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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \wedge h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

[A] $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ [B] $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$ [C] $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ [D] $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (0.28, -0.58, 1.79, -0.75)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\boldsymbol{\Phi},\mathbf{y})$?

[A] -5.69 [B] -4.73 [C] -12.63 [D] -10.64

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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0,1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

[A] $L(1,0) > L(0,1)$ i $L(1,1) = L(0,0) = 0$
[C] $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) > 0$
[B] $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$
[D] $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

4 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

[A] Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina

[B] Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0

[C] Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0

[D] Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0

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 $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.50 B 0.22 C 1.42 D 0.09

6 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

A Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$

B Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$

C Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

D Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika

7 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 1:3 B OVO:OVR $\approx$ 4:5 C OVO:OVR $\approx$ 32:50 D OVO:OVR $\approx$ 1:405

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

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

9 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

B Što je vrijednost γ veća, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

C Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

D Što je vrijednost γ veća, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

10 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{3}{5}\sqrt{10}$ puta B $\frac{1}{6}\sqrt{2}$ puta C $\frac{2}{5}\sqrt{10}$ puta D $\frac{4}{5}\sqrt{10}$ puta

11 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x}$ = “fire” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 0$ B $y = 0$ i $y = 1$ C $y = 1$ i $y = 1$ D $y = 0$ i $y = 2$

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

A $(1, 3), (5, 4), (7, 9)$ B $(1, 6), (3, 3), (5, 0)$ C $(3, 2), (4, 5), (9, 7)$ D $(1, 2), (3, 4), (6, 7)$

13 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

A Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani

B Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

C Funkcija koja uzoku $\mathcal{D}$ pridjeluje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

D Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

14 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

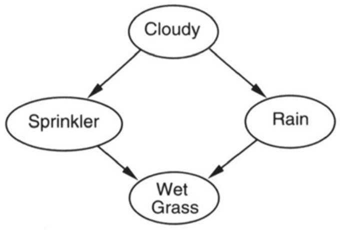
i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{dipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.588 B 0.488 C 0.741 D 0.322

Probabilistički grafički modeli *(3 pitanja)*

15 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S=1\|C)$
0	0.5
1	0.3

C	$P(R=1\|C)$
0	0.2
1	0.8

S	R	$P(W=1\|S,R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

A 0.309 B 0.069 C 0.223 D 0.144

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

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

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

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

17 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest "efekt objašnjavanja", koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

A U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y zavisni, ali postaju nezavisni ako je opažena posljedica z

B U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y zavisni, ali postaju nezavisni ako je opažen posredni uzrok z

C U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y nezavisni, ali postaju zavisni ako je opažen posredni uzrok z

D U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z

Grupiranje *(3 pitanja)*

18 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$\begin{array}{c} \\ \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array}
\begin{array}{c}
\begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\
\begin{pmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.5 \\ & 1 & 0.9 & 0.4 & 0.6 \\ & & 1 & 0.8 & 0.1 \\ & & & 1 & 0.6 \\ & & & & 1 \end{pmatrix}
\end{array}$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

A $((((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(4)}),\mathbf{x}^{(1)}),\mathbf{x}^{(5)})$ C $(((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(4)}),(\mathbf{x}^{(1)},\mathbf{x}^{(5)}))$

B $((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(1)}),(\mathbf{x}^{(4)},\mathbf{x}^{(5)}))$ D $((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(4)}),(\mathbf{x}^{(5)},\mathbf{x}^{(1)}))$

19 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina

$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3$, $LL^*_1 \geqslant LL^*_2$ C $LL^0_2 \leqslant LL^0_4$, $LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

B $LL^0_3 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$ D $LL^0_2 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$

20 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

A Oba algoritma imaju vremensku složenost veću od algoritma k-sredina, ali mogu raditi s općenitom mjerom sličnosti između primjera

B Oba algoritma imaju složenost linearnu u broju primjera, no algoritam k-medoida provodi čvrsto a algoritam HAC hijerarhijsko grupiranje

C Oba algoritma imaju vremenska složenost veću od linearne u broju primjera, no kod k-medoida primjer može pripada u više grupa istovremeno

D Oba algoritma iziskuju odabir početnih medoida grupa, no kod algoritma HAC taj se odabir može napraviti jednostrukim ili potpunim povezivanjem

Vrednovanje modela *(2 pitanja)*

21 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\big\{(y^{(i)}, h(\mathbf{x}^{(i)})\big\}_{i=1}^{10} = \big\{(1, 0.2), (0, 0.2), (0, 0.6), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.6), (1, 0.2), (0, 0.6), (1, 0.8)\big\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0 B 0.16 C 0.24 D 0.35

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \dots, 2^5\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 4096 B 2820 C 5585 D 2980

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

A Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0

B Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0

C Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0

D Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-1.91, 0.22, -0.47, -0.06)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\boldsymbol{\Phi},\mathbf{y})$?

A -4.73 B -12.63 C -5.69 D -10.64

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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0,1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

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

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

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

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

4 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \wedge h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

A $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ B $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ C $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$ D $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$

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 $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.50 B 1.42 C 0.09 D 0.22

6 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

A Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika

B Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$

C Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$

D Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

7 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 32:50 B OVO:OVR $\approx$ 1:405 C OVO:OVR $\approx$ 1:3 D OVO:OVR $\approx$ 4:5

Jezgrene i neparametarske metode *(4 pitanja)*

8 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{3}{5}\sqrt{10}$ puta B $\frac{2}{5}\sqrt{10}$ puta C $\frac{1}{6}\sqrt{2}$ puta D $\frac{4}{5}\sqrt{10}$ puta

9 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

B Što je vrijednost γ veća, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

C Što je vrijednost γ manja, to veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

D Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

10 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{"water"}, 0), (\text{"voda"}, 1), (\text{"zrak"}, 1), (\text{"luft"}, 2), (\text{"feuer"}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{"water"}, \text{"voda"}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} =$ "love" pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 2$ i $y = 1$ B $y = 0$ i $y = 2$ C $y = 1$ i $y = 1$ D $y = 0$ i $y = 0$

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

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

Lagrangeovi koeficijenti su $\alpha_1 = 3.496 \cdot 10^{-6}$, $\alpha_2 = 4.136 \cdot 10^{-7}$ i $\alpha_3 = 3.910 \cdot 10^{-6}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

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

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

A $(1, 3), (5, 4), (7, 9)$ B $(3, 2), (4, 5), (9, 7)$ C $(1, 2), (3, 4), (6, 7)$ D $(1, 6), (3, 3), (5, 0)$

13 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

A Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani

B Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

C Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

D Funkcija koja uzoku $\mathcal{D}$ pridjeluje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

14 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

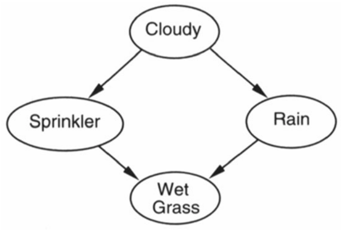
i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.322 B 0.488 C 0.588 D 0.741

Probabilistički grafički modeli *(3 pitanja)*

15 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S=1\|C)$
0	0.5
1	0.2

C	$P(R=1\|C)$
0	0.2
1	0.8

S	R	$P(W=1\|S,R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

[A] 0.223 [B] 0.309 [C] 0.144 [D] 0.069

16 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest "efekt objašnjavanja", koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

[A] U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y zavisni, ali postaju nezavisni ako je opažen posredni uzrok z

[B] U strukturi $x \leftarrow z \to y$, gdje su posljedice x i y nezavisni, ali postaju zavisni ako je opažen uzrok z

[C] U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z

[D] U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y zavisni, ali postaju nezavisni ako je opažena posljedica z

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

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

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

[A] $w\perp y|z$ [B] $x\perp z|y$ [C] $v\perp z|x$ [D] $v\perp y|x$

Grupiranje *(3 pitanja)*

18 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$\begin{array}{c} \\ \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\ \begin{pmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.5 \\ & 1 & 0.9 & 0.4 & 0.6 \\ & & 1 & 0.8 & 0.1 \\ & & & 1 & 0.6 \\ & & & & 1 \end{pmatrix} \end{array}$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

[A] $((((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(4)}),\mathbf{x}^{(1)}),\mathbf{x}^{(5)})$ [C] $((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(1)}),(\mathbf{x}^{(4)},\mathbf{x}^{(5)}))$

[B] $((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(4)}),(\mathbf{x}^{(5)},\mathbf{x}^{(1)}))$ [D] $(((\mathbf{x}^{(2)},\mathbf{x}^{(3)}),\mathbf{x}^{(4)}),(\mathbf{x}^{(1)},\mathbf{x}^{(5)}))$

19 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

A Oba algoritma mogu grupirati primjere koji nisu vektorizirani, no algoritam HAC daje hijerarhiju dok algoritam k-medoida daje particiju grupa

B Oba algoritma imaju vremenska složenost veću od linearne u broju primjera, no kod k-medoida primjer može pripada u više grupa istovremeno

C Oba algoritma nužno konvergiraju, no algoritam k-medoida konvergira u globalni optimum dok HAC može konvergirati u lokalni optimum

D Oba algoritma mogu grupirati na temelju mjere udaljenosti koja nije euklidska, no algoritam k-medoida ima veću vremensku složenost od algoritma HAC

20 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina

$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_3 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$ C $LL^0_2 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$

B $LL^0_2 \leqslant LL^0_4,\ LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$ D $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3,\ LL^*_1 \geqslant LL^*_2$

Vrednovanje modela *(2 pitanja)*

21 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\{(y^{(i)}, h(\mathbf{x}^{(i)})\}_{i=1}^{10} = \{(1, 0.2), (0, 0.2), (0, 0.6), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.6), (1, 0.2), (0, 0.6), (1, 0.8)\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0.35 B 0.16 C 0.24 D 0

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \ldots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 5636 B 6820 C 5956 D 8855

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (2.2, -0.66, 0.77, -0.24)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\boldsymbol{\Phi},\mathbf{y})$?

[A] -4.73 [B] -5.69 [C] -12.63 [D] -10.64

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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

[A] $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$
[C] $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$
[B] $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) > 0$
[D] $L(1,0) > L(0,1)$ i $L(1,1) = L(0,0) = 0$

3 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \wedge h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

[A] $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ [B] $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ [C] $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$ [D] $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$

4 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

[A] Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0
[B] Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina
[C] Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0
[D] Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0

Linearni klasifikacijski modeli *(3 pitanja)*

5 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 1:3   B OVO:OVR $\approx$ 32:50   C OVO:OVR $\approx$ 1:405   D OVO:OVR $\approx$ 4:5

6 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

A Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$

B Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

C Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika

D Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$

7 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, 0.5, 2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.50   B 0.09   C 0.22   D 1.42

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

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

9 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{2}{5}\sqrt{10}$ puta   B $\frac{3}{5}\sqrt{10}$ puta   C $\frac{1}{6}\sqrt{2}$ puta   D $\frac{4}{5}\sqrt{10}$ puta

10 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ manja, to veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

B Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

C Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

D Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

11 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x}$ = “zemlja” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

[A] $y = 0$ i $y = 0$ [B] $y = 1$ i $y = 1$ [C] $y = 0$ i $y = 2$ [D] $y = 0$ i $y = 1$

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{dipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

[A] 0.741 [B] 0.588 [C] 0.322 [D] 0.488

13 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

[A] Funkcija koja uzoku $\mathcal{D}$ pridjeljuje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

[B] Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani

[C] Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

[D] Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

14 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

[A] $(3, 2), (4, 5), (9, 7)$ [B] $(1, 3), (5, 4), (7, 9)$ [C] $(1, 2), (3, 4), (6, 7)$ [D] $(1, 6), (3, 3), (5, 0)$

Probabilistički grafički modeli *(3 pitanja)*

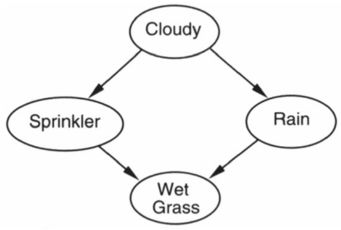
15 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

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

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

16 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S=1\|C)$
0	0.5
1	0.4

C	$P(R=1\|C)$
0	0.2
1	0.8

S	R	$P(W=1\|S,R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

A 0.144 B 0.309 C 0.223 D 0.069

17 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest "efekt objašnjavanja", koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

A U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y nezavisni, ali postaju zavisni ako je opažen posredni uzrok z

B U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z

C U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y zavisni, ali postaju nezavisni ako je opažena posljedica z

D U strukturi $x \leftarrow z \to y$, gdje su posljedice x i y zavisne, ali postaju nezavisne ako je opažen uzrok z

Grupiranje *(3 pitanja)*

18 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina
$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom
$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom
$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_2 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$ C $LL^0_2 \leqslant LL^0_4,\ LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

B $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3,\ LL^*_1 \geqslant LL^*_2$ D $LL^0_3 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$

19 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

- A Oba algoritma izvode se onoliko iteracija koliko ima grupa, no algoritam k-medoida može raditi s općenitom mjerom sličnosti ili udaljenosti
- B Oba algoritma mogu grupirati na temelju mjere udaljenosti koja nije euklidska, no algoritam k-medoida ima veću vremensku složenost od algoritma HAC
- C Oba algoritma mogu grupirati primjere koji nisu vektorizirani, no algoritam HAC daje hijerarhiju dok algoritam k-medoida daje particiju grupa
- D Oba algoritma imaju složenost linearnu u broju primjera, no algoritam k-medoida provodi čvrsto a algoritam HAC hijerarhijsko grupiranje

20 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$\begin{array}{c} \\ \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array}
\begin{array}{c} \begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\
\begin{pmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.5 \\ & 1 & 0.9 & 0.4 & 0.6 \\ & & 1 & 0.8 & 0.1 \\ & & & 1 & 0.6 \\ & & & & 1 \end{pmatrix} \end{array}$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

- A $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$
- B $((((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), \mathbf{x}^{(1)}), \mathbf{x}^{(5)})$
- C $(((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(5)}))$
- D $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$

Vrednovanje modela *(2 pitanja)*

21 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \dots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 2820 B 4096 C 2980 D 5585

22 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\big\{(y^{(i)}, h(\mathbf{x}^{(i)})\big\}_{i=1}^{10} = \big\{(1, 0.2), (0, 0.2), (0, 0.6), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.6), (1, 0.2), (0, 0.6), (1, 0.8)\big\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0.24 B 0.35 C 0 D 0.16

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-1.91, 0.22, -0.47, -0.06)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln\mathcal{L}(\mathbf{w}|\mathbf{\Phi},\mathbf{y})$?

A -4.73 B -10.64 C -12.63 D -5.69

2 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

- A Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0
- B Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0
- C Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina
- D Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0

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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0,1\}$, gdje $y=1$ označava postojanje karcinoma, a $y=0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

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

4 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \wedge h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

A $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ B $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$ C $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ D $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$

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 $\mathbf{w} = (0.1, 0.5, -2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.09 B 1.42 C 0.50 D 0.22

6 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

A Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$

B Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

C Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika

D Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$

7 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 4:5 B OVO:OVR $\approx$ 32:50 C OVO:OVR $\approx$ 1:405 D OVO:OVR $\approx$ 1:3

Jezgrene i neparametarske metode *(4 pitanja)*

8 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{4}{5}\sqrt{10}$ puta B $\frac{1}{6}\sqrt{2}$ puta C $\frac{2}{5}\sqrt{10}$ puta D $\frac{3}{5}\sqrt{10}$ puta

9 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x}$ = “fire” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 0$ B $y = 0$ i $y = 1$ C $y = 0$ i $y = 2$ D $y = 1$ i $y = 1$

10 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ veća, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

B Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

C Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

D Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

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

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

Lagrangeovi koeficijenti su $\alpha_1 = 2.214 \cdot 10^{-8}$, $\alpha_2 = 3.803 \cdot 10^{-8}$ i $\alpha_3 = 6.017 \cdot 10^{-8}$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

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

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

A Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

B Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

C Funkcija koja uzoku $\mathcal{D}$ pridjeluje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

D Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani

13 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

A $(1, 6), (3, 3), (5, 0)$ B $(3, 2), (4, 5), (9, 7)$ C $(1, 2), (3, 4), (6, 7)$ D $(1, 3), (5, 4), (7, 9)$

14 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{da}, \text{zimski}, \text{dipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.588 B 0.741 C 0.488 D 0.322

Probabilistički grafički modeli *(3 pitanja)*

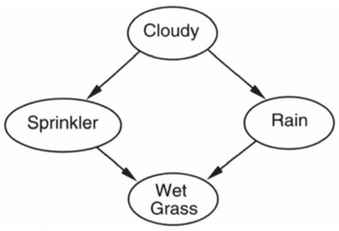
15 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

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

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

16 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S = 1\|C)$
0	0.5
1	0.1

C	$P(R = 1\|C)$
0	0.2
1	0.8

S	R	$P(W = 1\|S, R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

A 0.223 B 0.309 C 0.069 D 0.144

17 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest "efekt objašnjavanja", koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

A U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y zavisni, ali postaju nezavisni ako je opažen posredni uzrok z

B U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z

C U strukturi $x \leftarrow z \to y$, gdje su posljedice x i y nezavisni, ali postaju zavisni ako je opažen uzrok z

D U strukturi $x \leftarrow z \to y$, gdje su posljedice x i y zavisne, ali postaju nezavisne ako je opažen uzrok z

Grupiranje *(3 pitanja)*

18 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

A Oba algoritma mogu grupirati na temelju mjere udaljenosti koja nije euklidska, no algoritam k-medoida ima veću vremensku složenost od algoritma HAC

B Oba algoritma iziskuju odabir početnih medoida grupa, no kod algoritma HAC taj se odabir može napraviti jednostrukim ili potpunim povezivanjem

C Oba algoritma imaju složenost linearnu u broju primjera, no algoritam k-medoida provodi čvrsto a algoritam HAC hijerarhijsko grupiranje

D Oba algoritma imaju vremensku složenost veću od algoritma k-sredina, ali mogu raditi s općenitom mjerom sličnosti između primjera

19 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$\begin{array}{c c} & \begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\ \begin{array}{c} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array} & \begin{pmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.5 \\ & 1 & 0.9 & 0.4 & 0.6 \\ & & 1 & 0.8 & 0.1 \\ & & & 1 & 0.6 \\ & & & & 1 \end{pmatrix} \end{array}$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

A $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$ C $(((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(5)}))$

B $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$ D $((((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), \mathbf{x}^{(1)}), \mathbf{x}^{(5)})$

20 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina

$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_3 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$ C $LL^0_2 \leqslant LL^0_4,\ LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

B $LL^0_2 \geqslant LL^0_4,\ LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$ D $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3,\ LL^*_1 \geqslant LL^*_2$

Vrednovanje modela *(2 pitanja)*

21 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\big\{(y^{(i)}, h(\mathbf{x}^{(i)})\big\}_{i=1}^{10} = \big\{(1, 0.6), (0, 0.2), (0, 0.6), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.8), (0, 0.2), (0, 0.2), (1, 0.8)\big\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0.24 B 0 C 0.16 D 0.35

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \ldots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 4096 B 2820 C 2980 D 5585

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

A Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina

B Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0

C Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0

D Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-1.91, 0.22, -0.47, -0.06)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln\mathcal{L}(\mathbf{w}|\mathbf{\Phi},\mathbf{y})$?

A -10.64 B -5.69 C -4.73 D -12.63

3 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \wedge h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

A $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ B $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$ C $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$ D $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$

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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0,1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

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

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

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 $\mathbf{w} = (0.1, -0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.50 B 0.22 C 1.42 D 0.09

6 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 32:50 B OVO:OVR $\approx$ 1:405 C OVO:OVR $\approx$ 4:5 D OVO:OVR $\approx$ 1:3

7 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

A Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

B Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$

C Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$

D Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

A +0.947 B +1.434 C −0.676 D −2.330

9 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ manja, to veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

B Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

C Što je vrijednost γ veća, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

D Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

10 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{3}{5}\sqrt{10}$ puta B $\frac{2}{5}\sqrt{10}$ puta C $\frac{4}{5}\sqrt{10}$ puta D $\frac{1}{6}\sqrt{2}$ puta

11 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} = \text{“fire”}$ pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 1$ B $y = 0$ i $y = 2$ C $y = 1$ i $y = 1$ D $y = 0$ i $y = 0$

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

A Funkcija koja uzoku $\mathcal{D}$ pridjeljuje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

B Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

C Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

D Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani

13 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

A $(3, 2), (4, 5), (9, 7)$ B $(1, 6), (3, 3), (5, 0)$ C $(1, 3), (5, 4), (7, 9)$ D $(1, 2), (3, 4), (6, 7)$

14 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{da}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.322 B 0.488 C 0.588 D 0.741

Probabilistički grafički modeli *(3 pitanja)*

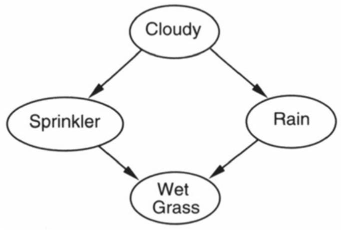
15 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

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

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

16 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S=1\|C)$
0	0.5
1	0.2

C	$P(R=1\|C)$
0	0.2
1	0.8

S	R	$P(W=1\|S,R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

A 0.223 B 0.309 C 0.069 D 0.144

17 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest "efekt objašnjavanja", koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

A U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y zavisni, ali postaju nezavisni ako je opažen posredni uzrok z

B U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z

C U strukturi $x \leftarrow z \to y$, gdje su posljedice x i y zavisne, ali postaju nezavisne ako je opažen uzrok z

D U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y zavisni, ali postaju nezavisni ako je opažena posljedica z

Grupiranje *(3 pitanja)*

18 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina

$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3$, $LL^*_1 \geqslant LL^*_2$

C $LL^0_2 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$

B $LL^0_2 \leqslant LL^0_4$, $LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

D $LL^0_3 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$

19 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

A Oba algoritma imaju složenost linearnu u broju primjera, no algoritam k-medoida provodi čvrsto a algoritam HAC hijerarhijsko grupiranje

B Oba algoritma iziskuju odabir početnih medoida grupa, no kod algoritma HAC taj se odabir može napraviti jednostrukim ili potpunim povezivanjem

C Oba algoritma imaju vremensku složenost veću od algoritma k-sredina, ali mogu raditi s općenitom mjerom sličnosti između primjera

D Oba algoritma nužno konvergiraju, no algoritam k-medoida konvergira u globalni optimum dok HAC može konvergirati u lokalni optimum

20 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$\begin{array}{c} \\ \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\ \begin{pmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.5 \\ & 1 & 0.9 & 0.4 & 0.6 \\ & & 1 & 0.8 & 0.1 \\ & & & 1 & 0.6 \\ & & & & 1 \end{pmatrix} \end{array}$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

A $(((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(5)}))$ C $((((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), \mathbf{x}^{(1)}), \mathbf{x}^{(5)})$

B $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$ D $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$

Vrednovanje modela *(2 pitanja)*

21 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \dots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 2820 B 2980 C 5585 D 4096

22 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\big\{(y^{(i)}, h(\mathbf{x}^{(i)})\big\}_{i=1}^{10} = \big\{(1, 0.6), (0, 0.2), (0, 0.6), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.8), (0, 0.2), (0, 0.2), (1, 0.8)\big\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0.24 B 0 C 0.16 D 0.35

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

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

2 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-1.91, 0.22, -0.47, -0.06)^\mathrm{T}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x}; \mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu, \sigma^2) = (\sqrt{2\pi}\sigma)^{-1} \exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\mathbf{\Phi}, \mathbf{y})$?

A -10.64 B -12.63 C -4.73 D -5.69

3 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

- A Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina
- B Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0
- C Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0
- D Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0

4 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^\mathrm{T}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = h_1(\mathbf{x}; \boldsymbol{\theta}_1) \wedge h_2(\mathbf{x}; \boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

A $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$ B $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ C $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$ D $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$

Linearni klasifikacijski modeli *(3 pitanja)*

5 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

- A Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$
- B Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$
- C Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika
- D Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

6 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 1:405   B OVO:OVR $\approx$ 4:5   C OVO:OVR $\approx$ 1:3   D OVO:OVR $\approx$ 32:50

7 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, -0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.09   B 0.50   C 1.42   D 0.22

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} =$ “zemlja” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 0$ B $y = 0$ i $y = 1$ C $y = 0$ i $y = 2$ D $y = 1$ i $y = 1$

9 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{2}{5}\sqrt{10}$ puta   B $\frac{1}{6}\sqrt{2}$ puta   C $\frac{3}{5}\sqrt{10}$ puta   D $\frac{4}{5}\sqrt{10}$ puta

10 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

B Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

C Što je vrijednost γ manja, to veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

D Što je vrijednost γ manja, to veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

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

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

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

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

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

A $(1, 3), (5, 4), (7, 9)$ B $(1, 2), (3, 4), (6, 7)$ C $(3, 2), (4, 5), (9, 7)$ D $(1, 6), (3, 3), (5, 0)$

13 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

A Funkcija koja uzoku $\mathcal{D}$ pridjeljuje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

B Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani

C Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

D Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

14 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.741 B 0.588 C 0.322 D 0.488

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest "efekt objašnjavanja", koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

A U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z

B U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y nezavisni, ali postaju zavisni ako je opažen posredni uzrok z

C U strukturi $x \leftarrow z \to y$, gdje su posljedice x i y nezavisni, ali postaju zavisni ako je opažen uzrok z

D U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y zavisni, ali postaju nezavisni ako je opažen posredni uzrok z

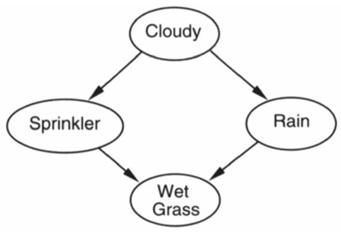
16 (P) Bayesova mreža ima pet varijabli, s topološkim uređajem v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće uvjetne nezavisnosti:

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

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

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

17 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S=1\|C)$
0	0.5
1	0.1

C	$P(R=1\|C)$
0	0.2
1	0.8

S	R	$P(W=1\|S,R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

A 0.223 B 0.069 C 0.144 D 0.309

Grupiranje *(3 pitanja)*

18 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina

$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_3 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$

B $LL^0_2 \leqslant LL^0_4$, $LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$

C $LL^0_2 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$

D $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3$, $LL^*_1 \geqslant LL^*_2$

19 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$\begin{array}{c} \\ \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\ \begin{pmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.5 \\ & 1 & 0.9 & 0.4 & 0.6 \\ & & 1 & 0.8 & 0.1 \\ & & & 1 & 0.6 \\ & & & & 1 \end{pmatrix} \end{array}$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

A $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$
C $(((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(5)}))$
B $((((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), \mathbf{x}^{(1)}), \mathbf{x}^{(5)})$
D $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$

20 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

A Oba algoritma iziskuju odabir početnih medoida grupa, no kod algoritma HAC taj se odabir može napraviti jednostrukim ili potpunim povezivanjem

B Oba algoritma izvode se onoliko iteracija koliko ima grupa, no algoritam k-medoida može raditi s općenitom mjerom sličnosti ili udaljenosti

C Oba algoritma imaju složenost linearnu u broju primjera, no algoritam k-medoida provodi čvrsto a algoritam HAC hijerarhijsko grupiranje

D Oba algoritma mogu grupirati primjere koji nisu vektorizirani, no algoritam HAC daje hijerarhiju dok algoritam k-medoida daje particiju grupa

Vrednovanje modela *(2 pitanja)*

21 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \dots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 2820 B 2980 C 4096 D 5585

22 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\{(y^{(i)}, h(\mathbf{x}^{(i)})\}_{i=1}^{10} = \{(1, 0.6), (0, 0.2), (1, 0.2), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.8), (0, 0.2), (0, 0.2), (1, 0.8)\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0.35 B 0.16 C 0.24 D 0

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

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

2 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

A Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0

B Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina

C Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0

D Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0

3 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\mathbf{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\mathbf{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-1.91, 0.22, -0.47, -0.06)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x}; \mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu, \sigma^2) = (\sqrt{2\pi}\sigma)^{-1} \exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\mathbf{\Phi}, \mathbf{y})$?

A -10.64 B -4.73 C -5.69 D -12.63

4 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0, 1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = h_1(\mathbf{x}; \boldsymbol{\theta}_1) \wedge h_2(\mathbf{x}; \boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

A $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$ B $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ C $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$ D $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$

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 $\mathbf{w} = (0.1, 0.5, 2, -0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.50 B 0.09 C 0.22 D 1.42

6 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

A Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika

B Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$

C Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

D Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$

7 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 1:3 B OVO:OVR $\approx$ 4:5 C OVO:OVR $\approx$ 1:405 D OVO:OVR $\approx$ 32:50

Jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

B Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

C Što je vrijednost γ manja, to veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

D Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

9 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2|/|\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x}$ = “fire” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 0$ B $y = 0$ i $y = 2$ C $y = 1$ i $y = 1$ D $y = 0$ i $y = 1$

10 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{3}{5}\sqrt{10}$ puta   B $\frac{1}{6}\sqrt{2}$ puta   C $\frac{2}{5}\sqrt{10}$ puta   D $\frac{4}{5}\sqrt{10}$ puta

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

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

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

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

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{dipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.488   B 0.588   C 0.322   D 0.741

13 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

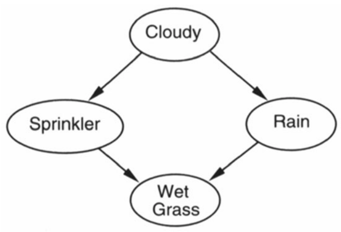
A $(3,2),(4,5),(9,7)$ B $(1,6),(3,3),(5,0)$ C $(1,3),(5,4),(7,9)$ D $(1,2),(3,4),(6,7)$

14 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

- A Funkcija koja uzoku $\mathcal{D}$ pridjeluje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x};\boldsymbol{\theta})$
- B Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x};\boldsymbol{\theta})$
- C Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani
- D Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x};\boldsymbol{\theta})$

Probabilistički grafički modeli *(3 pitanja)*

15 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S=1\|C)$
0	0.5
1	0.1

C	$P(R=1\|C)$
0	0.2
1	0.8

S	R	$P(W=1\|S,R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

A 0.223 B 0.069 C 0.144 D 0.309

16 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest “efekt objašnjavanja”, koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

- A U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y zavisni, ali postaju nezavisni ako je opažen posredni uzrok z
- B U strukturi $x \to z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z
- C U strukturi $x \to z \to y$, gdje su uzrok x i posljedica y nezavisni, ali postaju zavisni ako je opažen posredni uzrok z
- D U strukturi $x \leftarrow z \to y$, gdje su posljedice x i y zavisne, ali postaju nezavisne ako je opažen uzrok z

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

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

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

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

Grupiranje *(3 pitanja)*

18 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina
$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom
$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom
$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_3 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$
C $LL^0_2 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$
B $LL^0_2 \leqslant LL^0_4$, $LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$
D $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3$, $LL^*_1 \geqslant LL^*_2$

19 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$
\begin{array}{c c}
 & \begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\
\begin{array}{c} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array} &
\begin{pmatrix}
1 & 0.3 & 0.5 & 0.7 & 0.5 \\
 & 1 & 0.9 & 0.4 & 0.6 \\
 & & 1 & 0.8 & 0.1 \\
 & & & 1 & 0.6 \\
 & & & & 1
\end{pmatrix}
\end{array}
$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

A $((((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), \mathbf{x}^{(1)}), \mathbf{x}^{(5)})$
C $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$
B $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$
D $(((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(5)}))$

20 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

A Oba algoritma izvode se onoliko iteracija koliko ima grupa, no algoritam k-medoida može raditi s općenitom mjerom sličnosti ili udaljenosti

B Oba algoritma ne računaju centroide grupa, no grupiranje algoritmom k-medoida ovisi o odabiru početnih medoida dok kod algoritma HAC ne ovisi

C Oba algoritma mogu grupirati na temelju mjere udaljenosti koja nije euklidska, no algoritam k-medoida ima veću vremensku složenost od algoritma HAC

D Oba algoritma iziskuju odabir početnih medoida grupa, no kod algoritma HAC taj se odabir može napraviti jednostrukim ili potpunim povezivanjem

Vrednovanje modela *(2 pitanja)*

21 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\{(y^{(i)}, h(\mathbf{x}^{(i)})\}_{i=1}^{10} = \{(1, 0.8), (0, 0.2), (0, 0.6), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.6), (1, 0.2), (0, 0.6), (1, 0.8)\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0.16 B 0.35 C 0.24 D 0

22 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \dots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 2980 B 5585 C 4096 D 2820

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

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 70% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \wedge h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

A $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$ B $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$ C $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$ D $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$

2 (T) Optimizacijom regularizirane funkcije pogreške smanjuje se prenaučenost modela. **Kako je definirana L_2-regularizirana pogreška kod linearne regresije?**

A Zbroj neregularizirane pogreške i izraza propocionalnog s kvadratom norme vektora težina bez težine w_0

B Zbroj funkcije gubitka po svim primjerima i neregularizirane pogreške bez težine w_0

C Zbroj očekivanja funkcije gubitka unakrsne entropije i druge norme vektora težina

D Zbroj prosjeka kvadratnog gubitka na svim primjerima i kvadrata druge norme vektora težina bez težine w_0

3 (N) Model linearne regresije treniramo na skupu označenih primjera iz dvodimenzijskoga ulaznog prostora:

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

Za preslikavanje iz ulaznog prostora u prostor značajki $\boldsymbol{\Phi}$ koristimo funkciju $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Treniranjem modela na skupu $(\boldsymbol{\Phi}, \mathbf{y})$ dobili smo parametre $\mathbf{w} = (-0.81, 2.32, 0.66, 0.1)^{\mathrm{T}}$. Prisjetite se da probabilistički model linearne regresije šum oko $h(\mathbf{x};\mathbf{w})$ modelira normalnom distribucijom, čija je gustoća vjerojatnosti općenito definirana kao $p(x|\mu,\sigma^2) = (\sqrt{2\pi}\sigma)^{-1}\exp(-\frac{1}{2}\sigma^{-2}(x-\mu)^2)$. Pretpostavite $\sigma^2 = 1$. Uz takav model, zanima nas log-izglednost parametara $\mathbf{w}$ na skupu primjera $\boldsymbol{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\boldsymbol{\Phi},\mathbf{y})$?

A -5.69 B -4.73 C -12.63 D -10.64

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 karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0,1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **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 $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) > 0$

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

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

Linearni klasifikacijski modeli *(3 pitanja)*

5 (P) Na skupu od $N = 1000$ primjera rješavamo problem višeklasne klasifikacije u $K = 4$ klase. Dvije klase imaju svaka po 400 primjera, a dvije svaka po 100 primjera. Razmatramo bismo li koristili SVM u shemi OVO ili SVM u shemi OVR. Model treniramo s jezgrenom funkcijom, no zbog ograničenja na raspoloživu računalnu memoriju moramo pripaziti da Gramova matrica ne postane prevelika. Prisjetite se da je Gramova matrična simetrična, pa je dovoljno pohraniti samo polovicu matrice (bez dijagonale). **Koji je u ovom slučaju najveći omjer veličine Gramove matrice za sheme OVO i OVR**?

A OVO:OVR $\approx$ 1:3 B OVO:OVR $\approx$ 1:405 C OVO:OVR $\approx$ 4:5 D OVO:OVR $\approx$ 32:50

6 (T) Optimizacija parametara logističke regresije algoritmom grupnog gradijentnog spusta tipično u svakoj iteraciji uključuje i linijsko pretraživanje. **Što nam osigurava uporaba linijskog pretraživanja kod optimizacije logističke regresije?**

A Da postupak uvijek konvergira, neovisno o odabranoj stopi učenja η i početnim težinama $\mathbf{w}$

B Da postupak ne može zaglaviti u lokalnome minimumu, pod uvjetom da u skupu $\mathcal{D}$ nema multikolinearnosti

C Da postupak uvijek konvergira, pod uvjetom da su primjeri linearno neodvojivi ili da regulariziramo sa $\lambda > 0$

D Da postupak konvergira brže, pod uvjetom da su primjeri linearno odvojivi i da stopa učenja nije η prevelika

7 (N) Model logističke regresije treniramo stohastičkim gradijentnim spustom. Primjere iz dvodimenzijskog ulaznog prostora preslikali smo u prostor značajki funkcijom $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. U jednoj iteraciji treniranja modela vektor parametara jednak je $\mathbf{w} = (0.1, -0.5, -2, 0.5)$. **Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1, -1), 1)$?**

A 0.09 B 0.22 C 1.42 D 0.50

Jezgrene i neparametarske metode *(4 pitanja)*

8 (P) Razmatramo sljedeći skup označenih primjera u dvodimenzijskome ulaznom prostoru:

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

Na ovom skupu treniramo model SVM-a, i to model s tvrdom marginom te model s mekom marginom sa $C = 1$. Kod modela s mekom marginom za dualne koeficijente vrijedi $\alpha_1 = 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, $\alpha_4 > 0$ i $\alpha_5 = 1$. Skicirajte tvrdu i meku marginu u ulaznome prostoru. **Koliko je meka margina šira od tvrde margine?**

A $\frac{3}{5}\sqrt{10}$ puta B $\frac{2}{5}\sqrt{10}$ puta C $\frac{1}{6}\sqrt{2}$ puta D $\frac{4}{5}\sqrt{10}$ puta

9 (N) Algoritam k-NN koristimo za višeklasnu klasifikaciju riječi prema jeziku kojemu pripadaju. Skup za učenje sastoji se od sljedećih riječi i oznaka klasa:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{(\text{“water”}, 0), (\text{“voda”}, 1), (\text{“zrak”}, 1), (\text{“luft”}, 2), (\text{“feuer”}, 2)\}$$

Kao mjeru sličnosti između primjera koristimo jezgrenu funkciju nad znakovnim nizovima, definiranu kao $\kappa(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1 \cap \mathbf{x}_2| / |\mathbf{x}_1 \cup \mathbf{x}_2|$, gdje je su operacije unije i presjeka definirane nad skupovima slova od kojih se riječi sastoje. Npr., $\kappa(\text{“water”}, \text{“voda”}) = 1/8 = 0.125$. Razmatramo dvije varijante algoritma: 3-NN i težinski k-NN. Kod potonjeg u obzir uzimamo sve primjere, tj. $k = N$. Odredite klasifikaciju primjera $\mathbf{x} =$ “wasser” pomoću ova dva algoritma. U slučaju jednake sličnosti između dva primjera, kao susjed se uzima onaj koji je u skupu $\mathcal{D}$ naveden prvi. U slučaju izjednačenja glasova između klasa, prednost se daje klasi s numerički manjom oznakom y. **U koju će klasu biti klasificiran primjer x algoritmom 3-NN, a u koju algoritmom težinski k-NN?**

A $y = 0$ i $y = 1$ B $y = 1$ i $y = 1$ C $y = 0$ i $y = 0$ D $y = 0$ i $y = 2$

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

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

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

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

11 (T) Gaussova jezgrena funkcija $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ nad primjerima $\mathbf{x}_1$ i $\mathbf{x}_2$ definirana je s parametrom preciznosti γ, gdje $\gamma = 1/2\sigma^2$. Ovaj parametar ima utjecaj na vrijednost jezgrene funkcije, ali i na složenost (nelinearnost) modela jezgrenog stroja s Gaussovom jezgrenom funkcijom. **Kakav je utjecaj parametra γ na vrijednost Gaussove jezgrene funkcije $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, gdje $\mathbf{x}_1 \neq \mathbf{x}_2$, te na nelinearnost modela jezgrenog stroja?**

A Što je vrijednost γ manja, to veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

B Što je vrijednost γ manja, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model više nelinearan

C Što je vrijednost γ veća, to je manja vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

D Što je vrijednost γ veća, to je veća vrijednost $\kappa(\mathbf{x}_1, \mathbf{x}_2)$, i to je model manje nelinearan

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussov Bayesov klasifikator koristimo za klasifikaciju u dvije klase ($y = 1$ i $y = 2$) u dvodimenzijskome ulaznom prostoru ($\mathcal{X} = \mathbb{R}^2$). Apriorne vjerojatnosti klasa su jednake, dok su izglednosti klasa modelirane bivarijatnim Gaussovim distribucijama sa sljedećim parametrima:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \boldsymbol{\mu}_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Skicirajte gustoću zajedničke vjerojatnosti u ulaznome prostoru i granicu između klasa definiranu jednadžbom $h(x_1, x_2) = 0$. **Koje su od sljedećih točaka (x_1, x_2) najbliže točkama kroz koje prolazi ta granica?**

A $(3, 2), (4, 5), (9, 7)$ B $(1, 2), (3, 4), (6, 7)$ C $(1, 3), (5, 4), (7, 9)$ D $(1, 6), (3, 3), (5, 0)$

13 (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{da}, \text{zimski}, \text{preddipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.588 B 0.488 C 0.322 D 0.741

14 (T) Treniranje probabilističkih modela svodi se na procjenu njihovih parametara $\boldsymbol{\theta}$ na temelju funkcije log-izglednosti, $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. **Što je zapravo funkcija log-izglednosti?**

A Funkcija koja uzoku $\mathcal{D}$ pridjeljuje vjerojatnost parametara $\boldsymbol{\theta}$, uz pretpostavku da se parametri pokoravaju distibuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

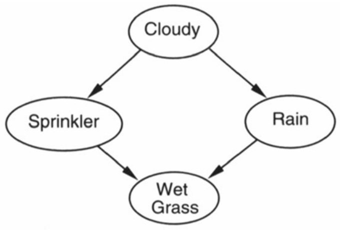
B Funkcija koja primjeru $(\mathbf{x}, y)$ pridjeljuje vjerojatnost pripadanja skupu označenih primjea $\mathcal{D}$, uz pretpostavku da su primjeri nezavisno i identično distribuirani

C Funkcija koja primjeru $\mathbf{x}$ pridjeljuje vjerojatnost $p(y|\mathbf{x})$, uz pretpostavku da se ta vjerojatnost pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

D Funkcija koja parametrima $\boldsymbol{\theta}$ pridjeljuje vjerojatnost uzorka $\mathcal{D}$, uz pretpostavku da se uzorak pokorava distribuciji definiranoj modelom $h(\mathbf{x}; \boldsymbol{\theta})$

Probabilistički grafički modeli *(3 pitanja)*

15 (N) Na slici ispod prikazana je Bayesova mreža za problem prskalice za travu. Varijable su: C (oblačno/*cloudy*), S (prskalica/*sprinkler*), R (kiša/*rain*) i W (mokra trava/*wet grass*). Dane su i tablice uvjetnih vjerojatnosti.



C	$P(S=1\|C)$
0	0.5
1	0.1

C	$P(R=1\|C)$
0	0.2
1	0.8

S	R	$P(W=1\|S,R)$
0	0	0.0
0	1	0.9
1	0	0.1
1	1	0.99

Izračunajte aposteriornu vjerojatnost da radi prskalica ako trava nije mokra i oblačno je.

A 0.223 B 0.144 C 0.069 D 0.309

16 (T) Bayesova mreža može se upotrijebiti za modeliranje kauzalnih odnosa između varijabli, odnosno za zaključivanje o uzrocima i posljedicama događaja. Jedan primjer takvog zaključivanja jest "efekt objašnjavanja", koji se primjenjuje kada postoji interakcija uzorka nekog događaja. **Gdje u Bayesovoj mreži nastupa efekt objašnjavanja i kako se on manifestira?**

A U strukturi $x \leftarrow z \rightarrow y$, gdje su posljedice x i y zavisne, ali postaju nezavisne ako je opažen uzrok z

B U strukturi $x \rightarrow z \rightarrow y$, gdje su uzrok x i posljedica y zavisni, ali postaju nezavisni ako je opažen posredni uzrok z

C U strukturi $x \rightarrow z \leftarrow y$, gdje su uzroci x i y zavisni, ali postaju nezavisni ako je opažena posljedica z

D U strukturi $x \rightarrow z \leftarrow y$, gdje su uzroci x i y nezavisni, ali postaju zavisni ako je opažena posljedica z

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

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

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

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

Grupiranje *(3 pitanja)*

18 (P) Za grupiranje skupa primjera $\mathcal{D}$ koristimo algoritam GMM. Koristimo nekoliko varijanti tog modela:

$\mathcal{H}_1$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina

$\mathcal{H}_2$: Model sa $K = 50$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

$\mathcal{H}_3$: Model sa $K = 50$ slučajno inicijaliziranim središtima i dijeljenom kovarijacijskom matricom

$\mathcal{H}_4$: Model sa $K = 10$ središta inicijaliziranima algoritmom k-sredina i dijeljenom kovarijacijskom matricom

Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je LL^0_α prosječna log-izglednost za model $\mathcal{H}_\alpha$ na početku izvođenja EM-algoritma, a neka je LL^*_α prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL^0_2 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_2 \geqslant LL^*_3$ C $LL^0_2 \geqslant LL^0_4 \geqslant LL^0_3$, $LL^*_1 \geqslant LL^*_2$

B $LL^0_2 \leqslant LL^0_4$, $LL^*_2 \leqslant LL^*_1 \geqslant LL^*_3$ D $LL^0_3 \geqslant LL^0_4$, $LL^*_1 \geqslant LL^*_3 \geqslant LL^*_4$

19 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, definiranu sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

$$\begin{array}{c} \\ \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \mathbf{x}^{(3)} \\ \mathbf{x}^{(4)} \\ \mathbf{x}^{(5)} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \mathbf{x}^{(3)} & \mathbf{x}^{(4)} & \mathbf{x}^{(5)} \end{array} \\ \begin{pmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.5 \\ & 1 & 0.9 & 0.4 & 0.6 \\ & & 1 & 0.8 & 0.1 \\ & & & 1 & 0.6 \\ & & & & 1 \end{pmatrix} \end{array}$$

Provedite grupiranje algoritmom HAC s jednostrukim povezivanjem te nacrtajte pripadni dendrogram. Primijetite da dendrogram odgovara binarnom stablu, s pojedinim primjerima u listovima. **Kojem binarnom stablu odgovara dobiveni dendrogram?**

A $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$
C $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$
B $((((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), \mathbf{x}^{(1)}), \mathbf{x}^{(5)})$
D $(((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(5)}))$

20 (T) Algoritam k-medoida proširenje je algoritma k-sredina. **Što algoritam k-medoida i algoritam hijerarhijskog grupiranja HAC imaju zajedničko, a po čemu se razlikuju?**

A Oba algoritma ne računaju centroide grupa, no grupiranje algoritmom k-medoida ovisi o odabiru početnih medoida dok kod algoritma HAC ne ovisi

B Oba algoritma mogu grupirati na temelju mjere udaljenosti koja nije euklidska, no algoritam k-medoida ima veću vremensku složenost od algoritma HAC

C Oba algoritma iziskuju odabir početnih medoida grupa, no kod algoritma HAC taj se odabir može napraviti jednostrukim ili potpunim povezivanjem

D Oba algoritma izvode se onoliko iteracija koliko ima grupa, no algoritam k-medoida može raditi s općenitom mjerom sličnosti ili udaljenosti

Vrednovanje modela *(2 pitanja)*

21 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-5}, 2^{-4}, \ldots, 2^{5}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 5 ponavljanja u vanjskoj petlji i 3 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 4096 B 2980 C 5585 D 2820

22 (N) Logističku regresiju vrednujemo na ispitnome skupu od $N = 10$ primjera. Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\{(y^{(i)}, h(\mathbf{x}^{(i)})\}_{i=1}^{10} = \{(1, 0.6), (0, 0.2), (0, 0.6), (0, 0.6), (1, 0.8), (0, 0.8), (1, 0.8), (0, 0.2), (0, 0.2), (1, 0.8)\}$$

Na temelju ovog uzorka želimo procijeniti mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma). Skicirajte krivulju ROC, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora. **Koliko je ovaj klasifikator prema mjeri AUC bolji od nasumičnog klasifikatora?**

A 0 B 0.16 C 0.24 D 0.35

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