čime se primjeri efektivno preslikavaju u beskonačnodimenzijski prostor značajki

10 (N) Rješavamo binarni klasifikacijski problem. Raspolažemo označenim skupom primjera. Odgovarajuća matrica dizajna je sljedeća:

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 16 & -8 & -11 \\ 1 & -5 & 4 & -8 & -7 \\ 1 & 7 & -4 & 11 & 9 \\ 1 & 15 & -20 & 25 & 25 \end{pmatrix}$$

Na ovom skupu treniramo model SVM-a s tvrdom marginom i linearnom jezgrenom funkcijom (tj. bez preslikavanja u prostor značajki). Model treniramo u primarnoj formulaciji. Za rješenje maksimalne margine dobili smo ovaj vektor težina (uključivo s težinom w_0):

$$\mathbf{w} = (-1.3214, +0.0714, +0.1071, 0, -0.0357)$$

Umjesto u primarnoj formulaciji, model smo mogli trenirati u dualnoj formulaciji, pa bismo umjesto vektora težina $\mathbf{w}$ dobili vektor dualnih parametara $\boldsymbol{\alpha}$, odnosno Lagrangeove multiplikatore. Prisjetite se da su vektori čiji su Lagrangeovi multiplikatori veći od nule potporni vektori. Premda to nije uvijek moguće, u ovom konkretnom slučaju dualni parametri modela mogu se izvesti iz rješenja primarnog modela. Izvedite vektor dualnih parametara $\boldsymbol{\alpha}$. **Koliko iznosi najveća vrijednost parametra u vektoru dualnih parametara $\boldsymbol{\alpha}$?** (Rezultate uspoređujte po prve tri decimale.)

A 0.0089 B 0.0013 C 0.0024 D 0.0045

11 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 1)^3$. Potporni vektori i njihove oznake su sljedeći:

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((11, 32, 21), -1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((5, 11, 6), -1)$$
$$(\mathbf{x}^{(3)}, y^{(3)}) = ((6, 12, 6), +1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 1.063 \cdot 10^{-7}$, $\alpha_2 = 1.022 \cdot 10^{-5}$ i $\alpha_3 = 1.032 \cdot 10^{-5}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A $+1.434$ B -0.676 C -2.330 D -3.553

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (N) Na skupu označenih primjera u ulaznome prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$. Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

gdje je Σ_j matrica kovarijacije za klasu j. Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.1, \quad \hat{\boldsymbol{\mu}}_1 = (1, 0, -2), \quad \hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{pmatrix}, \quad \hat{\mu}_2 = 0.9, \quad \hat{\boldsymbol{\mu}}_2 = (2, -1, 5), \quad \hat{\Sigma}_2 = \begin{pmatrix} 5.0 & -0.222 & -0.444 \\ -0.222 & 2.778 & -0.333 \\ -0.444 & -0.333 & 0.444 \end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1, 2$ (zaokružite elemente tako dobivene kovarijacijske matrice na jednu decimalu). Zanima nas klasifikacija modela za primjer $\mathbf{x} = (0, 0, 0)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

A -10.913 B -5.902 C -12.788 D -8.513

13 (T) Umjesto naivnog Bayesov klasifikatora, ponekad je bolje koristiti polunaivan Bayesov klasifikator? **Po čemu se polunaivan Bayesov klasifikator (SNBC) razlikuje od naivnog Bayesovog klasifikatora (NBC)?**

A SNBC ima više parametara nego NBC, ali manje bridova kada ga se prikaže kao Bayesovu mrežu

B SNBC zajedničku vjerojatnost faktorizira u manje faktora, pa ima i manje parametara od NBC

C SNBC pretpostavlja zavisnosti između značajki i klase, ali ima manje parametara od NBC

D SNBC ima više parametara od NBC te modelira zavisnosti između značajki

14 (P) MAP-procjeniteljem na skupu $\mathcal{D}$ procjenjujemo parametre $\boldsymbol{\mu}$ kategoričke (multinulijeve) varijable $\mathbf{x}$, tj. procjenjujemo parametre multinulijeve distribucije $P(\mathbf{x}|\boldsymbol{\mu})$. Varijabla $\mathbf{x}$ može proprimiti tri moguće vrijednosti ($K = 3$). Apriorna Dirichletova distribucija $P(\boldsymbol{\mu}|\boldsymbol{\alpha})$ definirana je parametrima $\boldsymbol{\alpha}$ definiranima kao $\boldsymbol{\alpha} = \alpha \cdot (2, 1, 4)$, gdje faktor α određuje vršnost Dirichletove distribucije, $\alpha \geqslant 1$. U skupu podataka $\mathcal{D}$ realizirane su sve tri vrijednosti kategoričke varijable $\mathbf{x}$, i to sa sljedećim brojem realizacija: $N_1 = 50$, $N_2 = 64$, $N_3 = 40$. **Koliko mora iznositi faktor α, a da bi vrijednost x_3 bila najvjerojatnija vrijednost kategoričke varijable x prema procijenjenoj distribuciji $P(\mathbf{x}|\boldsymbol{\mu})$?**

A $\alpha > 7$ B $\alpha > 5$ C $\alpha > 8$ D $\alpha > 2$

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Skriveni Markovljev model (HMM) posebna je vrsta Bayesove mreže. **Na što se odnosi pridjev "skriveni" u nazivu tog modela?**

A Opažene varijable ovise samo o trenutačnom stanju i prethodno opaženoj varijabli

B Model opisuje prijelaze između stanja, a trenutačno stanje ovisi samo o prethodnom stanju

C Neke varijable modela nisu opažene u podatcima, ali zavise o opaženim varijablama

D Stanja modela poredana su u lanac, a izlazi modela vidljivi su samo za posljednje stanje

16 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Sve varijable su binarne, osim varijabli v i x, koje su ternarne. Uz navedeni topološki uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

A 21 B 18 C 24 D 23

17 (N) Bayesovom mrežom s četiri varijable modeliramo konstrukte pozitivne psihologije. Koristimo binarne varijable *Ljubav* (L), *Sreća* (S), *Tjeskoba* (T), s vrijednostima 0 (nema) i 1 (ima), te ternarnu varijablu *Novac* (N), s vrijednostima 0 (nema), 1 (ima malo) i 2 (ima puno). Strukturu Bayesove mreže definirali smo tako da ona modelira sljedeće pretpostavljene kauzalne odnose: L uzrokuje S, a N uzrokuje S i T. Tako definiranu Bayesovu mrežu zatim treniramo na sljedećem skupu od $N = 7$ primjera:

L	N	S	T
1	0	1	0
1	0	1	0
0	2	0	1
1	2	1	1
1	1	1	0
0	0	0	0
0	2	1	0

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu), što je istovjetno Laplaceovom zaglađivanju MLE procjene. Na kraju nas, naravno, zanima koja je vjerojatnost života uz ljubav, sreću i malo novaca. Napravite potrebne MAP-procjene parametara. **Koliko iznosi zajednička vjerojatnost** $P(L = 1, S = 1, N = 1)$**?**

A 0.074 B 0.143 C 0.023 D 0.833

Grupiranje *(3 pitanja)*

18 (P) Skup neoznačenih primjera u dvodimenzijskome ulaznom prostoru neka je sljedeći:

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{5} = \{(0,0),(0,4),(2,0),(2,4),(4,2)\}$$

Primjere grupiramo algoritmom K-sredina sa $K = 3$ grupe. Za početna središta grupa odaberemo nasumično primjere iz $\mathcal{D}$, pri čemu, naravno, pazimo da odaberemo različita središta. Ishod grupiranja i konačan iznos kriterijske funkcije J ovisit će o odabiru početnih središta. Neka je J^* vrijednost kriterijske funkcije u točki globalnog minimuma, dakle vrijednost koja odgovara najboljem grupiranju. Neka je J^+ vrijednost kriterijske funkcije u točki lokalnog minimuma, i to onoj točki lokalnog minimuma s najvećom vrijednošću funkcije J. **Koliko iznosi razlika** $J^+ - J^*$?

A 8 B 4 C 12 D 6

19 (T) Algoritam hijerarhijskog aglomerativnog grupiranja (HAC) može koristiti različite vrste povezanosti, uključivo jednostruku (*single*), potpunu (*complete*) i prosječnu (*average*). Označimo funkcije udaljenosti koje odgovaraju ovim povezanostima sa d_s, d_c odnosno d_a. Neka su G_i i G_j dvije neprazne grupe primjera. **Koji odnosi vrijede između vrijednosti ovih funkcija?**

A $d_s(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_c(G_i, G_j)$

B $d_s(G_i, G_j) \leqslant d_a(G_i, G_j)$, $d_c(G_i, G_j) \leqslant d_a(G_i, G_j)$

C $d_s(G_i, G_j) \leqslant d_c(G_i, G_j) \geqslant d_a(G_i, G_j)$

D $d_c(G_i, G_j) \leqslant d_a(G_i, G_j) \leqslant d_s(G_i, G_j)$

20 (N) Particijskim algoritmom grupiranja grupiramo $N = 1000$ primjera. Na temelju znanja o problemu zaključili smo da bi primjeri trebali formirati $K = 3$ grupe, pa smo s tim brojem grupa proveli grupiranje. Kako bismo evaluirali točnost grupiranja, slučajnim odabirom smo iz skupa primjera uzorkovali 10 primjera, ručno smo označili primjere iz tog uzorka, i zatim na tom uzorku računamo Randov indeks. Označavanje smo proveli tako da smo svakom primjeru iz uzorka dodijelili oznaku točne grupe. Oznake grupe dobivene algoritmom grupiranja y_{pred} i oznake točnih grupa y_{true} za svih deset primjera u uzorku su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y_{pred}^{(i)}$	0	0	2	1	1	0	0	2	1	1
$y_{true}^{(i)}$	1	0	0	2	1	0	2	1	1	1

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.58 B 0.62 C 0.49 D 0.53

Vrednovanje modela *(2 pitanja)*

21 (N) Na ispitnome skupu evaluiramo model multinomijalne logističke regresije (MLR) za klasifikaciju u $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 20 & 18 & 3 \\ 21 & 25 & 2 \\ 4 & 2 & 5 \end{pmatrix} \end{array}$$

Klasifikator MLR uspoređujemo s klasifikatorom RAND koji primjere klasificira nasumično, i to tako da oznaku $y = j$ dodjeljuje s vjerojatnošću proporcionalnoj udjelu klase j u ispitnome skupu. Izračunajte mikro-F_1 za klasifikator MLR i očekivani mikro-F_1 za klasifikator RAND. **Koliko iznosi očekivana razlika u vrijednostima mikro-F_1 klasifikatora MLR i RAND?**

A 0.185 B 0.085 C 0.205 D 0.155

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:

$h_1 : (0.4, 0.2), (0.6, 0.2), (0.9, 0.4)$

$h_2 : (0.3, 0.1), (0.5, 0.4), (0.8, 0.6)$

$h_3 : (0.1, 0.5), (0.6, 0.6), (0.7, 0.8)$

$h_4 : (0.3, 0.8), (0.4, 0.9), (0.6, 1.0)$

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_3 i h_4 [C] h_1 i h_3 [D] h_2 i h_3

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22
Grupa A	B	D	B	A	C	C	C	B	B	B	D	C	C	A	A	A	B	B	B	A	A	C
Grupa B	B	C	D	B	C	A	D	D	B	D	A	D	C	B	B	C	D	C	D	C	C	C
Grupa C	C	D	A	A	D	B	C	D	D	D	B	B	D	C	C	C	A	A	D	C	C	C
Grupa D	B	A	C	B	C	B	A	D	D	A	C	D	B	B	D	C	B	D	C	B	A	D
Grupa E	D	D	B	A	B	B	B	A	B	B	A	D	B	B	B	D	B	A	B	A	D	B
Grupa F	D	A	B	A	B	C	A	B	C	A	D	D	D	C	C	C	A	A	A	A	B	A