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

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

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi $E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$. **Što iz toga možemo zaključiti?**

- A Model $\mathcal{H}_{\alpha_1}$ je podnaučen
- B Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$
- C Model $\mathcal{H}_{\alpha_2}$ je prenaučen
- D Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

2 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Što je induktivna pristranost preferencije linearnog modela regresije?**

- A Odabir linearnog modela $h(\mathbf{w};\mathbf{y}) = \mathbf{w}^\mathrm{T}\mathbf{x}$
- C Pretpostavka $P(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^N P(y^{(i)}|\mathbf{w})$
- B Maksimizacija iznosa $-\ln\mathcal{L}(\mathbf{w}|\mathbf{y})$
- D Minimizacija iznosa $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2$

3 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike i fizike na državnoj maturi (ukupno 6 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 28 B 12 C 22 D 13

4 (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 $\mathbf{w} = (2.2, -0.66, 0.77, -0.24)^\mathrm{T}$. 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 $\mathbf{\Phi}$ 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

Linearni klasifikacijski modeli *(3 pitanja)*

5 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 100-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\big)$$

Koliko ukupno parametara ima ovaj model?

A 256 B 506 C 511 D 261

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

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

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

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

Početne težine perceptrona neka su sljedeće:

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

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

A 6 B 16 C 22 D 32

7 (T) Postoji poveznica između algoritma logističke regresije (LR) i algoritma neuronske mreže sa sigmoidnim prijenosnim funkcijama (NN). **Koja je točno poveznica između ova dva algoritma?**

A Model dvoslojne NN istovjetan je modelu LR s poopćenim linearnim modelima sa sigmoidnim funkcijama kao baznim funkcijama

B NN i LR imaju istu funkciju pogreške, ali se samo LR može optimirati Newtonovim postupkom jer funkcija gubitka NN nije konveksna

C Model LR istovjetan je modelu dvoslojne NN sa sigmoidnim prijenosnim funkcijama i pogreškom unakrsne entropije

D Jezgreni stroj s Gaussovom jezgrenom funkcijom istovjetan je NN sa L_2-regulariziranom funkcijom pogreške

Jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Može se dokazati da je Gaussova jezgra s hiperparametrom γ Mercerova jezgra. U praktičnome smislu, to znači da Gaussovu jezgru možemo koristiti za jezgreni trik umjesto da eksplicitno koristimo funkciju preslikavanja $\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^m$. **Što to znači u matematičkome smislu?**

A $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ (\mathbf{x}_1 \neq \mathbf{x}_2) \Rightarrow \Big(\exp(-\gamma \|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2) = \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2\Big)$

B $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1)^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_2) = \exp(-\gamma\Delta^2)$ gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$

C $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = \exp(-\gamma \|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2)$

D $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2) = \exp(-\gamma\Delta^2)$, gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$

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

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((0,2,3), +1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-1,4,1), -1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 0.55$ i $\alpha_2 = 0.13$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3,0,-3)$.**

A $+1503.56$ B -1548.91 C -265.35 D $+328.88$

11 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (+1,+1,-1,-1,+1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 5$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.688, 5, 0.688, 5)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za peti primjer, $L(y^{(5)}, h(\mathbf{x}^{(5)}))$?**

A 0.03 B 2.06 C 1.86 D 0.40

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

Parametre μ i σ^2 gustoće vjerojatnosti $p(x|y)$ procjenjujemo MLE-om. Neka su μ_1 i σ_1^2 parametri gustoće vjerojatnosti $p(x|y=1)$ dobiveni MLE-om na podskupu primjera $\mathcal{D}_{y=1}$. **Koliko iznosi log-izglednost $\mathcal{L}(\mu_1, \sigma_1^2|\mathcal{D}_{y=1})$?**

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

13 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u $K = 10$ klasa sa $n = 5$ značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu Σ možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$: Značajke nisu korelirane, no imaju različite varijance unutar klase i između klasa

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

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

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

A $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_2 \supset \mathcal{H}_1$

B $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$

C $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$

D $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$, $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$

14 (T) Probabilistički modeli mogu biti generativni ili diskriminativni. U praksi su diskriminativni modeli nerijetko veće klasifikacijske točnosti od generativnih modela. **Zašto je tomu tako?**

A Diskriminativni modeli modeliraju zajedničku distribuciju primjera i oznaka, pa u slučaju preklapajućih distribucija u ulaznome prostoru ostvaruju veću točnost od generativnih modela

B Kod generativnih modela parametri se procjenjuju metodom najveće izglednosti koja je pristrana, dok se diskriminativni modeli uče gradijentnim spustom koji je statistički nepristran

C Generativni modeli mogu modelirati nelinearne zavisnosti između značajki, međutim kada su značajne stohastički nezavisne, to dovodi do prenaučenosti modela

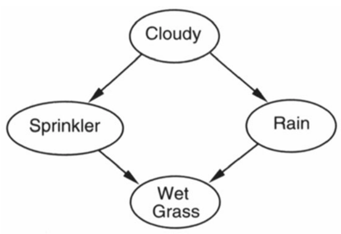
D Diskriminativni modeli s manje parametara mogu modelirati istu granicu između klasa kao i generativni modeli, pa trebaju manje primjera da ih se nauči, a i teže ih je prenaučiti

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koji minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A V, N, T B T C V, P D V, P, T

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	0.3

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 0.144 C 0.069 D 0.223

17 (T) Bayesove mreže možemo upotrijebiti za procjenu aposteriorne vjerojatnosti $P(\mathbf{x}_q|\mathbf{x}_o)$, gdje je $\mathbf{x}_q$ vektor varijabli upita a $\mathbf{x}_o$ vektor opaženih varijabli. U tu se svrhu često koriste metode približnog zaključivanja. Najjednostavnija takva metoda je uzorkovanje s odbacivanjem, međutim ta metoda ima nedostatak zbog koje se u praksi ne koristi. **Koji je nedostatak metode uzorkovanja s odbacivanjem?**

A Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, generiramo vektore iz apriorne distribucije $P(\mathbf{x}_q)$ i procjena je pristrana

B Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, treba generirati mnogo vektora da bi uzorak bio velik i procjena pouzdana

C Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, mnogo će generiranih vektora biti odbačeno i procjena će biti pristrana

D Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, generirani vektori nisu iz aposteriorne distribucije i procjena je netočna

Grupiranje *(3 pitanja)*

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

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

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

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

19 (T) Model Gaussove mješavine često treniramo algoritmom maksimizacije očekivanja. **Na što se odnosi pojam "očekivanje" u nazivu tog algoritma**?

A Izglednost parametara modela izračunata uz fiksiranu pripadnosti svakog primjera njemu najbližoj grupi

B Vjerojatnost skupa podataka pod modelom s fiksiranim parametrima izračunata na temelju vjerojatnosti pripadanja primjera svakoj grupi

C Vjerojatnost parametara modela s fiksiranim dodijeljivanjem primjera grupama izračunata na temelju maksimizacije log-izglednosti

D Vjerojatnost pripadanja primjera svakoj grupi izračunata Bayesovim pravilom na temelju modela Gaussove mješavine

20 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“waterloo”}, \text{“moon”}, \text{“air”}\}$$

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., $\kappa(\text{“water”}, \text{“waterloo”}) = 5/7 = 0.714$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.535 B 0.583 C 0.292 D 0.354

Vrednovanje modela *(2 pitanja)*

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

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

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

A 0.023 B 0.114 C 0.091 D 0.059

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 6820 B 5636 C 5956 D 8855

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

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

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Što je induktivna pristranost preferencije linearnog modela regresije?**

A Maksimizacija iznosa $-\ln \mathcal{L}(\mathbf{w}|\mathbf{y})$

B Odabir linearnog modela $h(\mathbf{w};\mathbf{y}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

C Pretpostavka $P(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^{N} P(y^{(i)}|\mathbf{w})$

D Minimizacija iznosa $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2$

2 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike i fizike na državnoj maturi (ukupno 6 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 12   B 22   C 28   D 13

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 $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1 x_2)$. 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 -4.73 B -10.64 C -5.69 D -12.63

4 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi $E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$. **Što iz toga možemo zaključiti?**

A Model $\mathcal{H}_{\alpha_1}$ je podnaučen

B Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

C Model $\mathcal{H}_{\alpha_2}$ je prenaučen

D Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

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

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

Početne težine perceptrona neka su sljedeće:

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

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

A 32 B 6 C 16 D 22

6 (T) Postoji poveznica između algoritma logističke regresije (LR) i algoritma neuronske mreže sa sigmoidnim prijenosnim funkcijama (NN). **Koja je točno poveznica između ova dva algoritma?**

A Jezgreni stroj s Gaussovom jezgrenom funkcijom istovjetan je NN sa L_2-regulariziranom funkcijom pogreške

B Model dvoslojne NN istovjetan je modelu LR s poopćenim linearnim modelima sa sigmoidnim funkcijama kao baznim funkcijama

C Model LR istovjetan je modelu dvoslojne NN sa sigmoidnim prijenosnim funkcijama i pogreškom unakrsne entropije

D NN i LR imaju istu funkciju pogreške, ali se samo LR može optimirati Newtonovim postupkom jer funkcija gubitka NN nije konveksna

7 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 50-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\big)$$

Koliko ukupno parametara ima ovaj model?

A 256 B 511 C 506 D 261

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

9 (T) Može se dokazati da je Gaussova jezgra s hiperparametrom γ Mercerova jezgra. U praktičnome smislu, to znači da Gaussovu jezgru možemo koristiti za jezgreni trik umjesto da eksplicitno koristimo funkciju preslikavanja $\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^m$. **Što to znači u matematičkome smislu?**

A $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2) = \exp(-\gamma\Delta^2)$, gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$

B $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ (\mathbf{x}_1 \neq \mathbf{x}_2) \Rightarrow \Big(\exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2) = \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2\Big)$

C $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1)^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_2) = \exp(-\gamma\Delta^2)$ gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$

D $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = \exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2)$

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((0,2,3), +1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-1,4,1), -1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 0.55$ i $\alpha_2 = 0.13$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3,0,-3)$.**

A $+1503.56$ B -265.35 C $+328.88$ D -1548.91

11 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (+1,+1,-1,-1,+1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.754, 0.754, 1, 1)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 0.40 B 2.06 C 1.86 D 0.03

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u $K = 10$ klasa sa $n = 5$ značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu Σ možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$: Značajke nisu korelirane, no imaju različite varijance unutar klase i između klasa

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

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

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

A $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_2 \supset \mathcal{H}_1$ C $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$

B $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$, $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$ D $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$

13 (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

Parametre μ i σ^2 gustoće vjerojatnosti $p(x|y)$ procjenjujemo MLE-om. Neka su μ_1 i σ_1^2 parametri gustoće vjerojatnosti $p(x|y=1)$ dobiveni MLE-om na podskupu primjera $\mathcal{D}_{y=1}$. **Koliko iznosi log-izglednost $\mathcal{L}(\mu_1, \sigma_1^2|\mathcal{D}_{y=1})$?**

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

14 (T) Probabilistički modeli mogu biti generativni ili diskriminativni. U praksi su diskriminativni modeli nerijetko veće klasifikacijske točnosti od generativnih modela. **Zašto je tomu tako?**

A Generativni modeli mogu modelirati nelinearne zavisnosti između značajki, međutim kada su značajne stohastički nezavisne, to dovodi do prenaučenosti modela

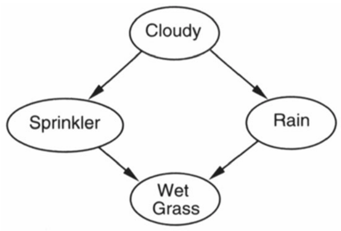
B Kod generativnih modela parametri se procjenjuju metodom najveće izglednosti koja je pristrana, dok se diskriminativni modeli uče gradijentnim spustom koji je statistički nepristran

C Diskriminativni modeli s manje parametara mogu modelirati istu granicu između klasa kao i generativni modeli, pa trebaju manje primjera da ih se nauči, a i teže ih je prenaučiti

D Diskriminativni modeli modeliraju zajedničku distribuciju primjera i oznaka, pa u slučaju preklapajućih distribucija u ulaznome prostoru ostvaruju veću točnost od generativnih modela

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	0.4

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

16 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koji minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A G, T B V, T C T D V, P

17 (T) Bayesove mreže možemo upotrijebiti za procjenu aposteriorne vjerojatnosti $P(\mathbf{x}_q|\mathbf{x}_o)$, gdje je $\mathbf{x}_q$ vektor varijabli upita a $\mathbf{x}_o$ vektor opaženih varijabli. U tu se svrhu često koriste metode približnog zaključivanja. Najjednostavnija takva metoda je uzorkovanje s odbacivanjem, međutim ta metoda ima nedostatak zbog koje se u praksi ne koristi. **Koji je nedostatak metode uzorkovanja s odbacivanjem?**

A Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, mnogo će generiranih vektora biti odbačeno i procjena će biti pristrana

B Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, generiramo vektore iz apriorne distribucije $P(\mathbf{x}_q)$ i procjena je pristrana

C Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, generirani vektori nisu iz aposteriorne distribucije i procjena je netočna

D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, treba generirati mnogo vektora da bi uzorak bio velik i procjena pouzdana

Grupiranje *(3 pitanja)*

18 (T) Model Gaussove mješavine često treniramo algoritmom maksimizacije očekivanja. **Na što se odnosi pojam "očekivanje" u nazivu tog algoritma**?

- A Vjerojatnost parametara modela s fiksiranim dodijeljivanjem primjera grupama izračunata na temelju maksimizacije log-izglednosti
- B Vjerojatnost pripadanja primjera svakoj grupi izračunata Bayesovim pravilom na temelju modela Gaussove mješavine
- C Vjerojatnost skupa podataka pod modelom s fiksiranim parametrima izračunata na temelju vjerojatnosti pripadanja primjera svakoj grupi
- D Izglednost parametara modela izračunata uz fiksiranu pripadnosti svakog primjera njemu najbližoj grupi

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

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

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

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

20 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{"water"}, \text{"waterloo"}, \text{"earth"}, \text{"air"}\}$$

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., $\kappa(\text{"water"}, \text{"waterloo"}) = 5/7 = 0.714$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.583 B 0.535 C 0.292 D 0.354

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 2820 C 2980 D 5585

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

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

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

A 0.059 B 0.023 C 0.091 D 0.114

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

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

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi $E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$. **Što iz toga možemo zaključiti?**

A Model $\mathcal{H}_{\alpha_2}$ je prenaučen

B Model $\mathcal{H}_{\alpha_1}$ je podnaučen

C Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

D Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

2 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike i fizike na državnoj maturi (ukupno 6 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 13 B 28 C 12 D 22

3 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Što je induktivna pristranost preferencije linearnog modela regresije?**

A Minimizacija iznosa $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2$

B Odabir linearnog modela $h(\mathbf{w}; \mathbf{y}) = \mathbf{w}^\mathrm{T}\mathbf{x}$

C Pretpostavka $P(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^N P(y^{(i)}|\mathbf{w})$

D Maksimizacija iznosa $-\ln \mathcal{L}(\mathbf{w}|\mathbf{y})$

4 (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 $\mathbf{w} = (0.28, -0.58, 1.79, -0.75)^\mathrm{T}$. 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 $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\mathbf{\Phi}, \mathbf{y})$?

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

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

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

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

Početne težine perceptrona neka su sljedeće:

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

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

A 32 B 6 C 22 D 16

6 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 50-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 256 B 511 C 506 D 261

7 (T) Postoji poveznica između algoritma logističke regresije (LR) i algoritma neuronske mreže sa sigmoidnim prijenosnim funkcijama (NN). **Koja je točno poveznica između ova dva algoritma?**

A Model dvoslojne NN istovjetan je modelu LR s poopćenim linearnim modelima sa sigmoidnim funkcijama kao baznim funkcijama

B NN i LR imaju istu funkciju pogreške, ali se samo LR može optimirati Newtonovim postupkom jer funkcija gubitka NN nije konveksna

C Model LR istovjetan je modelu dvoslojne NN sa sigmoidnim prijenosnim funkcijama i pogreškom unakrsne entropije

D Jezgreni stroj s Gaussovom jezgrenom funkcijom istovjetan je NN sa L_2-regulariziranom funkcijom pogreške

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (-1, -1, -1, +1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.733, 1, 1, 0.733)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 0.03 B 0.40 C 2.06 D 1.86

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

10 (T) Može se dokazati da je Gaussova jezgra s hiperparametrom γ Mercerova jezgra. U praktičnome smislu, to znači da Gaussovu jezgru možemo koristiti za jezgreni trik umjesto da eksplicitno koristimo funkciju preslikavanja $\boldsymbol{\phi}: \mathbb{R}^n \to \mathbb{R}^m$. **Što to znači u matematičkome smislu?**

- A $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2) = \exp(-\gamma\Delta^2)$, gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$
- B $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1)^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_2) = \exp(-\gamma\Delta^2)$ gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$
- C $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = \exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2)$
- D $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ (\mathbf{x}_1 \neq \mathbf{x}_2) \Rightarrow \Big(\exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2) = \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2\Big)$

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((0, 2, 3), +1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-1, 4, 1), -1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 0.55$ i $\alpha_2 = 0.13$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A -265.35 B $+328.88$ C $+1503.56$ D -1548.91

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u $K = 10$ klasa sa $n = 5$ značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu Σ možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$: Značajke nisu korelirane, no imaju različite varijance unutar klase i između klasa

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

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

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

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

13 (T) Probabilistički modeli mogu biti generativni ili diskriminativni. U praksi su diskriminativni modeli nerijetko veće klasifikacijske točnosti od generativnih modela. **Zašto je tomu tako?**

- A Generativni modeli mogu modelirati nelinearne zavisnosti između značajki, međutim kada su značajne stohastički nezavisne, to dovodi do prenaučenosti modela
- B Kod generativnih modela parametri se procjenjuju metodom najveće izglednosti koja je pristrana, dok se diskriminativni modeli uče gradijentnim spustom koji je statistički nepristran
- C Diskriminativni modeli modeliraju zajedničku distribuciju primjera i oznaka, pa u slučaju preklapajućih distribucija u ulaznome prostoru ostvaruju veću točnost od generativnih modela
- D Diskriminativni modeli s manje parametara mogu modelirati istu granicu između klasa kao i generativni modeli, pa trebaju manje primjera da ih se nauči, a i teže ih je prenaučiti

14 (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

Parametre μ i σ^2 gustoće vjerojatnosti $p(x|y)$ procjenjujemo MLE-om. Neka su μ_1 i σ_1^2 parametri gustoće vjerojatnosti $p(x|y = 1)$ dobiveni MLE-om na podskupu primjera $\mathcal{D}_{y=1}$. **Koliko iznosi log-izglednost $\mathcal{L}(\mu_1, \sigma_1^2|\mathcal{D}_{y=1})$?**

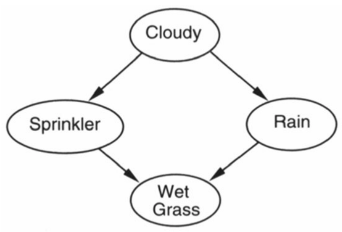
A -11.58 B -7.42 C -10.47 D -9.32

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Bayesove mreže možemo upotrijebiti za procjenu aposteriorne vjerojatnosti $P(\mathbf{x}_q|\mathbf{x}_o)$, gdje je $\mathbf{x}_q$ vektor varijabli upita a $\mathbf{x}_o$ vektor opaženih varijabli. U tu se svrhu često koriste metode približnog zaključivanja. Najjednostavnija takva metoda je uzorkovanje s odbacivanjem, međutim ta metoda ima nedostatak zbog koje se u praksi ne koristi. **Koji je nedostatak metode uzorkovanja s odbacivanjem?**

- A Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, mnogo će generiranih vektora biti odbačeno i procjena će biti pristrana
- B Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, generirani vektori nisu iz aposteriorne distribucije i procjena je netočna
- C Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, generiramo vektore iz apriorne distribucije $P(\mathbf{x}_q)$ i procjena je pristrana
- D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, treba generirati mnogo vektora da bi uzorak bio velik i procjena pouzdana

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	0.3

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 0.069 B 0.309 C 0.223 D 0.144

17 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koji minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A N, T B V, N, T C V, P D T

Grupiranje *(3 pitanja)*

18 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“watering”}, \text{“earth”}, \text{“air”}\}$$

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., $\kappa(\text{“water”}, \text{“watering”}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.535 B 0.583 C 0.478 D 0.292

19 (T) Model Gaussove mješavine često treniramo algoritmom maksimizacije očekivanja. **Na što se odnosi pojam “očekivanje” u nazivu tog algoritma**?

- A Izglednost parametara modela izračunata uz fiksiranu pripadnosti svakog primjera njemu najbližoj grupi
- B Vjerojatnost parametara modela s fiksiranim dodijeljivanjem primjera grupama izračunata na temelju maksimizacije log-izglednosti
- C Vjerojatnost skupa podataka pod modelom s fiksiranim parametrima izračunata na temelju vjerojatnosti pripadanja primjera svakoj grupi
- D Vjerojatnost pripadanja primjera svakoj grupi izračunata Bayesovim pravilom na temelju modela Gaussove mješavine

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

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

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

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

Vrednovanje modela *(2 pitanja)*

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

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

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

A 0.059 B 0.114 C 0.091 D 0.023

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 5956 B 6820 C 8855 D 5636

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

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

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (N) 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 $\mathbf{w} = (2.2, -0.66, 0.77, -0.24)^{\mathrm{T}}$. 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 $\mathbf{\Phi}$ 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

2 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi $E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$. **Što iz toga možemo zaključiti?**

A Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

B Model $\mathcal{H}_{\alpha_2}$ je prenaučen

C Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

D Model $\mathcal{H}_{\alpha_1}$ je podnaučen

3 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike i fizike na državnoj maturi (ukupno 6 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 12 B 22 C 13 D 28

4 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Što je induktivna pristranost preferencije linearnog modela regresije?**

A Pretpostavka $P(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^{N} P(y^{(i)}|\mathbf{w})$

B Minimizacija iznosa $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2$

C Odabir linearnog modela $h(\mathbf{w}; \mathbf{y}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

D Maksimizacija iznosa $-\ln \mathcal{L}(\mathbf{w}|\mathbf{y})$

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

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

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

Početne težine perceptrona neka su sljedeće:

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

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

A 16 B 32 C 6 D 22

6 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 50-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 506 B 256 C 261 D 511

7 (T) Postoji poveznica između algoritma logističke regresije (LR) i algoritma neuronske mreže sa sigmoidnim prijenosnim funkcijama (NN). **Koja je točno poveznica između ova dva algoritma?**

A NN i LR imaju istu funkciju pogreške, ali se samo LR može optimirati Newtonovim postupkom jer funkcija gubitka NN nije konveksna

B Jezgreni stroj s Gaussovom jezgrenom funkcijom istovjetan je NN sa L_2-regulariziranom funkcijom pogreške

C Model dvoslojne NN istovjetan je modelu LR s poopćenim linearnim modelima sa sigmoidnim funkcijama kao baznim funkcijama

D Model LR istovjetan je modelu dvoslojne NN sa sigmoidnim prijenosnim funkcijama i pogreškom unakrsne entropije

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (-1, -1, -1, +1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 10$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 1.098, 10, 3.418, 7.68)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za treći primjer, $L(y^{(3)}, h(\mathbf{x}^{(3)}))$?**

A 0.40 B 0.03 C 2.06 D 1.86

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

10 (T) Može se dokazati da je Gaussova jezgra s hiperparametrom γ Mercerova jezgra. U praktičnome smislu, to znači da Gaussovu jezgru možemo koristiti za jezgreni trik umjesto da eksplicitno koristimo funkciju preslikavanja $\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^m$. **Što to znači u matematičkome smislu?**

- A $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1)^\mathrm{T}\boldsymbol{\phi}(\mathbf{x}_2) = \exp(-\gamma\Delta^2)$ gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^\mathrm{T}(\mathbf{x}_1 - \mathbf{x}_2)$
- B $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ (\mathbf{x}_1 \neq \mathbf{x}_2) \Rightarrow \Big(\exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2) = \mathbf{x}_1^\mathrm{T}\mathbf{x}_2\Big)$
- C $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1^\mathrm{T}\mathbf{x}_2) = \exp(-\gamma\Delta^2)$, gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^\mathrm{T}(\mathbf{x}_1 - \mathbf{x}_2)$
- D $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \mathbf{x}_1^\mathrm{T}\mathbf{x}_2 = \exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2)$

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((0,2,3), +1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-1,4,1), -1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 0.55$ i $\alpha_2 = 0.13$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3,0,-3)$.**

A $+328.88$ B -265.35 C $+1503.56$ D -1548.91

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u $K = 10$ klasa sa $n = 5$ značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu Σ možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$: Značajke nisu korelirane, no imaju različite varijance unutar klase i između klasa

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

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

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

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

13 (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

Parametre μ i σ^2 gustoće vjerojatnosti $p(x|y)$ procjenjujemo MLE-om. Neka su μ_1 i σ_1^2 parametri gustoće vjerojatnosti $p(x|y=1)$ dobiveni MLE-om na podskupu primjera $\mathcal{D}_{y=1}$. **Koliko iznosi log-izglednost $\mathcal{L}(\mu_1, \sigma_1^2|\mathcal{D}_{y=1})$?**

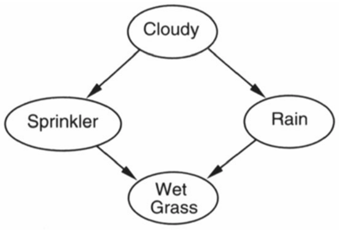
A -11.58 B -10.47 C -7.42 D -9.32

14 (T) Probabilistički modeli mogu biti generativni ili diskriminativni. U praksi su diskriminativni modeli nerijetko veće klasifikacijske točnosti od generativnih modela. **Zašto je tomu tako?**

- A Diskriminativni modeli s manje parametara mogu modelirati istu granicu između klasa kao i generativni modeli, pa trebaju manje primjera da ih se nauči, a i teže ih je prenaučiti
- B Generativni modeli mogu modelirati nelinearne zavisnosti između značajki, međutim kada su značajne stohastički nezavisne, to dovodi do prenaučenosti modela
- C Kod generativnih modela parametri se procjenjuju metodom najveće izglednosti koja je pristrana, dok se diskriminativni modeli uče gradijentnim spustom koji je statistički nepristran
- D Diskriminativni modeli modeliraju zajedničku distribuciju primjera i oznaka, pa u slučaju preklapajućih distribucija u ulaznome prostoru ostvaruju veću točnost od generativnih modela

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 0.223 C 0.069 D 0.144

16 (T) Bayesove mreže možemo upotrijebiti za procjenu aposteriorne vjerojatnosti $P(\mathbf{x}_q|\mathbf{x}_o)$, gdje je $\mathbf{x}_q$ vektor varijabli upita a $\mathbf{x}_o$ vektor opaženih varijabli. U tu se svrhu često koriste metode približnog zaključivanja. Najjednostavnija takva metoda je uzorkovanje s odbacivanjem, međutim ta metoda ima nedostatak zbog koje se u praksi ne koristi. **Koji je nedostatak metode uzorkovanja s odbacivanjem?**

A Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, mnogo će generiranih vektora biti odbačeno i procjena će biti pristrana

B Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, generiramo vektore iz apriorne distribucije $P(\mathbf{x}_q)$ i procjena je pristrana

C Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, treba generirati mnogo vektora da bi uzorak bio velik i procjena pouzdana

D Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, generirani vektori nisu iz aposteriorne distribucije i procjena je netočna

17 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koji minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A V, N, T B G, T C T D V, P

Grupiranje *(3 pitanja)*

18 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“watering”}, \text{“earth”}, \text{“air”}\}$$

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., $\kappa(\text{“water”}, \text{“watering”}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.292 B 0.478 C 0.535 D 0.583

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

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

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

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

20 (T) Model Gaussove mješavine često treniramo algoritmom maksimizacije očekivanja. **Na što se odnosi pojam "očekivanje" u nazivu tog algoritma**?

A Izglednost parametara modela izračunata uz fiksiranu pripadnosti svakog primjera njemu najbližoj grupi

B Vjerojatnost skupa podataka pod modelom s fiksiranim parametrima izračunata na temelju vjerojatnosti pripadanja primjera svakoj grupi

C Vjerojatnost parametara modela s fiksiranim dodijeljivanjem primjera grupama izračunata na temelju maksimizacije log-izglednosti

D Vjerojatnost pripadanja primjera svakoj grupi izračunata Bayesovim pravilom na temelju modela Gaussove mješavine

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 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 5956   B 6820   C 8855   D 5636

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

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

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

A 0.114   B 0.091   C 0.059   D 0.023

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

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

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (N) 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 $\mathbf{w} = (0.28, -0.58, 1.79, -0.75)^{\mathrm{T}}$. 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 $\mathbf{\Phi}$ 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

2 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi $E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$. **Što iz toga možemo zaključiti?**

[A] Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

[B] Model $\mathcal{H}_{\alpha_1}$ je podnaučen

[C] Model $\mathcal{H}_{\alpha_2}$ je prenaučen

[D] Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

3 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike i fizike na državnoj maturi (ukupno 6 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

[A] 13 [B] 28 [C] 22 [D] 12

4 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Što je induktivna pristranost preferencije linearnog modela regresije?**

[A] Odabir linearnog modela $h(\mathbf{w}; \mathbf{y}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

[B] Minimizacija iznosa $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2$

[C] Pretpostavka $P(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^N P(y^{(i)}|\mathbf{w})$

[D] Maksimizacija iznosa $-\ln \mathcal{L}(\mathbf{w}|\mathbf{y})$

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

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

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

Početne težine perceptrona neka su sljedeće:

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

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

A 32 B 16 C 6 D 22

6 (T) Postoji poveznica između algoritma logističke regresije (LR) i algoritma neuronske mreže sa sigmoidnim prijenosnim funkcijama (NN). **Koja je točno poveznica između ova dva algoritma?**

A Model LR istovjetan je modelu dvoslojne NN sa sigmoidnim prijenosnim funkcijama i pogreškom unakrsne entropije

B Jezgreni stroj s Gaussovom jezgrenom funkcijom istovjetan je NN sa L_2-regulariziranom funkcijom pogreške

C NN i LR imaju istu funkciju pogreške, ali se samo LR može optimirati Newtonovim postupkom jer funkcija gubitka NN nije konveksna

D Model dvoslojne NN istovjetan je modelu LR s poopćenim linearnim modelima sa sigmoidnim funkcijama kao baznim funkcijama

7 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 50-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

A 256 B 506 C 261 D 511

Jezgrene i neparametarske metode *(4 pitanja)*

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((0, 2, 3), +1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-1, 4, 1), -1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 0.55$ i $\alpha_2 = 0.13$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A -265.35 B $+1503.56$ C $+328.88$ D -1548.91

9 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (+1, +1, -1, -1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.754, 0.754, 1, 1)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

A 0.40 B 0.03 C 1.86 D 2.06

10 (T) Može se dokazati da je Gaussova jezgra s hiperparametrom γ Mercerova jezgra. U praktičnome smislu, to znači da Gaussovu jezgru možemo koristiti za jezgreni trik umjesto da eksplicitno koristimo funkciju preslikavanja $\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^m$. **Što to znači u matematičkome smislu?**

- A $\forall \mathbf{x}_1 \forall \mathbf{x}_2 .\ (\mathbf{x}_1 \neq \mathbf{x}_2) \Rightarrow \Big(\exp(-\gamma \|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2) = \mathbf{x}_1^\mathrm{T} \mathbf{x}_2 \Big)$
- B $\forall \mathbf{x}_1 \forall \mathbf{x}_2 .\ \mathbf{x}_1^\mathrm{T} \mathbf{x}_2 = \exp(-\gamma \|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2)$
- C $\forall \mathbf{x}_1 \forall \mathbf{x}_2 .\ \boldsymbol{\phi}(\mathbf{x}_1^\mathrm{T} \mathbf{x}_2) = \exp(-\gamma \Delta^2)$, gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^\mathrm{T}(\mathbf{x}_1 - \mathbf{x}_2)$
- D $\forall \mathbf{x}_1 \forall \mathbf{x}_2 .\ \boldsymbol{\phi}(\mathbf{x}_1)^\mathrm{T} \boldsymbol{\phi}(\mathbf{x}_2) = \exp(-\gamma \Delta^2)$ gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^\mathrm{T}(\mathbf{x}_1 - \mathbf{x}_2)$

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

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) Probabilistički modeli mogu biti generativni ili diskriminativni. U praksi su diskriminativni modeli nerijetko veće klasifikacijske točnosti od generativnih modela. **Zašto je tomu tako?**

- A Diskriminativni modeli s manje parametara mogu modelirati istu granicu između klasa kao i generativni modeli, pa trebaju manje primjera da ih se nauči, a i teže ih je prenaučiti
- B Kod generativnih modela parametri se procjenjuju metodom najveće izglednosti koja je pristrana, dok se diskriminativni modeli uče gradijentnim spustom koji je statistički nepristran
- C Generativni modeli mogu modelirati nelinearne zavisnosti između značajki, međutim kada su značajne stohastički nezavisne, to dovodi do prenaučenosti modela
- D Diskriminativni modeli modeliraju zajedničku distribuciju primjera i oznaka, pa u slučaju preklapajućih distribucija u ulaznome prostoru ostvaruju veću točnost od generativnih modela

13 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u $K = 10$ klasa sa $n = 5$ značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu Σ možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$: Značajke nisu korelirane, no imaju različite varijance unutar klase i između klasa

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

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

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

A $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$ C $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$

B $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_2 \supset \mathcal{H}_1$ D $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$, $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$

14 (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

Parametre μ i σ^2 gustoće vjerojatnosti $p(x|y)$ procjenjujemo MLE-om. Neka su μ_1 i σ_1^2 parametri gustoće vjerojatnosti $p(x|y=1)$ dobiveni MLE-om na podskupu primjera $\mathcal{D}_{y=1}$. **Koliko iznosi log-izglednost $\mathcal{L}(\mu_1, \sigma_1^2 | \mathcal{D}_{y=1})$?**

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

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koji minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A T B V, P C G, T D V, N, T

16 (T) Bayesove mreže možemo upotrijebiti za procjenu aposteriorne vjerojatnosti $P(\mathbf{x}_q|\mathbf{x}_o)$, gdje je $\mathbf{x}_q$ vektor varijabli upita a $\mathbf{x}_o$ vektor opaženih varijabli. U tu se svrhu često koriste metode približnog zaključivanja. Najjednostavnija takva metoda je uzorkovanje s odbacivanjem, međutim ta metoda ima nedostatak zbog koje se u praksi ne koristi. **Koji je nedostatak metode uzorkovanja s odbacivanjem?**

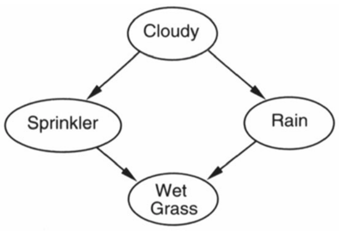
A Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, generiramo vektore iz apriorne distribucije $P(\mathbf{x}_q)$ i procjena je pristrana

B Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, mnogo će generiranih vektora biti odbačeno i procjena će biti pristrana

C Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, generirani vektori nisu iz aposteriorne distribucije i procjena je netočna

D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, treba generirati mnogo vektora da bi uzorak bio velik i procjena pouzdana

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	0.4

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 0.069 B 0.309 C 0.144 D 0.223

Grupiranje *(3 pitanja)*

18 (T) Model Gaussove mješavine često treniramo algoritmom maksimizacije očekivanja. **Na što se odnosi pojam "očekivanje" u nazivu tog algoritma**?

A Izglednost parametara modela izračunata uz fiksiranu pripadnosti svakog primjera njemu najbližoj grupi

B Vjerojatnost pripadanja primjera svakoj grupi izračunata Bayesovim pravilom na temelju modela Gaussove mješavine

C Vjerojatnost parametara modela s fiksiranim dodijeljivanjem primjera grupama izračunata na temelju maksimizacije log-izglednosti

D Vjerojatnost skupa podataka pod modelom s fiksiranim parametrima izračunata na temelju vjerojatnosti pripadanja primjera svakoj grupi

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

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3}\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

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

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

20 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{“water”}, \text{“watering”}, \text{“moon”}, \text{“air”}\}$$

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., $\kappa(\text{“water”}, \text{“watering”}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.583 B 0.535 C 0.292 D 0.354

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 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 5636 B 5956 C 6820 D 8855

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

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

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

A 0.114 B 0.091 C 0.059 D 0.023

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

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

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Što je induktivna pristranost preferencije linearnog modela regresije?**

A Pretpostavka $P(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^{N} P(y^{(i)}|\mathbf{w})$

C Minimizacija iznosa $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2$

B Maksimizacija iznosa $-\ln \mathcal{L}(\mathbf{w}|\mathbf{y})$

D Odabir linearnog modela $h(\mathbf{w}; \mathbf{y}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$

2 (P) Koristimo regresiju za predviđanje uspjeha studija na temelju prosjeka ocjena u četiri razreda srednje škole te uspjeha iz matematike i fizike na državnoj maturi (ukupno 6 značajki). Za preslikavanje u prostor značajki koristimo polinom drugog stupnja s kvadratnim, interakcijskim i linearnim značajkama. Od interakcijskih značajki uzimamo samo interakcije parova značajki. Pretpostavite da nema multikolinearnosti. **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 22 B 13 C 12 D 28

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 $\mathbf{w} = (0.28, -0.58, 1.79, -0.75)^{\mathrm{T}}$. 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 $\mathbf{\Phi}$ 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

4 (P) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi $E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$. **Što iz toga možemo zaključiti?**

A Model $\mathcal{H}_{\alpha_1}$ je podnaučen

B Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

C Model $\mathcal{H}_{\alpha_2}$ je prenaučen

D Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

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

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

Početne težine perceptrona neka su sljedeće:

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

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

[A] 22 [B] 6 [C] 16 [D] 32

6 (P) Poopćeni linearni modeli mogu koristiti adaptivne bazne funkcije. Prednost toga je da ne moramo ručno definirati preslikavanje $\boldsymbol{\phi}$ u prostor značajki, već se to preslikavanje može naučiti na temelju podataka. Rasplažemo podatcima iz $K = 2$ klase u 100-dimenzijskome ulaznom prostoru. Za taj klasifikacijski problem koristimo binarnu logističku regresiju, ali s adaptivnim baznim funkcijama. Svaka adaptivna bazna funkcija ϕ_j je i sama jedan model logističke regresije, pri čemu $\phi_0(\mathbf{x}) = 1$. Naš je model onda definiran ovako:

$$h(\mathbf{x}) = \sigma\Big(\sum_{j=0}^{5} w_j \phi_j(\mathbf{x})\Big)$$

Koliko ukupno parametara ima ovaj model?

[A] 506 [B] 261 [C] 256 [D] 511

7 (T) Postoji poveznica između algoritma logističke regresije (LR) i algoritma neuronske mreže sa sigmoidnim prijenosnim funkcijama (NN). **Koja je točno poveznica između ova dva algoritma?**

[A] Model dvoslojne NN istovjetan je modelu LR s poopćenim linearnim modelima sa sigmoidnim funkcijama kao baznim funkcijama

[B] Model LR istovjetan je modelu dvoslojne NN sa sigmoidnim prijenosnim funkcijama i pogreškom unakrsne entropije

[C] Jezgreni stroj s Gaussovom jezgrenom funkcijom istovjetan je NN sa L_2-regulariziranom funkcijom pogreške

[D] NN i LR imaju istu funkciju pogreške, ali se samo LR može optimirati Newtonovim postupkom jer funkcija gubitka NN nije konveksna

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Pomoću SVM-a rješavamo problem binarne klasifikacije grafova. Budući da su primjeri $\mathbf{x}$ grafovi, koristimo SVM s jezgrenom funkcijom nad grafovima. Model treniramo na skupu od $N = 5$ označenih primjera, s vektorom oznaka jednakim $\mathbf{y} = (-1, -1, -1, +1, +1)$ i sa sljedećom jezgrenom matricom:

$$\mathbf{K} = \begin{pmatrix} 1.0 & 0.97 & -0.949 & -0.555 & -0.986 \\ 0.97 & 1.0 & -0.844 & -0.336 & -0.917 \\ -0.949 & -0.844 & 1.0 & 0.789 & 0.988 \\ -0.555 & -0.336 & 0.789 & 1.0 & 0.684 \\ -0.986 & -0.917 & 0.988 & 0.684 & 1.0 \end{pmatrix}$$

Treniranjem uz $C = 1$ za vektor dualnih parametara dobili smo $\boldsymbol{\alpha} = (0, 0.733, 1, 1, 0.733)$. **Koliko iznosi gubitak zglobnice ovako naučenog modela SVM za četvrti primjer, $L(y^{(4)}, h(\mathbf{x}^{(4)}))$?**

[A] 2.06 [B] 0.40 [C] 1.86 [D] 0.03

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

10 (T) Može se dokazati da je Gaussova jezgra s hiperparametrom γ Mercerova jezgra. U praktičnome smislu, to znači da Gaussovu jezgru možemo koristiti za jezgreni trik umjesto da eksplicitno koristimo funkciju preslikavanja $\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^m$. **Što to znači u matematičkome smislu?**

- A $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2) = \exp(-\gamma\Delta^2)$, gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$
- B $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = \exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2)$
- C $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ \boldsymbol{\phi}(\mathbf{x}_1)^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_2) = \exp(-\gamma\Delta^2)$ gdje $\Delta^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}}(\mathbf{x}_1 - \mathbf{x}_2)$
- D $\forall \mathbf{x}_1 \forall \mathbf{x}_2.\ (\mathbf{x}_1 \neq \mathbf{x}_2) \Rightarrow \Big(\exp(-\gamma\|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)\|^2) = \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2\Big)$

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

$$(\mathbf{x}^{(1)}, y^{(1)}) = ((0, 2, 3), +1)$$
$$(\mathbf{x}^{(2)}, y^{(2)}) = ((-1, 4, 1), -1)$$

Lagrangeovi koeficijenti su $\alpha_1 = 0.55$ i $\alpha_2 = 0.13$. **Upotrijebite jezgreni trik da biste odredili vrijednost hipoteze $h(\mathbf{x})$ za primjer $\mathbf{x} = (3, 0, -3)$.**

A $+1503.56$ B $+328.88$ C -265.35 D -1548.91

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (P) Gaussovim Bayesovim klasifikatorom rješavamo problem klasifikacije u $K = 10$ klasa sa $n = 5$ značajki. Prisjetite se da kod Gaussovog Bayesovog klasifikatora uvođenjem odgovarajućih pretpostavki na kovarijacijsku matricu Σ možemo utjecati na broj parametara modela a time onda i na složenost modela. Razmatramo tri modela s kovarijacijskim matricama u koje smo ugradili sljedeće pretpostavke:

$\mathcal{H}_1$: Značajke nisu korelirane, no imaju različite varijance unutar klase i između klasa

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

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

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

A $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_2 \supset \mathcal{H}_1$ C $\mathcal{H}_1 > \mathcal{H}_2 > \mathcal{H}_3$, $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \mathcal{H}_3$

B $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$ D $\mathcal{H}_1 > \mathcal{H}_3 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$

13 (T) Probabilistički modeli mogu biti generativni ili diskriminativni. U praksi su diskriminativni modeli nerijetko veće klasifikacijske točnosti od generativnih modela. **Zašto je tomu tako?**

- A Kod generativnih modela parametri se procjenjuju metodom najveće izglednosti koja je pristrana, dok se diskriminativni modeli uče gradijentnim spustom koji je statistički nepristran
- B Diskriminativni modeli modeliraju zajedničku distribuciju primjera i oznaka, pa u slučaju preklapajućih distribucija u ulaznome prostoru ostvaruju veću točnost od generativnih modela
- C Generativni modeli mogu modelirati nelinearne zavisnosti između značajki, međutim kada su značajne stohastički nezavisne, to dovodi do prenaučenosti modela
- D Diskriminativni modeli s manje parametara mogu modelirati istu granicu između klasa kao i generativni modeli, pa trebaju manje primjera da ih se nauči, a i teže ih je prenaučiti

14 (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

Parametre μ i σ^2 gustoće vjerojatnosti $p(x|y)$ procjenjujemo MLE-om. Neka su μ_1 i σ_1^2 parametri gustoće vjerojatnosti $p(x|y = 1)$ dobiveni MLE-om na podskupu primjera $\mathcal{D}_{y=1}$. **Koliko iznosi log-izglednost $\mathcal{L}(\mu_1, \sigma_1^2|\mathcal{D}_{y=1})$?**

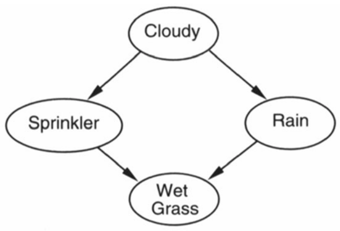
A -7.42 B -10.47 C -11.58 D -9.32

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Bayesove mreže možemo upotrijebiti za procjenu aposteriorne vjerojatnosti $P(\mathbf{x}_q|\mathbf{x}_o)$, gdje je $\mathbf{x}_q$ vektor varijabli upita a $\mathbf{x}_o$ vektor opaženih varijabli. U tu se svrhu često koriste metode približnog zaključivanja. Najjednostavnija takva metoda je uzorkovanje s odbacivanjem, međutim ta metoda ima nedostatak zbog koje se u praksi ne koristi. **Koji je nedostatak metode uzorkovanja s odbacivanjem?**

- A Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, generirani vektori nisu iz aposteriorne distribucije i procjena je netočna
- B Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, generiramo vektore iz apriorne distribucije $P(\mathbf{x}_q)$ i procjena je pristrana
- C Ako je vjerojatnost $P(\mathbf{x}_o)$ velika, mnogo će generiranih vektora biti odbačeno i procjena će biti pristrana
- D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, treba generirati mnogo vektora da bi uzorak bio velik i procjena pouzdana

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 0.069 B 0.223 C 0.309 D 0.144

17 (P) Bayesovom mrežom sa sedam binarnih varijabli modeliramo prometne (ne)prilike u gradu Zagrebu. U našoj mreži, strelice modeliraju pretpostavljeni smjer kauzalnosti. Jutarnje doba dana (J) ili loše vrijeme (V) mogu uzrokovati nastanak prometne gužve (G). Loše vrijeme također može uzrokovati nastanak prometne nesreće (N). Nadalje, nastanak prometne nesreće može uzrokovati nastanak prometne gužve. Loše vrijeme također može uzrokovati zastoj tramvaja (T), no i nestanak struje (S) može uzrokovati zastoj tramvaja. Konačno, zastoj tramvaja može uzrokovati masovno pješačenje putnika (P), što opet može uzrokovati prometnu nesreću. Primjenom algoritma d-odvajanja ispitujemo stohastičku zavisnost između jutarnjeg doba dana (J) i nestanka struje (S). **Koji minimalan skup varijabli treba biti opažen, a da vjerojatnost od S ne ovisi o vrijednosti J?**

A T B N, T C V, P D V, T

Grupiranje *(3 pitanja)*

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

$$p(\mathbf{x}) = \sum_{j=1}^{3} \frac{1}{3} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad \text{gdje} \quad \mu_1 = (5,5), \mu_2 = (5,10), \mu_3 = (-10,-10), \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2\mathbf{I}$$

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

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

19 (T) Model Gaussove mješavine često treniramo algoritmom maksimizacije očekivanja. **Na što se odnosi pojam "očekivanje" u nazivu tog algoritma**?

A Vjerojatnost skupa podataka pod modelom s fiksiranim parametrima izračunata na temelju vjerojatnosti pripadanja primjera svakoj grupi

B Vjerojatnost pripadanja primjera svakoj grupi izračunata Bayesovim pravilom na temelju modela Gaussove mješavine

C Izglednost parametara modela izračunata uz fiksiranu pripadnosti svakog primjera njemu najbližoj grupi

D Vjerojatnost parametara modela s fiksiranim dodijeljivanjem primjera grupama izračunata na temelju maksimizacije log-izglednosti

20 (N) Algoritmom HAC grupiramo riječi engleskog jezika. Neoznačeni skup podataka sastoji se od sljedećih riječi:

$$\mathcal{D} = \{\text{"water"}, \text{"watering"}, \text{"earth"}, \text{"air"}\}$$

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., $\kappa(\text{"water"}, \text{"watering"}) = 5/8 = 0.625$. Provedite prve dvije iteracije grupiranja algoritmom HAC uz prosječno povezivanje. **Na kojoj se razini sličnosti spajaju grupe u drugoj iteraciji algoritma HAC?**

A 0.583 B 0.535 C 0.478 D 0.292

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 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 5956 B 6820 C 5636 D 8855

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

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

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

A 0.091 B 0.114 C 0.059 D 0.023

	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	D	A	B	C	B	A	B	B	B	D	D	C	D	C	D	B	B	B	C	C	C
Grupa B	D	C	C	D	A	B	D	D	C	D	D	C	D	C	B	D	D	C	D	A	C	D
Grupa C	D	B	A	A	B	D	A	D	D	B	D	C	D	A	D	C	C	A	C	C	B	A
Grupa D	D	A	D	B	D	C	C	C	B	A	D	A	B	A	B	C	D	C	A	B	A	A
Grupa E	C	A	B	B	A	D	C	D	B	D	B	A	A	C	B	D	B	D	C	D	B	A
Grupa F	C	D	C	D	C	D	A	C	B	C	D	D	D	D	D	C	C	D	A	B	A	B