Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ i $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ pogreške hipoteze $h(\mathbf{x};\boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

- A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ po $\boldsymbol{\theta}$
- B Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ po $\boldsymbol{\alpha} \in \mathcal{A}$
- C Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$
- D Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

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

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 $\mathbf{w} = (-0.81, 2.32, 0.66, 0.1)^{\mathrm{T}}$. 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 $\mathbf{\Phi}$ s oznakama $\mathbf{y}$. **Koliko iznosi log-izglednost** $\ln \mathcal{L}(\mathbf{w}|\mathbf{\Phi}, \mathbf{y})$?

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

4 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva "hrbatna regresija". **Koji je učinak L2-regularizacije kod linearne regresije?**

- A Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara
- B Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu
- C Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara
- D Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara

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 261 B 506 C 511 D 256

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (1, 0, -1, 2, 3, 2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 100 B 40 C 18 D 96

7 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 10$ B $\lambda = 0$ C $\lambda = 25$ D $\lambda = 5$

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

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

9 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, +1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (0, 0.223, 10, 9.777)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.892$ te vrijedi $w_0^* = 1.306$. **Koliko iznosi širina margine za ovaj model?**

A 0.409 B 0.640 C 0.426 D 0.665

10 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

A Tvrda margina s linearnom jezgrom i velikom vrijednosti za C

B Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ

C Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ

D Meka margina s linearnom jezgrom i malom vrijednosti za C

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 1/2 B 0 C 3/4 D 1/4

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.2. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

A $\alpha = 2, \beta = 5$ B $\alpha = 2, \beta = 10$ C $\alpha = 5, \beta = 7$ D $\alpha = 4, \beta = 8$

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

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

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

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

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

A -3.03 B -4.13 C -2.75 D -3.84

14 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

- A Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela
- B Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela
- C Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela
- D Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model

Probabilistički grafički modeli *(3 pitanja)*

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

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

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

A 0.0786 B 0.1274 C 0.1877 D 0.0490

16 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

- A Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje
- B Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$
- C Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran
- D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen

17 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\begin{aligned}
\mathcal{H}_1 &: P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y)\\
\mathcal{H}_2 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z)\\
\mathcal{H}_3 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)
\end{aligned}$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

- A $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$
- B $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_1|\mathcal{D}_i)$
- C $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$
- D $E(h_1|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$

Grupiranje *(3 pitanja)*

18 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3: &\quad \big\{\{1,2,4,4\},\{2,2,3\},\{1,1,3,4\}\big\} \\ K = 4: &\quad \big\{\{1,2,2,4,4\},\{2,3,3\},\{1,1\},\{4\}\big\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

[A] 0.0827 [B] 0.1364 [C] 0.0799 [D] 0.0545

19 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

[A] Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k

[B] Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k

[C] Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k

[D] Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k

20 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N - 2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c c} & \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\ \begin{array}{c} G_1 \\ G_2 \\ G_3 \end{array} & \begin{pmatrix} 0 & 0.8 & 0.2 \\ 0.8 & 0 & 0.3 \\ 0.2 & 0.3 & 0 \end{pmatrix} \end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

[A] 0.589 [B] 0.411 [C] 0.744 [D] 0.878

Vrednovanje modela *(2 pitanja)*

21 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skupu sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.3, 0.8, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R = 0$ uzmite $P = 1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

[A] 0.60 [B] 0.56 [C] 0.50 [D] 0.66

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):

$$\begin{array}{c c} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 36 \end{pmatrix} \end{array}$$

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] 0.015 [B] 0.040 [C] 0.094 [D] 0.068

Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ i $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ pogreške hipoteze $h(\mathbf{x};\boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

- A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ po $\boldsymbol{\theta}$
- B Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$
- C Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$
- D Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ po $\boldsymbol{\alpha} \in \mathcal{A}$

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

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

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

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

3 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva “hrbatna regresija”. **Koji je učinak L2-regularizacije kod linearne regresije?**

- A Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu
- B Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara
- C Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara
- D Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara

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

Linearni klasifikacijski modeli *(3 pitanja)*

5 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 5$ B $\lambda = 25$ C $\lambda = 0$ D $\lambda = 10$

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 261 B 506 C 511 D 256

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (1, 0, -1, 2, 3, 2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 96 B 100 C 40 D 18

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

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

9 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

A Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ

B Tvrda margina s linearnom jezgrom i velikom vrijednosti za C

C Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ

D Meka margina s linearnom jezgrom i malom vrijednosti za C

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A $3/4$ B 0 C $1/4$ D $1/2$

11 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, -1, +1, +1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (9.654, 10, 0.346, 0)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.724$ te vrijedi $w_0^* = 1.376$. **Koliko iznosi širina margine za ovaj model?**

A 0.665 B 0.640 C 0.426 D 0.409

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

A Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model

B Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela

C Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela

D Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela

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

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

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

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

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

A -3.03 B -4.13 C -3.84 D -2.75

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

A $\alpha = 2, \beta = 5$ B $\alpha = 2, \beta = 10$ C $\alpha = 4, \beta = 8$ D $\alpha = 5, \beta = 7$

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\mathcal{H}_1 : P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y)$$
$$\mathcal{H}_2 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z)$$
$$\mathcal{H}_3 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

- A $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_1|\mathcal{D}_i)$
- B $E(h_1|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_1|\mathcal{D}_i)$
- C $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$
- D $E(h_1|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$

16 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

- A Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran
- B Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje
- C Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen
- D Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$

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

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

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

A 0.1274 B 0.1877 C 0.0786 D 0.0490

Grupiranje *(3 pitanja)*

18 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

- A Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k
- B Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k
- C Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k
- D Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k

19 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3: &\quad \big\{\{1,2,4,4\},\{2,2,3\},\{1,1,3,4\}\big\} \\ K = 4: &\quad \big\{\{1,2,2,4,4\},\{2,3,3\},\{1,1\},\{4\}\big\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

A 0.0545 B 0.1364 C 0.0799 D 0.0827

20 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N - 2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c} \\ G_1 \\ G_2 \\ G_3 \end{array} \begin{array}{c} \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\ \begin{pmatrix} 0 & 0.6 & 0.2 \\ 0.6 & 0 & 0.5 \\ 0.2 & 0.5 & 0 \end{pmatrix} \end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

A 0.589 B 0.744 C 0.411 D 0.878

Vrednovanje modela *(2 pitanja)*

21 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skupu sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.2, 0.6, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R = 0$ uzmite $P = 1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

A 0.66 B 0.50 C 0.56 D 0.60

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):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 26 \end{pmatrix} \end{array}$$

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 0.015 B 0.040 C 0.094 D 0.068

Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) 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 Optimalan model je onaj s hiperparametrom iz intervala $[\alpha_1, \alpha_2]$

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

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

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

2 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ i $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ pogreške hipoteze $h(\mathbf{x};\boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ po $\boldsymbol{\theta}$

B Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

C Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

D Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ po $\boldsymbol{\alpha} \in \mathcal{A}$

3 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva "hrbatna regresija". **Koji je učinak L2-regularizacije kod linearne regresije?**

A Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara

B Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara

C Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara

D Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu

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

A -10.64 B -4.73 C -12.63 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 506 B 256 C 511 D 261

6 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 0$ B $\lambda = 25$ C $\lambda = 5$ D $\lambda = 10$

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (-1, 0, -1, 2, 3, 2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 18 B 100 C 96 D 40

Jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

A Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ

B Meka margina s linearnom jezgrom i malom vrijednosti za C

C Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ

D Tvrda margina s linearnom jezgrom i velikom vrijednosti za C

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 1/2 B 0 C 3/4 D 1/4

10 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (-1, -1, -1, +1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (0, 0, 10, 10)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.882$ te vrijedi $w_0^* = -0.75$. **Koliko iznosi širina margine za ovaj model?**

A 0.665 B 0.426 C 0.409 D 0.640

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

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

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

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

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

A Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela

B Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela

C Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model

D Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela

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

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

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

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

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_1(\mathbf{x})$ za primjer $\mathbf{x} = (4, 0)$?**

A -3.84 B -2.75 C -3.03 D -4.13

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.4. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

[A] $\alpha = 5, \beta = 7$ [B] $\alpha = 2, \beta = 5$ [C] $\alpha = 4, \beta = 8$ [D] $\alpha = 2, \beta = 10$

Probabilistički grafički modeli *(3 pitanja)*

15 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

[A] Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen

[B] Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$

[C] Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje

[D] Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran

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

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

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

[A] 0.0490 [B] 0.1877 [C] 0.1274 [D] 0.0786

17 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\begin{aligned}
\mathcal{H}_1 &: P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y) \\
\mathcal{H}_2 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z) \\
\mathcal{H}_3 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)
\end{aligned}$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

[A] $E(h_1|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

[B] $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$

[C] $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

[D] $E(h_2|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

Grupiranje *(3 pitanja)*

18 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3: &\quad \big\{\{1,2,4,4\},\{2,2,3,3\},\{1,1,4\}\big\} \\ K = 4: &\quad \big\{\{1,2,2,4,4\},\{2,3,3\},\{1,1\},\{4\}\big\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

A 0.0182 B 0.0364 C 0.0727 D 0.0545

19 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

A Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k

B Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k

C Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k

D Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k

20 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N - 2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c} \\ G_1 \\ G_2 \\ G_3 \end{array} \begin{array}{c} \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\ \begin{pmatrix} 0 & 0.8 & 0.2 \\ 0.8 & 0 & 0.3 \\ 0.2 & 0.3 & 0 \end{pmatrix} \end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

A 0.744 B 0.411 C 0.878 D 0.589

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):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 26 \end{pmatrix} \end{array}$$

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 0.094 B 0.040 C 0.015 D 0.068

22 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skup sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.2, 0.8, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R = 0$ uzmite $P = 1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

A 0.66 B 0.60 C 0.56 D 0.50

Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

Ispit sadrži ***22 pitanja*** *i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti* ***20 pitanja****, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je* ***180 minuta****. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva “hrbatna regresija”. **Koji je učinak L2-regularizacije kod linearne regresije?**

A Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara

B Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu

C Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara

D Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara

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

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

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

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

3 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ i $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ pogreške hipoteze $h(\mathbf{x};\boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

B Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

C Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ po $\boldsymbol{\alpha} \in \mathcal{A}$

D Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ po $\boldsymbol{\theta}$

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 -10.64 B -4.73 C -12.63 D -5.69

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 40 B 96 C 18 D 100

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 261 B 511 C 506 D 256

7 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 10$ B $\lambda = 5$ C $\lambda = 25$ D $\lambda = 0$

Jezgrene i neparametarske metode *(4 pitanja)*

8 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

A Tvrda margina s linearnom jezgrom i velikom vrijednosti za C

B Meka margina s linearnom jezgrom i malom vrijednosti za C

C Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ

D Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ

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

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

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

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

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 1/2 B 1/4 C 0 D 3/4

11 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (-1, -1, -1, +1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (0, 0, 10, 10)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.882$ te vrijedi $w_0^* = -0.75$. **Koliko iznosi širina margine za ovaj model?**

A 0.426 B 0.409 C 0.640 D 0.665

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

A $\alpha = 4, \beta = 8$ B $\alpha = 2, \beta = 5$ C $\alpha = 5, \beta = 7$ D $\alpha = 2, \beta = 10$

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

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

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

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

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

A -3.84 B -3.03 C -2.75 D -4.13

14 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

A Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela

B Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model

C Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela

D Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\begin{aligned}
\mathcal{H}_1 &: P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y)\\
\mathcal{H}_2 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z)\\
\mathcal{H}_3 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)
\end{aligned}$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

- A $E(h_1|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$
- B $E(h_1|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$
- C $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$
- D $E(h_2|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

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

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

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

A 0.0786 B 0.1877 C 0.0490 D 0.1274

17 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

- A Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje
- B Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran
- C Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$
- D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen

Grupiranje *(3 pitanja)*

18 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N-2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c c}
 & \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\
\begin{array}{c} G_1 \\ G_2 \\ G_3 \end{array} & \begin{pmatrix} 0 & 0.9 & 0.2 \\ 0.9 & 0 & 0.7 \\ 0.2 & 0.7 & 0 \end{pmatrix}
\end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

A 0.744 B 0.589 C 0.411 D 0.878

19 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3: &\quad \big\{\{1,2,4,4\},\{2,2,3\},\{1,1,3,4\}\big\} \\ K = 4: &\quad \big\{\{1,2,2,4,4\},\{2,3,3\},\{1,1\},\{4\}\big\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

A 0.1364 B 0.0799 C 0.0827 D 0.0545

20 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

A Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k

B Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k

C Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k

D Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k

Vrednovanje modela *(2 pitanja)*

21 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skupu sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.2, 0.4, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R = 0$ uzmite $P = 1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

A 0.50 B 0.60 C 0.56 D 0.66

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):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 46 \end{pmatrix} \end{array}$$

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 0.094 B 0.068 C 0.015 D 0.040

Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

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

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

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

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

2 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva "hrbatna regresija". **Koji je učinak L2-regularizacije kod linearne regresije?**

A Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara

B Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara

C Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu

D Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara

3 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ i $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ pogreške hipoteze $h(\mathbf{x}; \boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ po $\boldsymbol{\theta}$

B Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

C Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ po $\boldsymbol{\alpha} \in \mathcal{A}$

D Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

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

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

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 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 10$ B $\lambda = 5$ C $\lambda = 0$ D $\lambda = 25$

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 100 B 40 C 96 D 18

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 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 261 B 506 C 256 D 511

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

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

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A $1/2$ B 0 C $1/4$ D $3/4$

10 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

- A Tvrda margina s linearnom jezgrom i velikom vrijednosti za C
- B Meka margina s linearnom jezgrom i malom vrijednosti za C
- C Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ
- D Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ

11 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (-1, -1, -1, +1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (0, 0, 10, 10)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.882$ te vrijedi $w_0^* = -0.75$. **Koliko iznosi širina margine za ovaj model?**

A 0.426 B 0.640 C 0.665 D 0.409

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

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

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

Procijenite parametre modela na ovom skupu primjera. Budući da je skup primjera malen, za procjenu kovarijacijske matrice koristite nepristran procjenitelj. Izlaz modela za klasu $y = j$ neka je zajednička gustoća vjerojatnosti, $h_j(\mathbf{x}) = \ln p(\mathbf{x}, y = j)$. **Koliko iznosi $h_1(\mathbf{x})$ za primjer $\mathbf{x} = (4, 0)$?**

A -3.84 B -4.13 C -3.03 D -2.75

13 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

- A Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela
- B Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela
- C Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model
- D Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.2. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

A $\alpha = 5, \beta = 7$ B $\alpha = 2, \beta = 5$ C $\alpha = 2, \beta = 10$ D $\alpha = 4, \beta = 8$

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\begin{aligned}\mathcal{H}_1 &: P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y)\\ \mathcal{H}_2 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z)\\ \mathcal{H}_3 &: P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)\end{aligned}$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

- A $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_1|\mathcal{D}_i)$
- B $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$
- C $E(h_1|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$
- D $E(h_2|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

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

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

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

A 0.1274 B 0.1877 C 0.0786 D 0.0490

17 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

- A Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen
- B Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje
- C Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$
- D Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran

Grupiranje *(3 pitanja)*

18 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1,2,3,4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned}K=3 &: \big\{\{1,2,4,4\},\{2,2,3\},\{1,1,3,4\}\big\}\\ K=4 &: \big\{\{1,2,2,4,4\},\{2,3\},\{1,1\},\{3,4\}\big\}\end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

A 0.0582 B 0.0364 C 0.0127 D 0.0245

19 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

A Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k

B Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k

C Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k

D Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k

20 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N-2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c} \\ G_1 \\ G_2 \\ G_3 \end{array} \begin{array}{c} \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\ \begin{pmatrix} 0 & 0.8 & 0.2 \\ 0.8 & 0 & 0.3 \\ 0.2 & 0.3 & 0 \end{pmatrix} \end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

A 0.411 B 0.589 C 0.744 D 0.878

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):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 46 \end{pmatrix} \end{array}$$

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 0.015 B 0.068 C 0.094 D 0.040

22 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skupu sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.3, 0.8, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R=0$ uzmite $P=1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

A 0.60 B 0.50 C 0.56 D 0.66

Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (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 -4.73 B -10.64 C -12.63 D -5.69

2 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ i $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ pogreške hipoteze $h(\mathbf{x};\boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

- A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ po $\boldsymbol{\theta}$
- B Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ po $\boldsymbol{\alpha} \in \mathcal{A}$
- C Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$
- D Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

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

4 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva "hrbatna regresija". **Koji je učinak L2-regularizacije kod linearne regresije?**

- A Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara
- B Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara
- C Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara
- D Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (1, 0, -1, 2, 3, 2)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 40 B 96 C 100 D 18

6 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 10$ B $\lambda = 25$ C $\lambda = 0$ D $\lambda = 5$

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 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 261 C 511 D 506

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

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

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 1/2 B 1/4 C 3/4 D 0

10 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

- A Tvrda margina s linearnom jezgrom i velikom vrijednosti za C
- B Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ
- C Meka margina s linearnom jezgrom i malom vrijednosti za C
- D Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ

11 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1, +1, -1, +1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (0, 0.223, 10, 9.777)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.892$ te vrijedi $w_0^* = 1.306$. **Koliko iznosi širina margine za ovaj model?**

A 0.409 B 0.665 C 0.426 D 0.640

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

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

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

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

A -2.75 B -3.84 C -4.13 D -3.03

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.1. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

A $\alpha = 2, \beta = 10$ B $\alpha = 5, \beta = 7$ C $\alpha = 2, \beta = 5$ D $\alpha = 4, \beta = 8$

14 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

- A Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela
- B Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela
- C Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela
- D Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\mathcal{H}_1 : P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y)$$
$$\mathcal{H}_2 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z)$$
$$\mathcal{H}_3 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

- A $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_1|\mathcal{D}_i)$
- B $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$
- C $E(h_1|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_2|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$
- D $E(h_2|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

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

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

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

A 0.1274 B 0.0786 C 0.1877 D 0.0490

17 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

- A Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje
- B Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran
- C Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$
- D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen

Grupiranje *(3 pitanja)*

18 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N-2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c} \\ G_1 \\ G_2 \\ G_3 \end{array} \begin{array}{c} \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\ \begin{pmatrix} 0 & 0.4 & 0.2 \\ 0.4 & 0 & 0.5 \\ 0.2 & 0.5 & 0 \end{pmatrix} \end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

A 0.744 B 0.411 C 0.878 D 0.589

19 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

- A Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k
- B Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k
- C Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k
- D Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k

20 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3: &\quad \big\{\{1,2,4,4\},\{2,2,3,3\},\{1,1,4\}\big\} \\ K = 4: &\quad \big\{\{1,2,2,4,4\},\{2,3,3\},\{1,1\},\{4\}\big\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

A 0.0545 B 0.0727 C 0.0182 D 0.0364

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):

$$\begin{array}{c c} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 36 \end{pmatrix} \end{array}$$

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 0.015 B 0.094 C 0.040 D 0.068

22 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skupu sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.2, 0.4, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R = 0$ uzmite $P = 1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

A 0.66 B 0.60 C 0.50 D 0.56

Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (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 -10.64 C -5.69 D -12.63

2 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva "hrbatna regresija". **Koji je učinak L2-regularizacije kod linearne regresije?**

A Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu

B Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara

C Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara

D Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara

3 (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 $[\alpha_1, \alpha_2]$

4 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_\text{train})$ i $E(\boldsymbol{\theta}|\mathcal{D}_\text{test})$ pogreške hipoteze $h(\mathbf{x};\boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_\text{test})$

B Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_\text{train})$ po $\boldsymbol{\theta}$

C Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_\text{train})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_\text{test})$

D Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_\text{test})$ po $\boldsymbol{\alpha} \in \mathcal{A}$

Linearni klasifikacijski modeli *(3 pitanja)*

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (1, 0, -1, 2, -3, 0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 40 B 100 C 18 D 96

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 511 B 256 C 261 D 506

7 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 10$ B $\lambda = 25$ C $\lambda = 5$ D $\lambda = 0$

Jezgrene i neparametarske metode *(4 pitanja)*

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

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

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

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

9 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

A Meka margina s linearnom jezgrom i malom vrijednosti za C

B Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ

C Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ

D Tvrda margina s linearnom jezgrom i velikom vrijednosti za C

10 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (-1, -1, -1, +1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (0, 0, 10, 10)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.882$ te vrijedi $w_0^* = -0.75$. **Koliko iznosi širina margine za ovaj model?**

A 0.640 B 0.426 C 0.665 D 0.409

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 1/2 B 3/4 C 0 D 1/4

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

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

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

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

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

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

A -2.75 B -3.84 C -4.13 D -3.03

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

A $\alpha = 2, \beta = 5$ B $\alpha = 5, \beta = 7$ C $\alpha = 4, \beta = 8$ D $\alpha = 2, \beta = 10$

14 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

A Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model

B Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela

C Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela

D Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela

Probabilistički grafički modeli *(3 pitanja)*

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

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

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

A 0.1877 B 0.0786 C 0.1274 D 0.0490

16 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

A Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran

B Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$

C Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen

D Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje

17 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\mathcal{H}_1 : P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y)$$
$$\mathcal{H}_2 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z)$$
$$\mathcal{H}_3 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

A $E(h_1|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

B $E(h_2|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

C $E(h_3|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

D $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$

Grupiranje *(3 pitanja)*

18 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

A Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k

B Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k

C Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k

D Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k

19 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1, 2, 3, 4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$\begin{aligned} K = 3: &\quad \big\{\{1,2,4,4\},\{2,3,3\},\{1,1,3,4\}\big\} \\ K = 4: &\quad \big\{\{1,2,2,4,4\},\{3,3\},\{1,1\},\{3,4\}\big\} \end{aligned}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

A 0.0727 B 0.0164 C 0.0945 D 0.0682

20 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N - 2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c} \\ G_1 \\ G_2 \\ G_3 \end{array} \begin{array}{c} \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\ \begin{pmatrix} 0 & 0.8 & 0.2 \\ 0.8 & 0 & 0.3 \\ 0.2 & 0.3 & 0 \end{pmatrix} \end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

A 0.744 B 0.411 C 0.589 D 0.878

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):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 6 \end{pmatrix} \end{array}$$

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 0.040 B 0.015 C 0.068 D 0.094

22 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skupu sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.2, 0.4, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R = 0$ uzmite $P = 1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

A 0.66 B 0.50 C 0.60 D 0.56

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

20. veljače 2024.

Ispit iz Strojnog učenja 1 (ak. god. 2023./2024.)

– NEKORIGIRANA VERZIJA –

*Ispit sadrži **22 pitanja** i ukupno nosi najviše 20 bodova (za 60% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **180 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Osnovni koncepti i linearna regresija *(4 pitanja)*

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

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

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

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

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

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

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

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

3 (T) Pomoću regularizacije može se smanjiti prenaučenost modela. Jedna od tehnika regularizacije je L2-regularizacija, koja se linearne regresije naziva "hrbatna regresija". **Koji je učinak L2-regularizacije kod linearne regresije?**

A Povećava normu vektora težina i tako privlači težine prema ishodištu prostora parametara

B Zakrivljuje područje oko točke minimuma funkcije pogreške u prostoru parametara

C Povećava kondicijski broj matrice dizajna pritezanjem težina na nulu

D Formira dugačak hrbat na površini funkcije pogreške u prostoru parametara

4 (T) Za odabir modela optimalne složenosti tipično se koristi metoda unakrsne provjere. Neka $\boldsymbol{\alpha}$ i $\boldsymbol{\theta}$ označavaju hiperparametre odnosno parametre nekog modela te neka je $\mathcal{A}$ skup razmatranih hiperparametara. Neka su $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ i $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ pogreške hipoteze $h(\mathbf{x};\boldsymbol{\theta})$ na skupu za učenje odnosno treniranje. **Kako funkcionira metoda unakrsne provjere?**

A Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ odabire se $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

B Za svaki $\boldsymbol{\theta}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ odabire se $\boldsymbol{\alpha} \in \mathcal{A}$ koji minimizira $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$

C Za svaki $\boldsymbol{\theta}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{test}})$ po $\boldsymbol{\alpha} \in \mathcal{A}$

D Za svaki $\boldsymbol{\alpha} \in \mathcal{A}$ provodi se minimizacija $E(\boldsymbol{\theta}|\mathcal{D}_{\text{train}})$ po $\boldsymbol{\theta}$

Linearni klasifikacijski modeli *(3 pitanja)*

5 (P) Treniramo model logističke regresije. Skup označenih primjera razdijelili smo na skup za učenje i skup za provjeru. Ako na skupu za učenje treniramo neregularizirani model ($\lambda = 0$), za vektor težina dobivamo $\mathbf{w}_1^*$. Ako isti taj model treniramo na skupu za provjeru, za vektor težina dobivamo $\mathbf{w}_2^* \approx 10 \cdot \mathbf{w}_1^*$. Na temelju ovoga želimo procijeniti vrijednost regularizacijskog faktora λ za koji bi model, naučen na skupu za učenje, najbolje generalizirao. **Koju vrijednost za λ trebamo odabrati?**

A $\lambda = 10$ B $\lambda = 0$ C $\lambda = 25$ D $\lambda = 5$

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

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

Na ovom skupu treniramo perceptron. Pritom koristimo funkciju preslikavanja u šesterodimenzijski prostor značajki, definiranu 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 $\mathbf{w} = (1,0,-1,2,3,0)$. **Koliko iznosi empirijska pogreška perceptrona na skupu za učenje prije početka treniranja (dakle, s početnim težinama)?**

A 40 B 18 C 96 D 100

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 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 511 C 506 D 261

Jezgrene i neparametarske metode *(4 pitanja)*

8 (N) Treniramo SVM s Gaussovom jezgrenom funkcijom. Model treniramo na skupu od $N = 4$ označenih primjera. Vektor oznaka je $\mathbf{y} = (+1,-1,+1,+1)$. Jezgrena matrica je sljedeća:

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

Treniranjem uz $C = 10$ za optimalan vektor dualnih parametara dobili smo $\boldsymbol{\alpha}^* = (9.654, 10, 0.346, 0)$. Za te parametre regularizirana pogreška na skupu za učenje iznosi $E(\boldsymbol{\alpha}^*|\mathcal{D}) = E(\mathbf{w}^*|\mathcal{D}) = 1.724$ te vrijedi $w_0^* = 1.376$. **Koliko iznosi širina margine za ovaj model?**

A 0.409 B 0.426 C 0.640 D 0.665

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

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

Na ovom skupu treniramo jezgreni stroj dimenzije $m = 2$ s Gaussovim baznim funkcijama, koje mjere sličnost između primjera. Za model koristimo logističku regresiju. Središta baznih funkcija su primjeri $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(4)}$. Preciznost jezgre odabrana je tako da je primjer $\mathbf{x}^{(3)}$ u prostoru značajki preslikan u vektor $\boldsymbol{\phi}(\mathbf{x}^{(3)}) = (1, 0.1, 0.2)$. Neka je vektor parametara modela $\mathbf{w}$ inicijalno postavljen na $(w_0, w_1, w_2) = (0.2, 1, -1)$. **Koliko iznosi točnost tako inicijaliziranog modela na skupu $\mathcal{D}$?**

A 0 B 1/4 C 1/2 D 3/4

10 (T) Stroj potpornih vektora (SVM) može se koristiti kod linearno neodvojivih problema. Pretpostavite linearnu neodvojivost i razmotrite SVM za slučaj tvrde margine i meke margine. **Koji od sljedećih slučajeva će rezultirati najužom marginom?**

- A Meka margina s Gaussovom jezgrenom funkcijom i malim vrijednostima za C i γ
- B Tvrda margina s linearnom jezgrom i velikom vrijednosti za C
- C Tvrda margina s Gaussovom jezgrenom funkcijom i velikom vrijednosti za γ
- D Meka margina s linearnom jezgrom i malom vrijednosti za C

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

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

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

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

Procjena parametara i Bayesov klasifikator *(3 pitanja)*

12 (T) U nekim situacijama umjesto naivnog Bayesovog klasifikatora koristimo polunaivan Bayesov klasifikator. Pretpostavite da podatci dolaze iz distribucije u kojoj vrijedi uvjetna nezavisnost značajki x_j uz opaženu varijablu klase y. **Što će se dogoditi ako na skupu za učenje iz takve distribucije treniramo polunaivan Bayesov klasifikator**?

- A Ako u skupu ima šuma, polunaivan model će očekivano imati manju ispitnu pogrešku od naivnog modela
- B Ako u skupu nema šuma, naivan model će očekivano imati veću ispitnu pogrešku od polunaivnog modela
- C Ako u skupu nema šuma, polunaivan model će imati jednaku pogrešku učenja kao i naivan model
- D Ako u skupu nema mnogo šuma, naivan model će imati manju pogrešku učenja od polunaivnog modela

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

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

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

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

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

A -3.84 B -3.03 C -4.13 D -2.75

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

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

Na temelju beta-Bernoullijevog modela na skupu $\mathcal{D}$ računamo MAP procjenu parametra μ Bernoulijeve distribucije. MLE procjena za isti parametar na skupu $\mathcal{D}$ iznosi 0.3. MAP i MLE procjene mogu se poklopiti i onda kada ne koristimo uniformnu apriornu razdiobu. **Uz koje parametre neuniformne beta-distribucije će MLE i MAP procjene biti identične?**

A $\alpha = 4, \beta = 8$ B $\alpha = 2, \beta = 5$ C $\alpha = 2, \beta = 10$ D $\alpha = 5, \beta = 7$

Probabilistički grafički modeli *(3 pitanja)*

15 (P) Bayesovom mrežom modeliramo zajedničku distribuciju šest varijabli: u, v, w, x, y, z. Razmatramo tri modela, $\mathcal{H}_1$, $\mathcal{H}_2$ i $\mathcal{H}_3$, koji odgovaraju sljedećim faktorizacijama:

$$\mathcal{H}_1 : P(u)P(v)P(w|u)P(x|u,v,w)P(y|x,w)p(z|y)$$
$$\mathcal{H}_2 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w,z)p(z)$$
$$\mathcal{H}_3 : P(u)P(v)P(w)P(x|u,v,w)P(y|x,w)p(z|y)$$

Modele učimo na skupu $\mathcal{D}_u$ i ispitujemo na skupu D_i. U oba skupa podataka približno vrijede sljedeće (ne)zavisnosti između parova varijabli: $v \perp w$, $u \perp y|x$ i $w \not\perp z$. Nakon treniranja modela, računamo empirijske pogreške $E(h|\mathcal{D}_u)$ i $E(h|\mathcal{D}_i)$. **Koji su očekivani odnosi vrijednosti empirijskih pogrešaka naučenih modela?**

A $E(h_3|\mathcal{D}_u) < E(h_2|\mathcal{D}_u)$, $E(h_1|\mathcal{D}_i) < E(h_3|\mathcal{D}_i)$

B $E(h_1|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

C $E(h_2|\mathcal{D}_u) < E(h_1|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_1|\mathcal{D}_i)$

D $E(h_2|\mathcal{D}_u) < E(h_3|\mathcal{D}_u)$, $E(h_3|\mathcal{D}_i) < E(h_2|\mathcal{D}_i)$

16 (T) Za izvođenje probabilističkih upita nad PGM-ovima potrebno je uzorkovati iz uvjetne distribucije $P(\mathbf{x}_q|\mathbf{x}_o)$. Naivan pristup bio bi fiksirati vrijednosti svih varijabli $\mathbf{x}_o$ na opažene vrijednosti, a za sve ostale varijable primijeniti unaprijedno uzorkovanje. **Zbog čega takvo uzorkovanje općenito nije dobro?**

A Ako čvorovi iz $\mathbf{x}_o$ nemaju roditelje, uzorak će biti pristran

B Postupak nije primjenjiv ako čvorovi iz $\mathbf{x}_q$ imaju roditelje

C Uzorkujemo iz distribucije koja je različita od $P(\mathbf{x}_q|\mathbf{x}_o)$

D Ako je vjerojatnost $P(\mathbf{x}_o)$ mala, uzorak će biti malen

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

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

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

A 0.1274 B 0.0786 C 0.0490 D 0.1877

Grupiranje *(3 pitanja)*

18 (N) Algoritmom K-sredina grupiramo $N = 1000$ primjera. U tom skupu nalazi se i uzorak od 11 primjera označenih oznakama $\mathcal{Y} = \{1,2,3,4\}$. Međutim, nismo sigurni hoće li grupiranje u četiri grupe doista dati optimalne rezultate, pa isprobavamo grupiranje sa $K = 3$ i $K = 4$ grupe. Rezultati su sljedeći:

$$K = 3 : \quad \big\{\{1,2,4,4\},\{2,3,3\},\{1,1,3,4\}\big\}$$
$$K = 4 : \quad \big\{\{1,2,2,4,4\},\{3,3\},\{1,1\},\{3,4\}\big\}$$

gdje podskupovi odgovaraju grupama, a brojke oznakama primjera. Izračunajte Randov indeks za oba ova grupiranja. **Koliko je Randov indeks za $K = 4$ veći od Randovog indeksa za $K = 3$?**

A 0.0727 B 0.0164 C 0.0945 D 0.0682

19 (T) Prije primjene algoritma maksimizacije očekivanja na model GMM, model je potrebno proširiti latentnim varijablama $\mathbf{z}^{(i)}$, i to jednom takvom varijablom za svaki primjer $\mathbf{x}^{(i)}$. Neka je slučajna varijabla $z_k^{(i)}$ k-ta komponenta slučajnog vektora $\mathbf{z}^{(i)}$. **Koju informaciju kodira varijabla $z_k^{(i)}$?**

A Vjerojatnost da primjer $\mathbf{x}^{(i)}$ pripada grupi k

B Pripadnost primjera $\mathbf{x}^{(i)}$ grupi k

C Vjerojatnost primjera $\mathbf{x}^{(i)}$, ako je njegova grupa k

D Odgovornost primjera $\mathbf{x}^{(i)}$ za grupu k

20 (P) Podatke grupiramo algoritmom hijerarhijskog grupiranja (HAC) s prosječnom povezanošću. U koraku $N-2$ podatci su grupirani u tri grupe (G_1, G_2 i G_3), a pripadna matrica udaljenosti je

$$\begin{array}{c} \\ G_1 \\ G_2 \\ G_3 \end{array} \begin{array}{c} \begin{array}{ccc} G_1 & G_2 & G_3 \end{array} \\ \begin{pmatrix} 0 & 0.8 & 0.2 \\ 0.8 & 0 & 0.3 \\ 0.2 & 0.3 & 0 \end{pmatrix} \end{array}$$

Grupa G_3 sadrži samo jedan primjer, dok se grupe G_1 i G_2 mogu prikazati dendrogramima koji odgovaraju savršenim binarnim stablima (svi unutarnji čvorovi imaju dvoje djece i svi su listovi na istoj dubini). Pritom je stablo za grupu G_1 dubine 3, a stablo za grupu G_2 je dubine 2. **Na kojoj će se udaljenosti provesti spajanje posljednjih dviju grupa?**

A 0.411 B 0.878 C 0.744 D 0.589

Vrednovanje modela *(2 pitanja)*

21 (P) Binarni klasifikator s vjerojatnosnim izlazom vrednovan je pomoću krivulje ROC. Ispitni skup sadrži dva puta više negativnih primjera od pozitivnih. Ispitane su četiri vrijednosti klasifikacijskog praga te je za te vrijednosti izračunat FPR (stopa lažnog alarma) i TPR (odziv). FPR vrijednosti su $0, 0.1, 0.2, 1$, a njima odgovarajuće TPR vrijednosti su $0, 0.2, 0.8, 1$. Međutim, sada za ovaj klasifikator želimo izračunati prosječnu preciznost, pa dobivenu krivulju ROC trebamo pretvoriti u krivulju preciznost-odziv. Pretvorite krivulju ROC u krivulju preciznost-odziv, linearno interpolirajući između izmjerenih točaka. Pritom za $R=0$ uzmite $P=1$. **Koliko iznosi prosječna preciznost ovog klasifikatora?**

A 0.66 B 0.56 C 0.60 D 0.50

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):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 5 & 13 & 1 \\ 6 & 15 & 4 \\ 14 & 2 & 46 \end{pmatrix} \end{array}$$

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 0.015 B 0.094 C 0.068 D 0.040

Grupa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
A	A	C	B	C	C	A	B	D	D	B	D	A	B	D	B	B	A	D	D	C	D	D
B	A	D	C	A	C	A	B	B	A	C	C	A	B	C	B	D	B	B	A	A	C	B
C	B	A	B	D	C	A	C	C	D	D	A	C	B	A	B	A	A	A	B	A	B	B
D	C	B	D	B	C	A	D	C	B	A	C	A	D	B	A	A	C	D	D	A	A	A
E	A	D	A	D	C	D	D	B	A	C	B	A	C	B	C	C	C	B	B	C	C	D
F	C	A	C	B	C	C	C	D	A	D	B	D	A	D	B	C	C	B	A	C	D	C
G	A	C	A	B	C	A	D	A	C	A	A	D	C	A	A	B	A	B	A	A	B	B
H	D	D	B	D	B	A	B	B	C	C	B	C	B	A	B	C	C	A	B	C	C	B