

Pismeni ispit iz Strojnog učenja (ak. god. 2020./2021.)

*Ispit sadrži **24 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Za linearan model u $\mathcal{X} = \{0,1\}^3$ zadan je sljedeći skup primjera za učenje:

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

Optimizacijski postupak klasifikatora funkcionira tako da minimizira empirijsku pogrešku, definiranu kao očekivanje funkcije gubitka 0-1, i postupak u tome uvijek uspijeva. Želimo znati koju bi klasu ovaj klasifikator dodijelio primjeru $\mathbf{x} = (1,1,1)$. **Možemo li, na temelju iznesenih informacija, odrediti klasifikaciju dotičnog primjera i što nam to govori o induktivnoj pristranosti ovog algoritma**?

A Možemo, klasifikacija je $y = 1$, i ovaj klasifikator ima definiranu induktivnu pristranost pomoću koje može jednoznačno odrediti klasifikaciju svakog primjera

B Ne možemo, jer nije definirana induktivna pristranost preferencijom, pa činjenica da je model linearan nije dovoljan skup pretpostavki da bismo jednoznačno odredili klasifikaciju svih novih primjera

C Možemo, klasifikacija je $y = 1$, premda dane informacije nisu dovoljne za definicju induktivne pristranosti, što se vidi iz toga da je za ovaj skup primjera prostor inačica veći od jedan

D Možemo, $y = 1$, jer klasifikator ima induktivnu pristranost jezikom (linearan model) i preferencijom (primjeri za koje je $h(x) \geq 0$ klasificiraju se pozitivno)

2 (P) U ulaznom prostoru $\mathcal{X} = \{0,1\}^3$ definiramo sljedeći klasifikacijski model:

$$h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \geq 0\}$$

Koja je dimenzija prostora parametara te koliko različitih hipoteza postoji u ovom modelu?

A Dimenzija prostora parametara je 4, a hipoteza ima manje od 256

B Dimenzija prostora parametara je 256, a hipoteza ima 14

C Dimenzija prostora parametara je 4, a hipoteza ima beskonačno mnogo

D Dimenzija prostora parametara i broj hipoteza su beskonačni

3 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Koja je razlika između induktivnih pristranosti regularizirane i neregularizirane linearne regresije?**

A Algoritmi se ne razlikuju po pristranosti preferencijom budući da koriste istu funkciju gubitka (kvadratni gubitak), međutim regularizirana regresija ima jaču induktivnu pristranost jezika od regularizirane regresije budući da prvi model uključuje drugi model

B Za razliku od neregularizirane regresije, regularizirana regresija preferira jednostavnije hipoteze, međutim pristranosti su im identične jer su oba algoritma definirana kao linearna kombinacija značajki i težina te oba koriste identičan optimizacijski postupak (pseudoinverz matrice dizajna)

C Algoritmi imaju različite pristranosti, i to različitu pristranost preferencije jer regularizirana regresija preferira jednostavnije hipoteze, a onda i različitu pristranost jezika jer je model neregularizirane regresije nadskup modela regularizirane regresije

D Oba algoritma imaju isti model, definiran kao linearnu kombinaciju značajki i težina, pa dakle imaju istu pristranost jezika, ali se razlikuju u pristranosti preferencije jer imaju različito definiranu empirijsku pogrešku (osim ako je regularizacijski faktor jednak nuli)

4 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 79 B 48 C 86 D 92

Cjelina 2: Logistička regresija *(4 pitanja)*

5 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, -1, -3)$ s oznakom $\mathbf{y} = (0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 6.00 B 8.00 C 4.02 D 12.02

6 (T) Na predavanjima smo za klasifikaciju pokušali upotrijebiti algoritam regresije. Zaključili smo da to ne funkcionira, tj. da algoritam linearne regesije jednostavno nije klasifikacijski algoritam. **Koje bismo minimalne preinake trebale učiniti u algoritmu linearne regresije, a da on dobro funkcionira kao klasifikacijski algoritam?**

A promijeniti model i optimizacijski postupak

B promijeniti optimizacijski postupak

C promijeniti model, funkciju gubitka i optimizacijski postupak

D promijeniti funkciju gubitka

7 (P) Razvijamo sustav za automatsku klasifikaciju novinskih članaka u jednu od pet kategorija. Tih pet kategorija su "sport", "politika", "kriminal", "znanost" i "lifestyle". Najveća razlika u veličini klasa je između kategorija "politika" i "znanost". Očekivano, u kategoriji "politika" ima najviše članaka, dok ih u kategoriji "znanost" ima 5x manje, što je u redu jer to ionako nitko ne čita. Svaki novinski članak prikazujemo kao vektor riječi, gdje su komponente vektora broj pojavljivanja pojedine riječi. Problem rješavamo algoritmom SVM, i to u primarnoj formulaciji. Budući da je SVM binaran klasifikator, odlučili smo primijeniti shemu OVR ili shemu OVO za dekompoziciju višeklasnog klasifikacijskog problema u skup binarnih klasifikacijskih problema. **Što možemo očekivati?**

A SVM+OVR će imati 2x manje značajki od SVM+OVO, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

B SVM+OVO će imati 5x puta manje značajki od SVM+OVR, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

C SVM+OVO će imati 2x puta manje značajki od SVM+OVR, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

D SVM+OVR će imati 5x manje značajki od SVM+OVO, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

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

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

U jednoj iteraciji treniranja modela vektor parametara jednak je

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

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

A 0.70 B 4.00 C 2.48 D 1.28

Cjelina 3: Jezgrene i neparametarske metode *(5 pitanja)*

9 (T) Algoritam SVM može biti parametarski i neparametarski, ovisno o tome provodimo li optimizaciju u primarnoj ili dualnoj formulaciji. U oba slučaja preferiramo da je model rijedak, tj. da je nakon treniranja što više parametara postavljeno na nulu. **Kako rijetkost modela ovisi o hiperparametru C?**

A Što je C manji, to je neparametarski model rjeđi, ali to nema utjecaja na rijetkost parametarskog modela jer on nema potporne vektore

B Što je C manji, to je neparametarski model rjeđi, a također je to rjeđi i parametarski model jer λ raste

C Što je C veći, to je neparametarski model manje rijedak, dok parametarski model nije rijedak jer ima L_2-regularizaciju a ne L_1-regularizaciju

D Što je C veći, to je neparametarski model manje rijedak, dok je parametarski to rjeđi jer λ pada

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

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

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

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

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

A 0.0024 B 0.0045 C 0.0089 D 0.0013

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

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

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

A +1.434 B +0.947 C −2.330 D −0.676

12 (T) Kod izvoda algoritma SVM s tvrdom marginom, pretpostavili smo da za primjere $\mathbf{x} \in \mathbb{R}^n$ vrijedi sljedeći uvjet linearne odvojivosti:

$$\forall (\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}.\ y^{(i)} h(\mathbf{x}^{(i)}) \geq 0$$

Koliko hipoteza zadovoljava ovaj uvjet, i kako algoritam SVM odabire jednu od njih?

A Uvjet zadovoljava beskonačno mnogo hipoteza, a SVM odabire onu jednu koja minimizira kvadrat vektora težina te koja ispravno klasificira sve primjere, uz uvjet da $h(\mathbf{x})$ nije u intervalu $(-1, +1)$

B Uvjet zadovoljava beskonačno mnogo hipoteza, međutim samo za jednu vrijedi $yh(\mathbf{x}) = 1$ za najbliže primjere, i to je hipoteza koju odabire SVM

C Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, a SVM između njih odabire onu jednu koja minimizira kvadrat vektora težina

D Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, no one se razlikuju samo po faktoru koji množi težine $(\mathbf{w}, w_0)$, pa SVM odabire onu jednu za koju vrijedi $yh(\mathbf{x}) \geq 1$ za sve primjere

13 (N) Raspolažemo sljedećim skupom označenih primjera u trodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((1, -3, 6), -1), ((4, -4, 4), -1), ((-2, 4, 1), +1))\}$$

Na ovom skupu primjera treniramo model SVM-a s linearnom jezgrenom funkcijom i sa $C = 0.001$. Postupak treniranja algoritmom SMO završio je s vektorom Lagrangeovih koeficijenata $\boldsymbol{\alpha} = (0.001, 0, 0.001)$. Iz ovoga se da izračunati da vrijedi $w_0 = -0.943$. Umjesto algoritma SMO, za optimizaciju smo mogli upotrijebiti gradijentni spust i optimirati težine u primarnoj formulaciji problema. U tom slučaju koristili bismo empirijsku pogrešku SVM-a definiranu kao L_2-regularizirani gubitak zglobnice. Međutim, tu pogrešku možemo izračunati i naknadno, nakon što smo naučili model. **Koliko iznosi empirijska pogreška ovog SVM-a na skupu primjera $\mathcal{D}$?**

A 1.135 B 1.585 C 1.650 D 1.959

Cjelina 4: Probabilistički modeli *(6 pitanja)*

14 (N) Na sljedećem skupu treniramo naivan Bayesov model za binarnu klasifikaciju *"Skupo ljetovanje na Jadranu"*:

i	Mjesto	Otok	Smještaj	Prijevoz	$y^{(i)}$
1	Kvarner	da	privatni	auto	1
2	Kvarner	ne	kamp	bus	0
3	Dalmacija	da	hotel	avion	1
4	Dalmacija	ne	privatni	avion	0
5	Istra	da	kamp	auto	0
6	Istra	ne	kamp	bus	1
7	Dalmacija	da	hotel	auto	1

Procjene radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija primjera $\mathbf{x} = (\text{Istra}, \text{ne}, \text{kamp}, \text{bus})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.685 B 0.318 C 0.706 D 0.237

15 (P) Bayesova mreža ima pet varijabli, od kojih su v, w i z binarne, a x i y ternarne varijable. Topološki uređaj varijabli neka je v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće marginalne i uvjetne nezavisnosti:

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

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

A 27 B 10 C 25 D 22

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

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

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

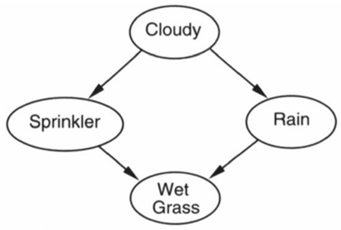
$\mathcal{H}_3$: Između značajki postoje korelacije, ali se one ne razlikuju između klasa

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

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

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

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



C	$P(C)$
0	0.5
1	0.5

S	C	$P(S\|C)$
0	0	0.5
0	1	0.9
1	0	0.5
1	1	0.1

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

W	R	S	$P(W\|R,S)$
0	0	0	1.0
0	0	1	0.9
0	1	0	0.1
0	1	1	0.01
1	0	0	0.0
1	0	1	0.1
1	1	0	0.9
1	1	1	0.99

Izračunajte aposteriornu vjerojatnost da pada kiša ako je trava mokra i nije oblačno.

A 0.112 B 0.709 C 0.825 D 0.491

18 (P) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x,y,z) = P(x)P(y)P(z|x,y)$. Sve varijable su binarne. Vrijedi $P(x=1) = 0.2$ i $P(y=1) = 0.3$. Tablica uvjetne vjerojatnosti za čvor z je sljedeća:

z	x	y	$p(z\|x,y)$	z	x	y	$p(z\|x,y)$
0	0	0	0.1	1	0	0	0.9
0	0	1	0.2	1	0	1	0.8
0	1	0	0.5	1	1	0	0.5
0	1	1	0.9	1	1	1	0.1

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(y|x=0, z=1)$. Uzorkovanje smo ponovili ukupno $N = 500$ puta. **Koliki je očekivani broj vektora koje ćemo pri uzorkovanju morati odbaciti?**

A 152 B 180 C 89 D 404

19 (P) Bacanje igraće kocke modeliramo kategoričkom varijablom $\mathbf{x}$, gdje indikatorske varijable $x_1, \dots, x_6$ odgovaraju vrijednosti koju dobivamo bacanjem kocke. Za procjenu parametara $\boldsymbol{\mu}$ kategoričke distribucije koristimo MAP-procjenitelj s Dirichletovom distribucijom za apriornu gustoću vjerojatnosti. U stvarnosti, kocka je modificirana tako da će nešto češće davati šesticu, odnosno realizaciju $x_6 = 1$, međutim mi to ne znamo. Naprotiv, na temelju manjeg broja opažanja ranijih bacanja kocke utvrdili smo da je kocka najčešće davala peticu, no svjesni smo da je naša procjena ograničena na manji broj opažanja. **Uz koje parametre Dirichletove distribucije će naša procjena za $\boldsymbol{\mu}$ biti najbliža stvarnoj vrijednosti tih parametara?**

A $\boldsymbol{\alpha} = (1,1,1,1,1,1)$ B $\boldsymbol{\alpha} = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ C $\boldsymbol{\alpha} = (2,2,2,2,2,2)$ D $\boldsymbol{\alpha} = (1,1,1,1,3,1)$

Cjelina 5: Grupiranje *(3 pitanja)*

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

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

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

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

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

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

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

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.62 B 0.49 C 0.58 D 0.53

22 Za procjenu parametara modela GMM tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). To je iterativan optimizacijski algoritam. **Pod kojim uvjetima EM-algoritam (primijenjen na model GMM) konvergira, i kamo?**

A Krenuvši od nekih početnih parametara, algoritam uvijek konvergira do parametara koji maksimiziraju očekivanje log-izglednosti, međutim to ne moraju biti parametri koji maksimiziraju vjerojatnost podataka

B Algoritam konvergira samo ako su primjeri u ulaznom prostoru sferični, ako su zavisnosti između značajki linearne, i ako nema multikolinearnosti, jer u protivnom zavisnosti nije moguće modelirati kovarijacijskom matricom

C Algoritm uvijek konvergira, i to do točke u prostoru parametara koja maksimizira log-izglednost parametara, no brzina konvergencije ovisi o toma kako su inicijalizirani parametri

D Algoritam uvijek konvergira, međutim globalni maksimum log-izglednosti parametara doseže samo ako je broj grupa postavljen na pravi broj grupa ili tako da je broj grupa jednak broju primjera

Cjelina 6: Vrednovanje modela *(2 pitanja)*

23 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 44640 B 49600 C 35721 D 69201

24 (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):

$$
\begin{array}{c c}
 & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\
\begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 5 & 13 & 1 \\ 6 & 5 & 4 \\ 4 & 2 & 16 \end{pmatrix}
\end{array}
$$

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

A 0.023 B 0.091 C 0.114 D 0.059

Pismeni ispit iz Strojnog učenja (ak. god. 2020./2021.)

*Ispit sadrži **24 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Za linearan model u $\mathcal{X} = \{0,1\}^3$ zadan je sljedeći skup primjera za učenje:

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

Optimizacijski postupak klasifikatora funkcionira tako da minimizira empirijsku pogrešku, definiranu kao očekivanje funkcije gubitka 0-1, i postupak u tome uvijek uspijeva. Želimo znati koju bi klasu ovaj klasifikator dodijelio primjeru $\mathbf{x} = (1,1,1)$. **Možemo li, na temelju iznesenih informacija, odrediti klasifikaciju dotičnog primjera i što nam to govori o induktivnoj pristranosti ovog algoritma**?

A Možemo, klasifikacija je $y = 1$, i ovaj klasifikator ima definiranu induktivnu pristranost pomoću koje može jednoznačno odrediti klasifikaciju svakog primjera

B Možemo, $y = 1$, jer klasifikator ima induktivnu pristranost jezikom (linearan model) i preferencijom (primjeri za koje je $h(x) \geq 0$ klasificiraju se pozitivno)

C Možemo, klasifikacija je $y = 1$, premda dane informacije nisu dovoljne za definicju induktivne pristranosti, što se vidi iz toga da je za ovaj skup primjera prostor inačica veći od jedan

D Ne možemo, jer nije definirana induktivna pristranost preferencijom, pa činjenica da je model linearan nije dovoljan skup pretpostavki da bismo jednoznačno odredili klasifikaciju svih novih primjera

2 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Koja je razlika između induktivnih pristranosti regularizirane i neregularizirane linearne regresije?**

A Za razliku od neregularizirane regresije, regularizirana regresija preferira jednostavnije hipoteze, međutim pristranosti su im identične jer su oba algoritma definirana kao linearna kombinacija značajki i težina te oba koriste identičan optimizacijski postupak (pseudoinverz matrice dizajna)

B Algoritmi imaju različite pristranosti, i to različitu pristranost preferencije jer regularizirana regresija preferira jednostavnije hipoteze, a onda i različitu pristranost jezika jer je model neregularizirane regresije nadskup modela regularizirane regresije

C Algoritmi se ne razlikuju po pristranosti preferencijom budući da koriste istu funkciju gubitka (kvadratni gubitak), međutim regularizirana regresija ima jaču induktivnu pristranost jezika od regularizirane regresije budući da prvi model uključuje drugi model

D Oba algoritma imaju isti model, definiran kao linearnu kombinaciju značajki i težina, pa dakle imaju istu pristranost jezika, ali se razlikuju u pristranosti preferencije jer imaju različito definiranu empirijsku pogrešku (osim ako je regularizacijski faktor jednak nuli)

3 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 92 B 86 C 48 D 79

4 (P) U ulaznom prostoru $\mathcal{X} = \{0, 1\}^3$ definiramo sljedeći klasifikacijski model:

$$h(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \geq 0\}$$

Koja je dimenzija prostora parametara te koliko različitih hipoteza postoji u ovom modelu?

A Dimenzija prostora parametara je 4, a hipoteza ima beskonačno mnogo

B Dimenzija prostora parametara je 256, a hipoteza ima 14

C Dimenzija prostora parametara je 4, a hipoteza ima manje od 256

D Dimenzija prostora parametara i broj hipoteza su beskonačni

Cjelina 2: Logistička regresija *(4 pitanja)*

5 (T) Na predavanjima smo za klasifikaciju pokušali upotrijebiti algoritam regresije. Zaključili smo da to ne funkcionira, tj. da algoritam linearne regesije jednostavno nije klasifikacijski algoritam. **Koje bismo minimalne preinake trebale učiniti u algoritmu linearne regresije, a da on dobro funkcionira kao klasifikacijski algoritam?**

A promijeniti model i optimizacijski postupak

B promijeniti optimizacijski postupak

C promijeniti model i funkciju gubitka

D promijeniti model, funkciju gubitka i optimizacijski postupak

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

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

U jednoj iteraciji treniranja modela vektor parametara jednak je

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

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

A 1.28 B 0.70 C 2.48 D 4.00

7 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, -1, -3)$ s oznakom $\mathbf{y} = (0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 8.00 B 6.00 C 4.02 D 12.02

8 (P) Razvijamo sustav za automatsku klasifikaciju novinskih članaka u jednu od pet kategorija. Tih pet kategorija su "sport", "politika", "kriminal", "znanost" i "lifestyle". Najveća razlika u veličini klasa je između kategorija "politika" i "znanost". Očekivano, u kategoriji "politika" ima najviše članaka, dok ih u kategoriji "znanost" ima 5x manje, što je u redu jer to ionako nitko ne čita. Svaki novinski članak prikazujemo kao vektor riječi, gdje su komponente vektora broj pojavljivanja pojedine riječi. Problem rješavamo algoritmom SVM, i to u primarnoj formulaciji. Budući da je SVM binaran klasifikator, odlučili smo primijeniti shemu OVR ili shemu OVO za dekompoziciju višeklasnog klasifikacijskog problema u skup binarnih klasifikacijskih problema. **Što možemo očekivati?**

A SVM+OVO će imati 5x puta manje značajki od SVM+OVR, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

B SVM+OVR će imati 5x manje značajki od SVM+OVO, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

C SVM+OVR će imati 2x manje značajki od SVM+OVO, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

D SVM+OVO će imati 2x puta manje značajki od SVM+OVR, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

Cjelina 3: Jezgrene i neparametarske metode *(5 pitanja)*

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

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

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

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

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

A 0.0024 B 0.0045 C 0.0089 D 0.0013

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

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

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

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

11 (N) Raspolažemo sljedećim skupom označenih primjera u trodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((1, -3, 6), -1), ((4, -4, 4), -1), ((-2, 4, 1), +1))\}$$

Na ovom skupu primjera treniramo model SVM-a s linearnom jezgrenom funkcijom i sa $C = 0.01$. Postupak treniranja algoritmom SMO završio je s vektorom Lagrangeovih koeficijenata $\boldsymbol{\alpha} = (0.01, 0, 0.01)$. Iz ovoga se da izračunati da vrijedi $w_0 = -0.43$. Umjesto algoritma SMO, za optimizaciju smo mogli upotrijebiti gradijentni spust i optimirati težine u primarnoj formulaciji problema. U tom slučaju koristili bismo empirijsku pogrešku SVM-a definiranu kao L_2-regularizirani gubitak zglobnice. Međutim, tu pogrešku možemo izračunati i naknadno, nakon što smo naučili model. **Koliko iznosi empirijska pogreška ovog SVM-a na skupu primjera $\mathcal{D}$?**

A 1.650 B 33.135 C 1.135 D 1.585

12 (T) Kod izvoda algoritma SVM s tvrdom marginom, pretpostavili smo da za primjere $\mathbf{x} \in \mathbb{R}^n$ vrijedi sljedeći uvjet linearne odvojivosti:

$$\forall(\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}.\ y^{(i)} h(\mathbf{x}^{(i)}) \geq 0$$

Koliko hipoteza zadovoljava ovaj uvjet, i kako algoritam SVM odabire jednu od njih?

- A Uvjet zadovoljava beskonačno mnogo hipoteza, međutim samo za jednu vrijedi $yh(\mathbf{x}) = 1$ za najbliže primjere, i to je hipoteza koju odabire SVM
- B Uvjet zadovoljava beskonačno mnogo hipoteza, a SVM odabire onu jednu koja minimizira kvadrat vektora težina te koja ispravno klasificira sve primjere, uz uvjet da $h(\mathbf{x})$ nije u intervalu $(-1, +1)$
- C Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, no one se razlikuju samo po faktoru koji množi težine $(\mathbf{w}, w_0)$, pa SVM odabire onu jednu za koju vrijedi $yh(\mathbf{x}) \geq 1$ za sve primjere
- D Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, a SVM između njih odabire onu jednu koja minimizira kvadrat vektora težina

13 (T) Algoritam SVM može biti parametarski i neparametarski, ovisno o tome provodimo li optimizaciju u primarnoj ili dualnoj formulaciji. U oba slučaja preferiramo da je model rijedak, tj. da je nakon treniranja što više parametara postavljeno na nulu. **Kako rijetkost modela ovisi o hiperparametru C?**

- A Što je C veći, to je neparametarski model manje rijedak, dok parametarski model nije rijedak jer ima L_2-regularizaciju a ne L_1-regularizaciju
- B Što je C veći, to je neparametarski model manje rijedak, dok je parametarski to rjeđi jer λ pada
- C Što je C manji, to je neparametarski model rjeđi, ali to nema utjecaja na rijetkost parametarskog modela jer on nema potporne vektore
- D Što je C manji, to je neparametarski model rjeđi, a također je to rjeđi i parametarski model jer λ raste

Cjelina 4: Probabilistički modeli *(6 pitanja)*

14 (P) Bacanje igraće kocke modeliramo kategoričkom varijablom $\mathbf{x}$, gdje indikatorske varijable $x_1, \ldots, x_6$ odgovaraju vrijednosti koju dobivamo bacanjem kocke. Za procjenu parametara $\boldsymbol{\mu}$ kategoričke distribucije koristimo MAP-procjenitelj s Dirichletovom distribucijom za apriornu gustoću vjerojatnosti. U stvarnosti, kocka je modificirana tako da će nešto češće davati šesticu, odnosno realizaciju $x_6 = 1$, međutim mi to ne znamo. Naprotiv, na temelju manjeg broja opažanja ranijih bacanja kocke utvrdili smo da je kocka najčešće davala peticu, no svjesni smo da je naša procjena ograničena na manji broj opažanja. **Uz koje parametre Dirichletove distribucije će naša procjena za $\boldsymbol{\mu}$ biti najbliža stvarnoj vrijednosti tih parametara?**

A $\boldsymbol{\alpha} = (1, 1, 1, 1, 1, 1)$ B $\boldsymbol{\alpha} = (1, 1, 1, 1, 3, 1)$ C $\boldsymbol{\alpha} = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ D $\boldsymbol{\alpha} = (2, 2, 2, 2, 2, 2)$

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

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

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

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

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

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

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

16 (P) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x, y, z) = P(x)P(y)P(z|x, y)$. Sve varijable su binarne. Vrijedi $P(x = 1) = 0.2$ i $P(y = 1) = 0.3$. Tablica uvjetne vjerojatnosti za čvor z je sljedeća:

z	x	y	$p(z\|x,y)$	z	x	y	$p(z\|x,y)$
0	0	0	0.1	1	0	0	0.9
0	0	1	0.2	1	0	1	0.8
0	1	0	0.5	1	1	0	0.5
0	1	1	0.9	1	1	1	0.1

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(y|x=0,z=1)$. Uzorkovanje smo ponovili ukupno $N=500$ puta. **Koliki je očekivani broj vektora koje ćemo pri uzorkovanju morati odbaciti?**

A 152 B 89 C 180 D 404

17 (P) Bayesova mreža ima pet varijabli, od kojih su v, w i z binarne, a x i y ternarne varijable. Topološki uređaj varijabli neka je v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće marginalne i uvjetne nezavisnosti:

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

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

A 25 B 22 C 27 D 10

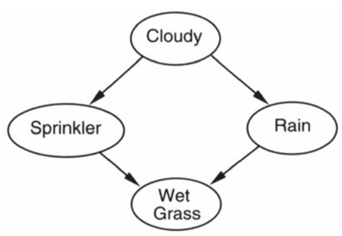
18 (N) Na sljedećem skupu treniramo naivan Bayesov model za binarnu klasifikaciju *"Skupo ljetovanje na Jadranu"*:

i	Mjesto	Otok	Smještaj	Prijevoz	$y^{(i)}$
1	Kvarner	da	privatni	auto	1
2	Kvarner	ne	kamp	bus	0
3	Dalmacija	da	hotel	avion	1
4	Dalmacija	ne	privatni	avion	0
5	Istra	da	kamp	auto	0
6	Istra	ne	kamp	bus	1
7	Dalmacija	da	hotel	auto	1

Procjene radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija primjera $\mathbf{x} = (\text{Dalmacija, ne, kamp, bus})$. **Koliko iznosi aposteriorna vjerojatnost $P(y=1|\mathbf{x})$?**

A 0.318 B 0.685 C 0.706 D 0.237

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



C	$P(C)$
0	0.5
1	0.5

S	C	$P(S\|C)$
0	0	0.5
0	1	0.9
1	0	0.5
1	1	0.1

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

W	R	S	$P(W\|R,S)$
0	0	0	1.0
0	0	1	0.9
0	1	0	0.1
0	1	1	0.01
1	0	0	0.0
1	0	1	0.1
1	1	0	0.9
1	1	1	0.99

Izračunajte aposteriornu vjerojatnost da pada kiša ako je trava mokra i nije oblačno.

A 0.709 B 0.825 C 0.491 D 0.112

Cjelina 5: Grupiranje *(3 pitanja)*

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

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

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.49   B 0.58   C 0.62   D 0.53

21 (N) Algoritmom hijerarhijskog aglomerativnog grupiranja (HAC) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru sličnosti, koja je za naših pet primjera definirana sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

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

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

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

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

22 Za procjenu parametara modela GMM tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). To je iterativan optimizacijski algoritam. **Pod kojim uvjetima EM-algoritam (primijenjen na model GMM) konvergira, i kamo?**

A Algoritam uvijek konvergira, međutim globalni maksimum log-izglednosti parametara doseže samo ako je broj grupa postavljen na pravi broj grupa ili tako da je broj grupa jednak broju primjera

B Krenuvši od nekih početnih parametara, algoritam uvijek konvergira do parametara koji maksimiziraju očekivanje log-izglednosti, međutim to ne moraju biti parametri koji maksimiziraju vjerojatnost podataka

C Algoritam konvergira samo ako su primjeri u ulaznom prostoru sferični, ako su zavisnosti između značajki linearne, i ako nema multikolinearnosti, jer u protivnom zavisnosti nije moguće modelirati kovarijacijskom matricom

D Algoritm uvijek konvergira, i to do točke u prostoru parametara koja maksimizira log-izglednost parametara, no brzina konvergencije ovisi o toma kako su inicijalizirani parametri

Cjelina 6: Vrednovanje modela *(2 pitanja)*

23 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 69201   B 44640   C 35721   D 49600

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

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

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

A 0.023   B 0.059   C 0.114   D 0.091

Pismeni ispit iz Strojnog učenja (ak. god. 2020./2021.)

*Ispit sadrži **24 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Koja je razlika između induktivnih pristranosti regularizirane i neregularizirane linearne regresije?**

A Oba algoritma imaju isti model, definiran kao linearnu kombinaciju značajki i težina, pa dakle imaju istu pristranost jezika, ali se razlikuju u pristranosti preferencije jer imaju različito definiranu empirijsku pogrešku (osim ako je regularizacijski faktor jednak nuli)

B Za razliku od neregularizirane regresije, regularizirana regresija preferira jednostavnije hipoteze, međutim pristranosti su im identične jer su oba algoritma definirana kao linearna kombinacija značajki i težina te oba koriste identičan optimizacijski postupak (pseudoinverz matrice dizajna)

C Algoritmi se ne razlikuju po pristranosti preferencijom budući da koriste istu funkciju gubitka (kvadratni gubitak), međutim regularizirana regresija ima jaču induktivnu pristranost jezika od regularizirane regresije budući da prvi model uključuje drugi model

D Algoritmi imaju različite pristranosti, i to različitu pristranost preferencije jer regularizirana regresija preferira jednostavnije hipoteze, a onda i različitu pristranost jezika jer je model neregularizirane regresije nadskup modela regularizirane regresije

2 (P) Za linearan model u $\mathcal{X} = \{0,1\}^3$ zadan je sljedeći skup primjera za učenje:

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

Optimizacijski postupak klasifikatora funkcionira tako da minimizira empirijsku pogrešku, definiranu kao očekivanje funkcije gubitka 0-1, i postupak u tome uvijek uspijeva. Želimo znati koju bi klasu ovaj klasifikator dodijelio primjeru $\mathbf{x} = (1,1,1)$. **Možemo li, na temelju iznesenih informacija, odrediti klasifikaciju dotičnog primjera i što nam to govori o induktivnoj pristranosti ovog algoritma**?

A Možemo, klasifikacija je $y = 1$, i ovaj klasifikator ima definiranu induktivnu pristranost pomoću koje može jednoznačno odrediti klasifikaciju svakog primjera

B Ne možemo, jer nije definirana induktivna pristranost preferencijom, pa činjenica da je model linearan nije dovoljan skup pretpostavki da bismo jednoznačno odredili klasifikaciju svih novih primjera

C Možemo, klasifikacija je $y = 1$, premda dane informacije nisu dovoljne za definicju induktivne pristranosti, što se vidi iz toga da je za ovaj skup primjera prostor inačica veći od jedan

D Možemo, $y = 1$, jer klasifikator ima induktivnu pristranost jezikom (linearan model) i preferencijom (primjeri za koje je $h(x) \geq 0$ klasificiraju se pozitivno)

3 (P) U ulaznom prostoru $\mathcal{X} = \{0,1\}^3$ definiramo sljedeći klasifikacijski model:

$$h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \geq 0\big\}$$

Koja je dimenzija prostora parametara te koliko različitih hipoteza postoji u ovom modelu?

A Dimenzija prostora parametara je 4, a hipoteza ima beskonačno mnogo

B Dimenzija prostora parametara i broj hipoteza su beskonačni

C Dimenzija prostora parametara je 4, a hipoteza ima manje od 256

D Dimenzija prostora parametara je 256, a hipoteza ima 14

4 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 92 B 48 C 79 D 86

Cjelina 2: Logistička regresija *(4 pitanja)*

5 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, 1, -3)$ s oznakom $\mathbf{y} = (0, 1, 0)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 12.02 B 6.00 C 4.02 D 8.00

6 (T) Na predavanjima smo za klasifikaciju pokušali upotrijebiti algoritam regresije. Zaključili smo da to ne funkcionira, tj. da algoritam linearne regesije jednostavno nije klasifikacijski algoritam. **Koje bismo minimalne preinake trebale učiniti u algoritmu linearne regresije, a da on dobro funkcionira kao klasifikacijski algoritam?**

A promijeniti funkciju gubitka

B promijeniti funkciju gubitka i optimizacijski postupak

C promijeniti optimizacijski postupak

D promijeniti model, funkciju gubitka i optimizacijski postupak

7 (P) Razvijamo sustav za automatsku klasifikaciju novinskih članaka u jednu od pet kategorija. Tih pet kategorija su "sport", "politika", "kriminal", "znanost" i "lifestyle". Najveća razlika u veličini klasa je između kategorija "politika" i "znanost". Očekivano, u kategoriji "politika" ima najviše članaka, dok ih u kategoriji "znanost" ima 5x manje, što je u redu jer to ionako nitko ne čita. Svaki novinski članak prikazujemo kao vektor riječi, gdje su komponente vektora broj pojavljivanja pojedine riječi. Problem rješavamo algoritmom SVM, i to u primarnoj formulaciji. Budući da je SVM binaran klasifikator, odlučili smo primijeniti shemu OVR ili shemu OVO za dekompoziciju višeklasnog klasifikacijskog problema u skup binarnih klasifikacijskih problema. **Što možemo očekivati?**

A SVM+OVR će imati 2x manje značajki od SVM+OVO, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

B SVM+OVO će imati 2x puta manje značajki od SVM+OVR, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

C SVM+OVO će imati 5x puta manje značajki od SVM+OVR, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

D SVM+OVR će imati 5x manje značajki od SVM+OVO, ali bi mogao raditi bolje na člancima iz kategorije "znanost"

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

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

U jednoj iteraciji treniranja modela vektor parametara jednak je

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

Koliko u toj iteraciji iznosi L_2-norma gradijenta gubitka za primjer $(\mathbf{x}, y) = ((1.5, -2), 1)$?

A 0.70 B 1.28 C 2.48 D 4.00

Cjelina 3: Jezgrene i neparametarske metode *(5 pitanja)*

9 (N) Raspolažemo sljedećim skupom označenih primjera u trodimenzijskome ulaznom prostoru:

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

Na ovom skupu primjera treniramo model SVM-a s linearnom jezgrenom funkcijom i sa $C = 0.05$. Postupak treniranja algoritmom SMO završio je s vektorom Lagrangeovih koeficijenata $\boldsymbol{\alpha} = (0.05, 0, 0.05)$. Iz ovoga se da izračunati da vrijedi $w_0 = -0.675$. Umjesto algoritma SMO, za optimizaciju smo mogli upotrijebiti gradijentni spust i optimirati težine u primarnoj formulaciji problema. U tom slučaju koristili bismo empirijsku pogrešku SVM-a definiranu kao L_2-regularizirani gubitak zglobnice. Međutim, tu pogrešku možemo izračunati i naknadno, nakon što smo naučili model. **Koliko iznosi empirijska pogreška ovog SVM-a na skupu primjera $\mathcal{D}$?**

A 1.958 B 1.650 C 1.725 D 1.585

10 (T) Algoritam SVM može biti parametarski i neparametarski, ovisno o tome provodimo li optimizaciju u primarnoj ili dualnoj formulaciji. U oba slučaja preferiramo da je model rijedak, tj. da je nakon treniranja što više parametara postavljeno na nulu. **Kako rijetkost modela ovisi o hiperparametru C?**

A Što je C veći, to je neparametarski model manje rijedak, dok je parametarski to rjeđi jer λ pada

B Što je C manji, to je neparametarski model rjeđi, ali to nema utjecaja na rijetkost parametarskog modela jer on nema potporne vektore

C Što je C veći, to je neparametarski model manje rijedak, dok parametarski model nije rijedak jer ima L_2-regularizaciju a ne L_1-regularizaciju

D Što je C manji, to je neparametarski model rjeđi, a također je to rjeđi i parametarski model jer λ raste

11 (T) Kod izvoda algoritma SVM s tvrdom marginom, pretpostavili smo da za primjere $\mathbf{x} \in \mathbb{R}^n$ vrijedi sljedeći uvjet linearne odvojivosti:

$$\forall (\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}.\ y^{(i)} h(\mathbf{x}^{(i)}) \geq 0$$

Koliko hipoteza zadovoljava ovaj uvjet, i kako algoritam SVM odabire jednu od njih?

A Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, no one se razlikuju samo po faktoru koji množi težine $(\mathbf{w}, w_0)$, pa SVM odabire onu jednu za koju vrijedi $yh(\mathbf{x}) \geq 1$ za sve primjere

B Uvjet zadovoljava beskonačno mnogo hipoteza, a SVM odabire onu jednu koja minimizira kvadrat vektora težina te koja ispravno klasificira sve primjere, uz uvjet da $h(\mathbf{x})$ nije u intervalu $(-1, +1)$

C Uvjet zadovoljava beskonačno mnogo hipoteza, međutim samo za jednu vrijedi $yh(\mathbf{x}) = 1$ za najbliže primjere, i to je hipoteza koju odabire SVM

D Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, a SVM između njih odabire onu jednu koja minimizira kvadrat vektora težina

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

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

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

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

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

A 0.0024   B 0.0013   C 0.0045   D 0.0089

13 (N) Na skupu primjera za učenje iz ulaznog prostora $n = 4$ trenirali smo SVM s polinomijalnom jezgrenom funkcijom $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + 2)^3$. Potporni vektori i njihove oznake su sljedeći:

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

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

A +1.434   B −2.330   C −0.676   D +0.947

Cjelina 4: Probabilistički modeli *(6 pitanja)*

14 (N) Na sljedećem skupu treniramo naivan Bayesov model za binarnu klasifikaciju *"Skupo ljetovanje na Jadranu"*:

i	Mjesto	Otok	Smještaj	Prijevoz	$y^{(i)}$
1	Kvarner	da	privatni	auto	1
2	Kvarner	ne	kamp	bus	0
3	Dalmacija	da	hotel	avion	1
4	Dalmacija	ne	privatni	avion	0
5	Istra	da	kamp	auto	0
6	Istra	ne	kamp	bus	1
7	Dalmacija	da	hotel	auto	1

Procjene radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija primjera $\mathbf{x} = (\text{Istra}, \text{ne}, \text{kamp}, \text{bus})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.706   B 0.685   C 0.318   D 0.237

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

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

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

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

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

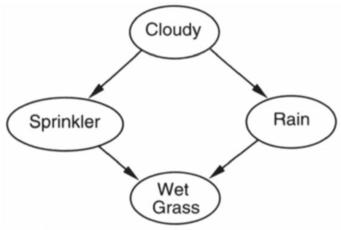
A $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_1 \supset \mathcal{H}_2$ C $\mathcal{H}_3 > \mathcal{H}_1 > \mathcal{H}_2$, $\mathcal{H}_3 \supset \mathcal{H}_2 \supset \mathcal{H}_1$

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

16 (P) Bacanje igraće kocke modeliramo kategoričkom varijablom $\mathbf{x}$, gdje indikatorske varijable $x_1, \ldots, x_6$ odgovaraju vrijednosti koju dobivamo bacanjem kocke. Za procjenu parametara $\boldsymbol{\mu}$ kategoričke distribucije koristimo MAP-procjenitelj s Dirichletovom distribucijom za apriornu gustoću vjerojatnosti. U stvarnosti, kocka je modificirana tako da će nešto češće davati šesticu, odnosno realizaciju $x_6 = 1$, međutim mi to ne znamo. Naprotiv, na temelju manjeg broja opažanja ranijih bacanja kocke utvrdili smo da je kocka najčešće davala peticu, no svjesni smo da je naša procjena ograničena na manji broj opažanja. **Uz koje parametre Dirichletove distribucije će naša procjena za μ biti najbliža stvarnoj vrijednosti tih parametara?**

A $\boldsymbol{\alpha} = (1,1,1,1,1,1)$ B $\boldsymbol{\alpha} = (2,2,2,2,2,2)$ C $\boldsymbol{\alpha} = (\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})$ D $\boldsymbol{\alpha} = (1,1,1,1,3,1)$

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



C	$P(C)$
0	0.5
1	0.5

S	C	$P(S\|C)$
0	0	0.5
0	1	0.9
1	0	0.5
1	1	0.1

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

W	R	S	$P(W\|R,S)$
0	0	0	1.0
0	0	1	0.9
0	1	0	0.1
0	1	1	0.01
1	0	0	0.0
1	0	1	0.1
1	1	0	0.9
1	1	1	0.99

Izračunajte aposteriornu vjerojatnost da pada kiša ako je trava mokra i nije oblačno.

A 0.709 B 0.825 C 0.112 D 0.491

18 (P) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x,y,z) = P(x)P(y)P(z|x,y)$. Sve varijable su binarne. Vrijedi $P(x=1) = 0.2$ i $P(y=1) = 0.3$. Tablica uvjetne vjerojatnosti za čvor z je sljedeća:

z	x	y	$p(z\|x,y)$	z	x	y	$p(z\|x,y)$
0	0	0	0.1	1	0	0	0.9
0	0	1	0.2	1	0	1	0.8
0	1	0	0.5	1	1	0	0.5
0	1	1	0.9	1	1	1	0.1

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(y|x=0,z=1)$. Uzorkovanje smo ponovili ukupno $N = 500$ puta. **Koliki je očekivani broj vektora koje ćemo pri uzorkovanju morati odbaciti?**

A 152 B 404 C 180 D 89

19 (P) Bayesova mreža ima pet varijabli, od kojih su v, w i z binarne, a x i y ternarne varijable. Topološki uređaj varijabli neka je v,w,x,y,z. Uz takav uređaj, u mreži vrijede sljedeće marginalne i uvjetne nezavisnosti:

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

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

A 22 B 27 C 10 D 25

Cjelina 5: Grupiranje *(3 pitanja)*

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

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

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

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

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

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

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

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.53 B 0.49 C 0.62 D 0.58

22 Za procjenu parametara modela GMM tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). To je iterativan optimizacijski algoritam. **Pod kojim uvjetima EM-algoritam (primijenjen na model GMM) konvergira, i kamo?**

A Algoritam uvijek konvergira, međutim globalni maksimum log-izglednosti parametara doseže samo ako je broj grupa postavljen na pravi broj grupa ili tako da je broj grupa jednak broju primjera

B Krenuvši od nekih početnih parametara, algoritam uvijek konvergira do parametara koji maksimiziraju očekivanje log-izglednosti, međutim to ne moraju biti parametri koji maksimiziraju vjerojatnost podataka

C Algoritm uvijek konvergira, i to do točke u prostoru parametara koja maksimizira log-izglednost parametara, no brzina konvergencije ovisi o toma kako su inicijalizirani parametri

D Algoritam konvergira samo ako su primjeri u ulaznom prostoru sferični, ako su zavisnosti između značajki linearne, i ako nema multikolinearnosti, jer u protivnom zavisnosti nije moguće modelirati kovarijacijskom matricom

Cjelina 6: Vrednovanje modela *(2 pitanja)*

23 (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):

$$\begin{array}{c c} & \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 5 & 13 & 1 \\ 6 & 5 & 4 \\ 4 & 2 & 16 \end{pmatrix} \end{array}$$

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

A 0.114 B 0.091 C 0.059 D 0.023

24 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 69201 B 35721 C 44640 D 49600

Pismeni ispit iz Strojnog učenja (ak. god. 2020./2021.)

*Ispit sadrži **24 pitanja** i ukupno nosi najviše 20 bodova (za 35% bodova na predmetu). Pitanja nose po 1 bod, a 1/3 boda oduzima se za pogrešan odgovor. Za maksimalan broj bodova dovoljno je točno riješiti **20 pitanja**, a višak bodova iznad 20 se zanemaruje. Trajanje ispita je **150 minuta**. Primjerak ispita morate predati zajedno sa svojim rješenjima.*

Cjelina 1: Osnovni koncepti i linearna regresija *(4 pitanja)*

1 (P) Za linearan model u $\mathcal{X} = \{0,1\}^3$ zadan je sljedeći skup primjera za učenje:

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

Optimizacijski postupak klasifikatora funkcionira tako da minimizira empirijsku pogrešku, definiranu kao očekivanje funkcije gubitka 0-1, i postupak u tome uvijek uspijeva. Želimo znati koju bi klasu ovaj klasifikator dodijelio primjeru $\mathbf{x} = (1,1,1)$. **Možemo li, na temelju iznesenih informacija, odrediti klasifikaciju dotičnog primjera i što nam to govori o induktivnoj pristranosti ovog algoritma**?

- A Možemo, $y = 1$, jer klasifikator ima induktivnu pristranost jezikom (linearan model) i preferencijom (primjeri za koje je $h(x) \geq 0$ klasificiraju se pozitivno)
- B Ne možemo, jer nije definirana induktivna pristranost preferencijom, pa činjenica da je model linearan nije dovoljan skup pretpostavki da bismo jednoznačno odredili klasifikaciju svih novih primjera
- C Možemo, klasifikacija je $y = 1$, i ovaj klasifikator ima definiranu induktivnu pristranost pomoću koje može jednoznačno odrediti klasifikaciju svakog primjera
- D Možemo, klasifikacija je $y = 1$, premda dane informacije nisu dovoljne za definicju induktivne pristranosti, što se vidi iz toga da je za ovaj skup primjera prostor inačica veći od jedan

2 (P) U ulaznom prostoru $\mathcal{X} = \{0,1\}^3$ definiramo sljedeći klasifikacijski model:

$$h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \geq 0\big\}$$

Koja je dimenzija prostora parametara te koliko različitih hipoteza postoji u ovom modelu?

- A Dimenzija prostora parametara i broj hipoteza su beskonačni
- B Dimenzija prostora parametara je 256, a hipoteza ima 14
- C Dimenzija prostora parametara je 4, a hipoteza ima manje od 256
- D Dimenzija prostora parametara je 4, a hipoteza ima beskonačno mnogo

3 (P) Koristimo regresiju za predviđanje kreditne sposobnosti klijenata. Na raspolaganju imamo podatke o dobi x_1, ukupnom prinosu na račune x_2, ukupnoj ušteđevini u kunama x_3, ukupnoj deviznoj ušteđevini u eurima x_4, ukupnom iznosu odobrenih kredita x_5, ukupnom iznosu do sada otplaćenih kredita x_6 te preostalom ukupnom dugovanju po svim kreditima x_7. Ove podatke želimo iskoristiti kao značajke, ali pritom naravno moramo paziti da ne uvedemo multikolinearnost, pa možda nećemo upotrijebiti sve značajke. Za funkciju preslikavanja $\boldsymbol{\phi}$ koristimo preslikavanje s linearnim, kvadratnim i interakcijskim značajkama. Za interakcije razmatramo parove i trojke značajki, uključivo kombinacije kvadriranih i nekvadriranih značajki (npr., x_1x_2, $x_1^2x_2$ i $x_1x_2^2x_3^2$). **Koliko minimalno primjera za učenje trebamo imati, a da bi rješenje bilo stabilno i bez regularizacije?**

A 86 B 48 C 79 D 92

4 (T) Svaki algoritam strojnog učenja ima neku induktivnu pristranost. Induktivna pristranost sastoji se od pristranosti jezika i pristranosti preferencije. **Koja je razlika između induktivnih pristranosti regularizirane i neregularizirane linearne regresije?**

- A Algoritmi imaju različite pristranosti, i to različitu pristranost preferencije jer regularizirana regresija preferira jednostavnije hipoteze, a onda i različitu pristranost jezika jer je model neregularizirane regresije nadskup modela regularizirane regresije
- B Za razliku od neregularizirane regresije, regularizirana regresija preferira jednostavnije hipoteze, međutim pristranosti su im identične jer su oba algoritma definirana kao linearna kombinacija značajki i težina te oba koriste identičan optimizacijski postupak (pseudoinverz matrice dizajna)
- C Oba algoritma imaju isti model, definiran kao linearnu kombinaciju značajki i težina, pa dakle imaju istu pristranost jezika, ali se razlikuju u pristranosti preferencije jer imaju različito definiranu empirijsku pogrešku (osim ako je regularizacijski faktor jednak nuli)
- D Algoritmi se ne razlikuju po pristranosti preferencijom budući da koriste istu funkciju gubitka (kvadratni gubitak), međutim regularizirana regresija ima jaču induktivnu pristranost jezika od regularizirane regresije budući da prvi model uključuje drugi model

Cjelina 2: Logistička regresija *(4 pitanja)*

5 (T) Na predavanjima smo za klasifikaciju pokušali upotrijebiti algoritam regresije. Zaključili smo da to ne funkcionira, tj. da algoritam linearne regesije jednostavno nije klasifikacijski algoritam. **Koje bismo minimalne preinake trebale učiniti u algoritmu linearne regresije, a da on dobro funkcionira kao klasifikacijski algoritam?**

- A promijeniti model, funkciju gubitka i optimizacijski postupak
- B promijeniti optimizacijski postupak
- C promijeniti funkciju gubitka
- D promijeniti model i optimizacijski postupak

6 (P) Razvijamo sustav za automatsku klasifikaciju novinskih članaka u jednu od pet kategorija. Tih pet kategorija su "sport", "politika", "kriminal", "znanost" i "lifestyle". Najveća razlika u veličini klasa je između kategorija "politika" i "znanost". Očekivano, u kategoriji "politika" ima najviše članaka, dok ih u kategoriji "znanost" ima 5x manje, što je u redu jer to ionako nitko ne čita. Svaki novinski članak prikazujemo kao vektor riječi, gdje su komponente vektora broj pojavljivanja pojedine riječi. Problem rješavamo algoritmom SVM, i to u primarnoj formulaciji. Budući da je SVM binaran klasifikator, odlučili smo primijeniti shemu OVR ili shemu OVO za dekompoziciju višeklasnog klasifikacijskog problema u skup binarnih klasifikacijskih problema. **Što možemo očekivati?**

- A SVM+OVO će imati 2x puta manje značajki od SVM+OVR, ali bi mogao raditi bolje na člancima iz kategorije "znanost"
- B SVM+OVR će imati 5x manje značajki od SVM+OVO, ali bi mogao raditi bolje na člancima iz kategorije "znanost"
- C SVM+OVR će imati 2x manje značajki od SVM+OVO, ali bi mogao raditi lošije na člancima iz kategorije "znanost"
- D SVM+OVO će imati 5x puta manje značajki od SVM+OVR, ali bi mogao raditi lošije na člancima iz kategorije "znanost"

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

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

U jednoj iteraciji treniranja modela vektor parametara jednak je

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

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

A 4.00 B 0.70 C 2.48 D 1.28

8 (N) Raspolažemo označenim skupom primjera iz triju klasa ($K = 3$) u trodimenzijskome ulaznom prostoru ($n = 3$). Na tom skupu treniramo model multinomijalne logističke regresije. Treniranje provodimo gradijentnim spustom. U nekoj od iteracija gradijentnog spusta matrica težina je sljedeća (stupci odgovaraju težinama za pojedine klase):

$$\mathbf{W} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 0 & -2 \\ 3 & -4 & 6 \\ -3 & 0 & 2 \end{pmatrix}$$

Jedan od primjera u skupu za učenje je primjer $\mathbf{x} = (-4, -1, -3)$ s oznakom $\mathbf{y} = (0, 0, 1)$. **Koliko iznosi gubitak unakrsne entropije koji u ovoj iteraciji optimizacijskog postupka nanosi dotični primjer?**

A 4.02 B 8.00 C 12.02 D 6.00

Cjelina 3: Jezgrene i neparametarske metode *(5 pitanja)*

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

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

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

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

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

A 0.0089 B 0.0045 C 0.0013 D 0.0024

10 (T) Kod izvoda algoritma SVM s tvrdom marginom, pretpostavili smo da za primjere $\mathbf{x} \in \mathbb{R}^n$ vrijedi sljedeći uvjet linearne odvojivosti:

$$\forall(\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}.\ y^{(i)}h(\mathbf{x}^{(i)}) \geq 0$$

Koliko hipoteza zadovoljava ovaj uvjet, i kako algoritam SVM odabire jednu od njih?

A Uvjet zadovoljava beskonačno mnogo hipoteza, a SVM odabire onu jednu koja minimizira kvadrat vektora težina te koja ispravno klasificira sve primjere, uz uvjet da $h(\mathbf{x})$ nije u intervalu $(-1, +1)$

B Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, a SVM između njih odabire onu jednu koja minimizira kvadrat vektora težina

C Uvjet zadovoljava beskonačno mnogo hipoteza, međutim samo za jednu vrijedi $yh(\mathbf{x}) = 1$ za najbliže primjere, i to je hipoteza koju odabire SVM

D Uvjet zadovoljava konačan broj hipoteza koje su linearno odvojive, no one se razlikuju samo po faktoru koji množi težine $(\mathbf{w}, w_0)$, pa SVM odabire onu jednu za koju vrijedi $yh(\mathbf{x}) \geq 1$ za sve primjere

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

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

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

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

12 (T) Algoritam SVM može biti parametarski i neparametarski, ovisno o tome provodimo li optimizaciju u primarnoj ili dualnoj formulaciji. U oba slučaja preferiramo da je model rijedak, tj. da je nakon treniranja što više parametara postavljeno na nulu. **Kako rijetkost modela ovisi o hiperparametru C?**

- A Što je C manji, to je neparametarski model rjeđi, ali to nema utjecaja na rijetkost parametarskog modela jer on nema potporne vektore
- B Što je C veći, to je neparametarski model manje rijedak, dok je parametarski to rjeđi jer λ pada
- C Što je C veći, to je neparametarski model manje rijedak, dok parametarski model nije rijedak jer ima L_2-regularizaciju a ne L_1-regularizaciju
- D Što je C manji, to je neparametarski model rjeđi, a također je to rjeđi i parametarski model jer λ raste

13 (N) Raspolažemo sljedećim skupom označenih primjera u trodimenzijskome ulaznom prostoru:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \{((1,-3,6),-1),((4,-4,4),-1),((-2,4,1),+1))\}$$

Na ovom skupu primjera treniramo model SVM-a s linearnom jezgrenom funkcijom i sa $C = 0.001$. Postupak treniranja algoritmom SMO završio je s vektorom Lagrangeovih koeficijenata $\boldsymbol{\alpha} = (0.001, 0, 0.001)$. Iz ovoga se da izračunati da vrijedi $w_0 = -0.943$. Umjesto algoritma SMO, za optimizaciju smo mogli upotrijebiti gradijentni spust i optimirati težine u primarnoj formulaciji problema. U tom slučaju koristili bismo empirijsku pogrešku SVM-a definiranu kao L_2-regularizirani gubitak zglobnice. Međutim, tu pogrešku možemo izračunati i naknadno, nakon što smo naučili model. **Koliko iznosi empirijska pogreška ovog SVM-a na skupu primjera $\mathcal{D}$?**

A 1.135 B 1.585 C 1.650 D 1.959

Cjelina 4: Probabilistički modeli *(6 pitanja)*

14 (P) Bacanje igraće kocke modeliramo kategoričkom varijablom $\mathbf{x}$, gdje indikatorske varijable $x_1, \dots, x_6$ odgovaraju vrijednosti koju dobivamo bacanjem kocke. Za procjenu parametara $\boldsymbol{\mu}$ kategoričke distribucije koristimo MAP-procjenitelj s Dirichletovom distribucijom za apriornu gustoću vjerojatnosti. U stvarnosti, kocka je modificirana tako da će nešto češće davati šesticu, odnosno realizaciju $x_6 = 1$, međutim mi to ne znamo. Naprotiv, na temelju manjeg broja opažanja ranijih bacanja kocke utvrdili smo da je kocka najčešće davala peticu, no svjesni smo da je naša procjena ograničena na manji broj opažanja. **Uz koje parametre Dirichletove distribucije će naša procjena za $\boldsymbol{\mu}$ biti najbliža stvarnoj vrijednosti tih parametara?**

A $\boldsymbol{\alpha} = (1,1,1,1,1,1)$ B $\boldsymbol{\alpha} = (\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})$ C $\boldsymbol{\alpha} = (1,1,1,1,3,1)$ D $\boldsymbol{\alpha} = (2,2,2,2,2,2)$

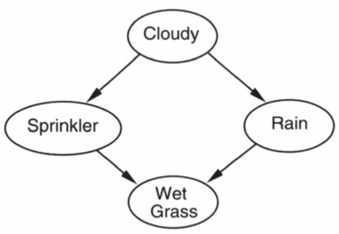
15 (P) Razmotrite jednostavnu Bayesovu mrežu koja odgovara faktorizaciji $P(x,y,z) = P(x)P(y)P(z|x,y)$. Sve varijable su binarne. Vrijedi $P(x=1) = 0.2$ i $P(y=1) = 0.3$. Tablica uvjetne vjerojatnosti za čvor z je sljedeća:

z	x	y	$p(z\|x,y)$	z	x	y	$p(z\|x,y)$
0	0	0	0.1	1	0	0	0.9
0	0	1	0.2	1	0	1	0.8
0	1	0	0.5	1	1	0	0.5
0	1	1	0.9	1	1	1	0.1

Postupkom uzorkovanja s odbijanjem uzorkujemo iz aposteriorne distribucije $P(y|x=0,z=1)$. Uzorkovanje smo ponovili ukupno $N = 500$ puta. **Koliki je očekivani broj vektora koje ćemo pri uzorkovanju morati odbaciti?**

A 404 B 152 C 89 D 180

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



C	$P(C)$
0	0.5
1	0.5

S	C	$P(S\|C)$
0	0	0.5
0	1	0.9
1	0	0.5
1	1	0.1

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

W	R	S	$P(W\|R,S)$
0	0	0	1.0
0	0	1	0.9
0	1	0	0.1
0	1	1	0.01
1	0	0	0.0
1	0	1	0.1
1	1	0	0.9
1	1	1	0.99

Izračunajte aposteriornu vjerojatnost da pada kiša ako je trava mokra i nije oblačno.

A 0.825 B 0.709 C 0.491 D 0.112

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

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

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

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

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

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

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

18 (P) Bayesova mreža ima pet varijabli, od kojih su v, w i z binarne, a x i y ternarne varijable. Topološki uređaj varijabli neka je v, w, x, y, z. Uz takav uređaj, u mreži vrijede sljedeće marginalne i uvjetne nezavisnosti:

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

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

A 25 B 10 C 27 D 22

19 (N) Na sljedećem skupu treniramo naivan Bayesov model za binarnu klasifikaciju *"Skupo ljetovanje na Jadranu"*:

i	Mjesto	Otok	Smještaj	Prijevoz	$y^{(i)}$
1	Kvarner	da	privatni	auto	1
2	Kvarner	ne	kamp	bus	0
3	Dalmacija	da	hotel	avion	1
4	Dalmacija	ne	privatni	avion	0
5	Istra	da	kamp	auto	0
6	Istra	ne	kamp	bus	1
7	Dalmacija	da	hotel	auto	1

Procjene radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija primjera $\mathbf{x} = (\text{Dalmacija}, \text{ne}, \text{kamp}, \text{bus})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.318 B 0.706 C 0.237 D 0.685

Cjelina 5: Grupiranje *(3 pitanja)*

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

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

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

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

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

21 Za procjenu parametara modela GMM tipično se koristi algoritam maksimizacije očekivanja (EM-algoritam). To je iterativan optimizacijski algoritam. **Pod kojim uvjetima EM-algoritam (primijenjen na model GMM) konvergira, i kamo?**

- A Algoritam uvijek konvergira, međutim globalni maksimum log-izglednosti parametara doseže samo ako je broj grupa postavljen na pravi broj grupa ili tako da je broj grupa jednak broju primjera
- B Algoritm uvijek konvergira, i to do točke u prostoru parametara koja maksimizira log-izglednost parametara, no brzina konvergencije ovisi o toma kako su inicijalizirani parametri
- C Algoritam konvergira samo ako su primjeri u ulaznom prostoru sferični, ako su zavisnosti između značajki linearne, i ako nema multikolinearnosti, jer u protivnom zavisnosti nije moguće modelirati kovarijacijskom matricom
- D Krenuvši od nekih početnih parametara, algoritam uvijek konvergira do parametara koji maksimiziraju očekivanje log-izglednosti, međutim to ne moraju biti parametri koji maksimiziraju vjerojatnost podataka

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

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

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

A 0.49 B 0.62 C 0.53 D 0.58

Cjelina 6: Vrednovanje modela *(2 pitanja)*

23 (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

A 49600 B 44640 C 35721 D 69201

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

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

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

A 0.091 B 0.059 C 0.023 D 0.114