SU 2020/21 Zadaci iz zadaća i kvizova

Ovaj dokument služi zajedničkom rješavanju zadataka iz domaćih zadaća i kvizova. Kopirajte tekst zadataka ili screenshot i zatim napišite rješenje zadatka. Mi (SU staff) ćemo provjeriti vaša rješenje i potvrditi rješenje ako je rješenje točno, odnosno ukazati na pogrešku ako je netočno. Provjeru ćemo raditi periodički, kad za to uhvatimo vremena. Možete koristiti komentare za diskusiju.

Slobodno editirajte i modificirajte ovaj dokument kako god želite, dok god to doprinosi lakšem i boljem savladavanju gradiva.

MOLIM VAS POMOZITE S FORMATIRANJEM DOKUMENTA GDJE GOD STIGNETE, PUNO NAS JE I POSTAJE TEŠKO ČITLJIVO yeet
AKO NE MOŽETE EDITIRATI DOKUMENT I VIDJETI KOMENTARE SA STRANE TO JE ZATO ŠTO NAS IMA PREKO 100 TRENUTNO NA DOKUMENTU. SVEJEDNO MOŽETE ČITATI, PRIČEKAJTE DA NETKO IZAĐE.

Domaće zadaće

DZ07: Procjena parametara

1. ***Neka je zajednicka vjerojatnost P(X,Y) varij ljedeća: P(1, 1) = 0.2, P(1, 2)=0.05, P(1, 3) = 0.3, P(2, 1) = 0.05, P(2, 2) = 0.3, P(2, 3) = 0.1.***
   - ***(a) Izracunajte ocekivanje E[X], varijancu Var(X), kovarijancu Cov(X, Y), koeficijent korelacije $\rho_{X,Y}$ i kovarijacijsku matricu Σ.***

     - $X = [[1,\ 2],\ [0.55,\ 0.45]],\ Y = [[1,\ 2,\ 3],\ [[0.25\ 0.35\ 0.4]]$
     - $E[X] = \sum_{i=1}^{2} p(x_i)x_i = 1.45,\ E[Y] = \sum_{i=1}^{3} p(y_i)y_i = 2.15$
     - $Var(X) = E[(X - E[X])]^2 = 0.55 \cdot (-0.45)^2 + 0.45 \cdot 0.55^2 = 0.2475$
       Drugi nacin (treba se dobit isto):
       $Var(X) = E[X^2] - (E[X])^2 = 0.55 \cdot 1 + 0.45 \cdot 4 - (1.45)^2 = 0.2475$
       $Var(Y) = 0.6275$***, dobiveno bilo kojom formulom***

- $Cov(X,\ Y) = \sum_{i=1}^{2}\sum_{j=1}^{3} p_{ij}(x_i, y_j)(x_i - E[X])(y_j - E[Y]) =$

  $0.2 \cdot (1 - 1.45)(Dokažite\ da\ su\ nezavisne\ varijable\ linearno\ nekorelirane.1 - 2.$
  $+ 0.05 \cdot (2 - 1.45)(1 - 2.15) + 0.3 \cdot (2 - 1.45)(2 - 2.15) + 0.1 \cdot (2 - 1.45)$
  $= -7/400 = -0.0175$

  Takoder, mozemo iskoristiti $Cov(X,Y) = E[XY] - E[X]E[Y]$ ***na nacin da prvo pomnozimo sve moguce vrijednosti X sa Y i vidimo vjerojatnosti dobivenih umnozaka, pa izracunamo ocekivanje E[XY], a a onda samo oduzmemo ocekivanje pojedinih***

  $XY = [[1,\ 2,\ 3,\ 4,\ 6], [0.2,\ 0.1,\ 0.3,\ 0.3,\ 0.1]]$
  $E[XY] = 0.2 \cdot 1 + 0.1 \cdot 2 + 0.3 \cdot 3 + 0.3 \cdot 4 + 0.1 \cdot 6 = 3.1$
  $Cor(X,Y) = 3.1 - 1.45 \cdot 2.15 = -0.0175$
- $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-0.0175}{\sqrt{0.2475 \cdot 0.6275}} =$***-0.0444***
- $\Sigma = [[0.2475,\ -0.0175], [-0.0175,\ 0.6275]]$

2. Razumjeti kako podaci odreduju izglednost parametara putem funkcije izglednosti

a. Definirajte nezavisnost slucajnih varijabli (preko zajednicke vjerojatnosti i preko uvjetne vjerojatnosti).

i. $P(X,Y) = P(X) \cdot P(Y)$***. Alternativno:***

$P(A \cap B) = P(A) \cdot P(B) \Rightarrow P(A) = \frac{P(A \cap B)}{P(B)} = P(A|B)$***, odnosno uvjetna vjerojatnost A, uz dani B, je ista kao vjerojatnost A (isto je i obratno), dakle B ne uvjetuje nista oko A, to jest, A i B su nezavisni.***

b. Sudeci po iznosu koeficijenta korelacije ρX,Y , jesu li varijable iz zadatka 1 linearno zavisne? Jesu li nezavisne?

i. Nisu linearno zavisne. Ne znamo jesu li nezavisne. (izgleda kao da postoji nekava zavisnost promatrajuci kov matricu).

c. Za koje od sljedecih varijabli ocekujete da su zavisne, a za koje da je ta zavisnost linearna: (i) dob i velicina cipela, (ii) dob i sati spavanja, (iii) razina buke i udaljenost od izvora buke, (iv) dob i prihodi?

i. Zavisne, ali ne linearno
ii. Zavisne, ali ne linearno
iii. Zavisne, i to linearno (ovisno o definiciji - decibeli su obično logaritam)
iv. Vrlo malo zavisne, ali ne linearno

3. [Svrha: Razumjeti kako podatci odreduju izglednost parametara putem funkcije izglednosti.]

a. ***Definirajte funkciju izglednosti L(θ|D). Na kojoj se pretpostavci o skupu D temelji ta definicija?***

   i. $L(\theta|D) = p(D|\theta) = \prod_{i=1}^{N} p(x^{(i)}|\theta)$ ***Pretpostavka iid.***

b. ***Raspolazemo skupom (neoznacenih) primjera D tx^(i) ui -2, -1, 1, 3, 5, 7u. Pretpostavljamo da se primjeri pokoravaju Gaussovoj distribuciji, x^(i) N (μ, σ2). Napiˇsite funkciju izglednosti L(μ, σ2 |D). Koliko iznosi izglednost parametara μ 0 i σ 2 1, a koliko vjerojatnost uzorka D uz te parametre?***

   i. $L(\mu, \sigma^2 | D) = p(-2 | \mu, \sigma^2)\, p(-1 | \mu, \sigma^2) \ldots p(7 | \mu, \sigma^2)$

   $L(\mu, \sigma^2 | D) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^6 exp\left(\frac{-(-2-\mu)^2}{2\sigma^2} + \frac{-(-1-\mu)^2}{2\sigma^2} + \ldots + \frac{-(7-\mu)^2}{2\sigma^2}\right)$

   $L(\mu, \sigma^2 | D) = \frac{1}{8\pi^3\sigma^6} exp\left(\frac{(-89+26\mu-6\mu^2)}{2\sigma^2}\right)$

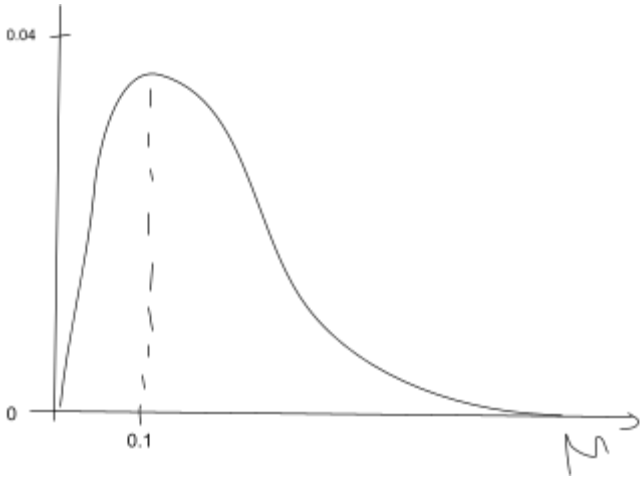
   $L(0, 1| D) = p(D | 0, 1) = \frac{1}{8\pi^3 1} exp\left(\frac{-89}{2}\right) = 1.9\,10^{-22}$

c. ***Novcic bacamo N puta, pri cemu smo m puta dobili glavu, a N m puta pismo. Ishodi bacanja novcica sacinavaju nas uzorak D. Napisite izraz za funkciju izglednosti parametra μ Bernoullijeve varijable, parametriziranu s N i m, tj. Lpμ|N, mq.***

   i. $L(\mu | N, m) = \mu^m (1 - \mu)^{N-m}$

d. ***Skicirajte funkciju izglednosti za sluˇcaj N 10 i m 1. Koja je vrijednost parametra μ najizglednija? Uz koju je vrijednost μ skup D najvjerojatniji?***

   i. $L(\mu | 10, 1) = \mu\,(1 - \mu)^9$**,** $\mu = 0.1$



DZ08: Bayesov klasifikator

Zadatak 2. Razmotrimo problem klasifikaciji neˇzeljene el. poˇste u klase spam (y 1), important (y 2) i normal (y 3). Neka su apriorne vjerojatnosti tih klasa Ppy 1q 0.2, Ppy 2q 0.05 i Ppy 3q 0.75. Za neku poruku el. poˇste x izglednosti iznose ppx|y 1q 0.8 i ppx|y 2q ppx|y 3q 0.5. Izraˇcunajte aposteriorne vjerojatnost za svaku od klasa te maksimalnu aposteriornu hipotezu za primjer x.

RJ.
P(x|y=1)*P(y=1) = 0.8*0.2 = 0.16

P(x|y=2)*P(y=2) = 0.5*0.05 = 0.025

P(x|y=3)*P(y=3) = 0.75*0.5 = 0.375 (primjer x ce biti svrstan u klasu 3, ali ne znamo s kolikom vjerojatnoscu.)

P(x) = P(x|y=1)*P(y=1) + P(x|y=2)*P(y=2) + P(x|y=3)*P(y=3) = 0.56

P(y=1|x) = P(x|y=1)*P(y=1) / P(x) = 0.16/0.56 = 0.2857

P(y=2|x) = P(x|y=2)*P(y=2) / P(x) = 0.025/0.56 = 0.0446

P(y=3|x) = P(x|y=3)*P(y=3) / P(x) = 0.375/0.56 = 0.6696 (primjer ce biti svrstan u klasu 3 s aposteriori vjerojatnoscu od 0.6696)

Zadatak 3.
(a) Definirajte naivan Bayesov klasifikator i pretpostavku na kojoj se temelji.
(b) Zaˇsto nam treba pretpostavka o uvjetnoj nezavisnosti znaˇcajki te kojoj vrsti induktivne pristranosti ona odgovara?

(c) Naivan Bayesov klasifikator koristimo za klasifikaciju rukom pisanih znamenki u jednu od deset klasa. Znamenke su prikazane kao vektor binarnih znaˇcajki (crno/bijeli slikovni elementi) u matrici s razlučivošću 32x32. Odredite ukupan broj parametara naivnog Bayesovog klasifikatora.
K = 10, n=32x32=1024, značajke se nezavisne i binarne
*Br.parametara = K*n = 10240? (provjeriti)*

Zadatak 4. Naivan Bayesov model ˇzelimo upotrijebiti za binarnu klasifikaciju "Skupo ljetovanje na Jadranu". Skup primjera za uˇcenje je sljede´ci:

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

(a) Izraˇcunajte MLE procjene svih parametara modela te klasificirajte primjere p(Istra, ne, kamp, bus) i p(Dalmacija, da, hotel, bus).
(b) Izraˇcunajte Laplaceove (zagladene) procjene za sve parametre modela te klasificajte nanovo iste primjere.

Zadatak 8.
Zelimo izgraditi klasifikator za klasifikaciju brucoˇsa u jednu od dvije klase: ˇ y 1 ñ "Završava FER u roku" i y 2 ñ "Produljuje studij". Svaki je primjer opisan sa ˇsest ulaznih varijabli: prosjek ocjena 1.–4. razreda (ˇcetiri varijable), bodovi državne mature iz matematike te bodovi državne mature iz fizike. Raspolaˇzemo trima modelima: modelom H1 s dijeljenom kovarijacijskom matricom, modelom H2 s dijagonalnom (i dijeljenom) kovarijacijskom matricom i modelom H3 s izotropnom kovarijacijskom matricom.
(a) Koliko svaki od ova tri modela ima parametara?
H1 : n(n+1)/2 +K*n + K-1 = 34*
*H2 : n+K*n + K-1 = 19*
*H3 : 1+K*n K-1= 14*

(b) Za koji od ova tri modela oˇcekujete da ´ce najbolje generalizirati u ovom konkretnom sluˇcaju (uzmite u obzir prirodu problema i oˇcekivane odnose izmedu znaˇcajki)? Zaˇsto?
(c) Nacrtajte skicu funkcije empirijske pogreˇske i pogreˇske generalizacije i naznaˇcite na njoj toˇcke koje oznaˇcavaju navedenim trima modelima.
(d) Kako biste u praksi odredili koji ´cete model upotrijebiti?

Zadatak 9.

a) Izvedite model logisti˘cke regresije krenuv˘si od generativne definicije za Ppy 1|xq. Izvod napravite korak po korak te se uvjerite da mo˘zete obrazlo˘ziti svaki korak u izvodu. Napi˘site sve pretpostavke koje ste ugradili u izvod.
(b) Model logisti˘cke regresije koristimo za binarnu klasifikaciju primjera s n 100 zna˘cajki. Odredite broj parametara modela logisti˘cke regresije te njemu odgovaraju´ceg generativnog modela.
Logreg: n+1 = 101 parametar
Bayes: n(n+1)/2 + Kn + (K-1) = 5250 parametara*
(c) Izra˘cunajte broj parametara za isti slu˘caj, ali sa K 5 klasa.
*Logreg: K*n+1 =501*
Bayes: n(n+1)/2 + Kn + (K-1) = 5554 parametara*
(d) Pretpostavite da klasificiramo u K 10 klasa. Izra˘cunajte koliko velika mora biti dimenzija prostora zna˘cajki n, a da bi se logisti˘cka regresija isplatila jer ima manje parametara od odgovaraju´cega generativnog modela.

DZ09: Probabilistički grafički modeli

Zadatak 2. Gradimo Bayesovu mrežu koja predviđa ho će li student/ica uspješno položiti SU. Mreža sadrži pet varijabli: pohađali osoba konzultacije (x1), je li osoba dobra u Pythonu (x2), rješava li osoba samostalno domaće zadaće i laboratorijske vježbe (x3), ocjenu iz predmeta UI (x4) te varijablu koja govori je li osoba položila SU (y). Pritom vrijedi x1, x2, x3, x5 e {T, F} i x4 e {2,3,4,5}. Vrijedi P(x1=T)=0.2 i P(x2=T)=0.6. Dane su i tablice uvjetnih vjerojatnosti.
(b)Koji je ukupan broj parametara ove mreže? 16?
Zar nije 14? Kod x4 ne treba 8 parametara nego 6 zato što se zadnji može izračunati kao 1 - svi ostali...

(c) Postupkom egzaktnog zaključivanja izračunajte P(y=T|x1=T, x4=3).
Na predavanju je rečeno da je odgovor 0.311, no ne dobivam tako. Vidi li netko gdje mi je greška?
P(y=T|x1=T, x4=3) = P(y=T,x1=T, x4=3)/P(x1=T, x4=3) = (suma po x2 i x3 od P(y=T,x1=T,x2, x3, x4=3))/(suma po x2, x3 i y od P(x1=T, x4=3, x2, x3, y)) = 0.02212/0.028 = 0.79
*suma po x2 i x3 od P(y=T,x1=T,x2, x3, x4=3) = suma po x2, x3 od P(x1=T)*P(x2)*P(x3|x1=T,x2)*P(x4=3|x2)*P(y=T|x3) =*
0.2(0.4*0.2*0.2*0.2+0.4*0.8*0.2*0.9+0.6*0.1*0.1*0.2+0.6*0.9*0.1*0.9) = 0.02212*
*suma po x2, x3 i y od P(x1=T, x4=3, x2, x3, y) = suma po x2, x3, y od P(x1=T)*P(x2)*P(x3|x1=T,x2)*P(x4=3|x2)*P(y|x3) = 0.2*(0.4*0.2*0.2 + 0.4*0.8*0.2 + 0.6*0.1*0.1 + 0.6*0.9*0.1) = 0.028*

(d) Koja je razlika izmedu posteriornog i MAP-upita? O kakvom tipu upita se radiu pro˘slom zadatku? Obrazlo˘zite.
(e) Utječe li broj varijabli u mreži na učinkovitost zaključivanja? Za˘sto?

Zadatak…

DZ10: Grupiranje

2. Zadatak
Raspolažemo skupom neoznačenih primjera: D{(a=5,2), b=(7,1), c=(1,4), d=(6,2), e=(2,8), f=(3,6), g=(0,4)}.
(a) Izvedite jedan korak algoritma k-sredina uz K = 3. Za početna središta odaberite μ_1=b,μ_2=c i μ_3=e
Nakon prve iteracije:
μ_1=(6,5/3) ,μ_2= (1/2,4) i μ_3=(5/2,7)
(b) Izvedite jedan korak algoritma k-medoida uz K = 3. Za početna središta odaberite primjere b,c i e.
Nakon prve iteracije:
μ_1= d,μ_2= c, μ_3=e

Jesu li ovo dobra rjesenja?
Ako su u grupi dva člana, mijenja li se medoid u sljedećoj iteraciji (npr. bilo bi svejedno odaberemo li c ili g u drugoj iteraciji, koji biramo?)

4. zadatak
U slučaju izjednačenosti, je li točno da su moguća dva različita dendrograma?

Koja pravila određuju na kojoj razini presjeći dendrogram? U skripti nisam ništa našao.

Zadatak…

DZ11: Vrednovanje modela

Zadatak…

Zadatak…

Moodle kvizovi

Cjelina 4A: Procjena Parametara + Bayesov klasifikator

1. ***(N) Raspolažemo sljedećim skupom označenih primjera:***
   D={x(i),y(i)}={(−2,1),(−2,1),(−1,0),(0,0),(1,1),(3,1)}
   Na ovom skupu treniramo univarijatni Bayesov klasifikator, za što trebamo procijeniti izglednosti klasa p(x\|y). Te su izglednosti definirane Gaussovom gustoćom vjerojatnosti:
   $p(x|\mu,\sigma 2) = 12\pi -- \sqrt{\sigma exp}\{- (x - \mu)22\sigma 2\}$

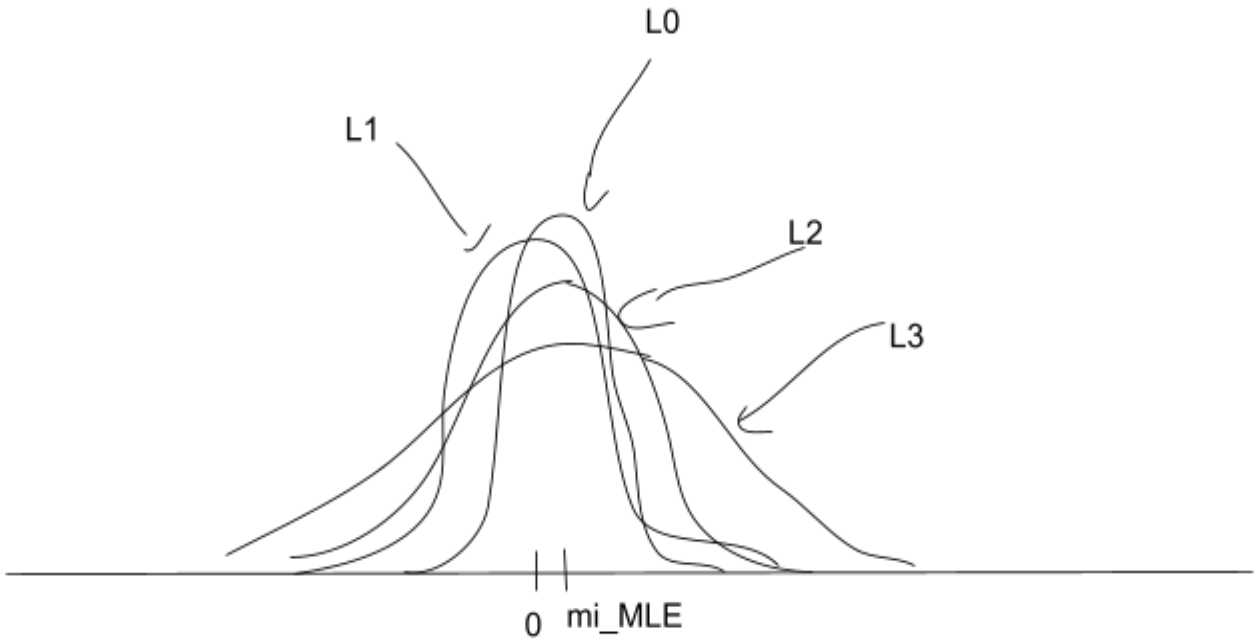
   Parametre μ i σ2 gustoće vjerojatnosti p(x\|y) procjenjujemo MLE-om. Neka su μ1 i σ21 parametri gustoće vjerojatnosti p(x\|y=1) dobiveni MLE-om na podskupu primjera Dy=1.
   Koliko iznosi log-izglednost L(μ1,σ21\|Dy=1)?

$$\mu_{MLE} = \frac{-2\ -2\ +1\ +3}{4} = 0$$
$$\sigma^2 = \frac{(-2+0)^2+(-2+0)^2+(3+0)^2+(1+0)^2}{4} = 4.5$$
$$L(\mu_1, \sigma^2{}_1|D_{y=1}) = \frac{-4}{2} ln(2\pi) - 4ln(\sqrt{4.5}) - \frac{(-2+0)^2+(-2+0)^2+(3+0)^2+(1+0)^2}{2\ 4.5} = -8.68$$

2. ***(P) Neka je L(μ,σ2\|D) log-izglednost parametara μ i σ2 normalne distribucije izračunata nad uzorkom D koji sadrži ukupno N opažanja normalne varijable x. Nadalje, neka su (μMLE,σ2MLE) parametri normalne razdiobe procijenjeni MLE-om nad uzorkom D, te neka je σ2UB nepristrana procjena varijance, izračunata kao σ2UB=NN−1σ2MLE. Konačno, neka je D' slučajno uzorkovan podskup uzorka D, tj. D'⊂D, pri čemu je poduzorkovanje načinjeno nakon procjene parametara.***

   Razmotrite sljedeće četiri vrijednosti funkcije log-izglednosti L(μ,σ2\|D):
   L0L1L2L3=L(μMLE,σ2MLE\|D)=L(0,1\|D)=L(μMLE,σ2UB\|D)=L(μMLE,σ2UB\|D')

   Što možemo zaključiti o odnosima između ovih vrijednosti funkcije log-izglednosti?



$L_0 \geq L_1 \quad L_0 > L_2 > L_3$

L0 je veci od L1 jer ima pristran procjenitelj pa je manja varijanca
L2 je manji jer ima nepristran procjenitelj (procjenitelj najveće izglednosti, koji ima L_0, je pristran)

L3 je manji jer ima nepristran procjenitelj i jer je izracunat nad manjim skupom podataka?

Zašto $L0 > L1, L2 \geq L3$ ne bi mogao biti točan?
Pretpostavljam da taj odgovor nije točan jer je L3 izračunat nad strogo manjim podskupom podataka.
oA zašto vrijedi L0 >= L1, a ne L0 > L1?

3. ***(N) U beta-Bernoullijevom modelu, apriornu vjerojatnost parametra μ modeliramo beta-distribucijom, čija je gustoća vjerojatnosti definirana kao: $p(\mu|\alpha, \beta) = 1 / B(\alpha, \beta) * \mu^{\alpha-1} (1 - \mu)^{\beta-1}$.***

   Mod (maksimizator) te distribucije jest: $\mu* = (\alpha - 1) / (\alpha + \beta - 2)$.

   Aposteriorna distribucija parametra definirana je kao: $p(\mu|D, \alpha, \beta) = \mu^{m+\alpha-1} (1 - \mu)^{N-m+\beta-1} * 1 / (B(\alpha, \beta) * p(D))$

   Neka α=β=2. Računamo MAP-procjenu za parametar μ Bernoulijeve varijable. To radimo na dva uzorka, D1=(N1,m1) i D2=(N2,m2), koji nam pristižu jedan za drugim. Pritom koristimo svojstvo konjugatnosti, na način da aposteriornu gustoću vjerojatnosti izračunatu na temelju prvog uzorka koristimo kao apriornu gustoću vjerojatnosti pri procjeni na temelju drugog uzorka. U prvom uzorku, veličine N1=50, Bernoullijeva varijabla realizirana je s vrijednošću y=1 ukupno

m1=42 puta. U drugom uzorku, veličine N2=15, Bernoullijeva varijabla realizirana je s vrijednošću y=1 ukupno m2=3 puta.

Izračunajte MAP-procjene za parametar μ na temelju ova dva uzorka. Koliko iznosi promjena u procjeni za μ između prve i druge procjene?

Nije teško, jedino što se mora znati je da se MAP procjena odnosi na mod (maksimum) aposteriorne distribucije.

Nakon prvog uzorka, parametri su $\alpha' = \alpha + m_1 = 44$ i $\beta' = \beta + (N_1 - m_1) = 10$ pa je procjena $\mu = (\alpha - 1) / (\alpha + \beta - 2) = 43/52$.

Nakon drugog uzorka, parametri su $\alpha'' = \alpha' + m_2 = 47$ i $\beta'' = \beta' + (N_2 - m_2) = 22$ pa je procjena $\mu = (\alpha - 1) / (\alpha + \beta - 2) = 46/67$.

Razlika je 46/67 - 43/52 = -0.14.

4. *(P) Koristimo MAP-procjenitelj kako bismo procijenili parametre distribucije kategoričke (multinulijeve) varijable X. Varijabla može poprimiti tri vrijednosti, x_1, x_2 i x_3, pa dakle trebamo procijeniti vektor parametara (μ_1, μ_2, μ_3). Budući da se ovdje radi o kategoričkoj varijabli, za MAP-procjenu koristimo Dirichlet-kategorički model.*

   Na temelju stručnog znanja o problemu koji rješavamo, u procjenu smo ugradili naše pretpostavke. To znači da smo na prikladan način definirali Dirichletovu apriornu gustoću vjerojatnosti, $p(\mu_1, \mu_2, \mu_3|\alpha_1, \alpha_2, \alpha_3)$, gdje je ($\alpha_1$, α_2, α_3) vektor hiperparametara (parametri Dirichletove distribucije). Konkretno, te smo hiperparametre definirali kao (α_1, α_2, α_3)=(2, 2, 1).

   Međutim, skup podataka D ne odgovara našoj pretpostavci. U tom skupu, varijabla X je u pola slučajeva realizirana s vrijednošću x_2, u pola slučajeva s vrijednošću x_3, no baš niti jednom s vrijednošću x_1.

   Kakva će biti naša MAP-procjena parametara (μ_1 ,μ_2, μ_3)?

   Ovdje mi je jasno da će vrijediti $\mu_2 > \mu_3$ zato što je $\alpha_2 > \alpha_3$ te da μ_1 neće doći do nule zbog apriornog znanja, tj. $\alpha_1 = 2$. To nameće $0 < \mu_1 < ⅓$, $⅓ < \mu_2 < 1$, $0 < \mu_3 < \mu_2 < 1$ kao odgovor ali nije mi jasno što tu radi ⅓?

$$\alpha_1 = 0 + 2 = 2\alpha_2 = k + 2\,\alpha_3 = k + 1$$

$$\mu_{k,MAP} = \frac{\alpha_k - 1}{\sum_k \alpha_k - K}$$

$$\mu_1 = \frac{2-1}{2+k+2+k+1-3} = \frac{1}{2(k+1)}$$

$$\mu_2 = \frac{k+2-1}{2(k+1)} = \frac{1}{2}$$

$$\mu_3 = \frac{k+1-1}{2(k+1)} = \frac{k}{2(k+1)}$$

Uvrstis za razlicite k i vidis da ce ti μ_1 ***uvijek biti izmedu 0 i ⅓,*** μ_2 ***½ i*** $\mu_3 = 1 - \mu_1 - \mu_2$ ***; što ako uvrstim k=0 i onda je*** μ_1 ***= ½ (se možda ne smije uzeti k=0 i koji je razlog ako ne? k--broj realizacija k-te vrijednosti)***

Zašto odgovor ne može biti 0 < $\mu_1 < \mu_3 < 1$, $\frac{1}{3} < \mu_2 < \frac{2}{3}$ ***? Tu je i dalje zadovoljeno*** $\mu_2 = \frac{1}{2}$ ***i*** $\mu_1 > 0$***?***

5. ***Funkcija izglednosti*** $L(\theta|D)$ ***nije isto što i vjerojatnost.***

   Po čemu se izglednost razlikuje od vjerojatnosti?

   Funkcija izglednosti $L(\theta|D)$ ***jednaka je gustoći vjerojatnosti*** $p(D|\theta)$***, samo što je izglednost funkcija parametara*** θ ***dok je*** $p(D|\theta)$ ***funkcija uzorka*** D***.***

   Evo onda pokušaj da obrazloži zašto su ostale tvrdnje krive:
   Ako su podatci diskretni (kategoričke značajke), onda je funkcija izglednosti parametara θ isto što i zajednička vjerojatnost uzorka D i parametara θ

   Za razliku od vjerojatnosti, funkcija izglednosti L(θ|D) je simetrična, u smislu da vrijedi L(θ|D)=p(D|θ)
   Obje su simetrične?
   Mislim da je ovdje "krivo" to što kaže da je simetričnost == L(θ|D)=p(D|θ)?

   Funkcija izglednosti L(θ|D) jednaka je gustoći vjerojatnosti p(θ|D), ali, za razliku od gustoće vjerojatnosti, nije odozgo ograničena sa 1
   Za ovo nemam ideju
   Ja bih rekla da je sve tocno osim p(θ|D), trebalo bi pisati p(D|θ) -> da, upravo zato je ovaj odg kriv

6. ***(N) Treniramo naivan Bayesov model za binarnu klasifikaciju \emph{``Skupo ljetovanje na Jadranu''}. Skup primjera za učenje je sljedeći:***
Procjene parametara radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija sljedećeg primjera:
x=(Istra,ne,kamp,bus)

(N) Treniramo naivan Bayesov model za binarnu klasifikaciju \emph{``Skupo ljetovanje na Jadranu''}. Skup primjera za učenje je sljedeći:

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

Procjene parametara radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija sljedećeg primjera:

$$\mathbf{x} = (\text{Istra}, \text{ne}, \text{kamp}, \text{bus})$$

Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?

- 0.1747
- 0.6856
- 0.0032
- 0.3144

Koliko iznosi aposteriorna vjerojatnost P(y=1|x)?

$$P(x,\ y = 1) = \frac{3}{7}\ \frac{0+1}{3+3}\ \frac{1+1}{3+2}\ \frac{1+1}{3+3}\ \frac{1+1}{3+3} = \frac{1}{315} = 0.00317$$

$$P(x,\ y = 0) = \frac{4}{7}\ \frac{2+1}{4+3}\ \frac{2+1}{4+2}\ \frac{2+1}{4+3}\ \frac{1+1}{4+3} = \frac{36}{2401} = 0.01499$$

$$P(x) = P(x,\ y = 1) + P(x,\ y = 0) = 0.01817$$

$$P(y = 1 \mid x) = \frac{P(x, y=1)}{P(x)} = \frac{343}{1963} = 0.1747$$

Drugi način...
Prior vjerojatnosti za klase: P(y=1) = 3/7 , P(y=0) = 4/7
Dalje...
P(x1=Istra | y=1) = (0+1) / (3+3)
P(x1=Istra | y=0) = (2+1) / (4+3)

P(x2=ne | y=1) = (1+1) / (3+2)
P(x2=ne | y=0) = (2+1) / (4+2)
P(x3=kamp | y=1) = (1+1) / (3+3)
P(x3=kamp | y=0) = (2+1) / (4+3)
P(x4=bus | y=1) = (1+1) / (3+3)
P(x4=bus | y=0) = (1+1) / (4+3)

x=(Istra, ne, kamp, bus); P(x|y=1)P(y=1) = 0.003; P(x|y=0)P(y=0) = 0.015
P(x) = P(x|y=1)P(y=1) + P(x|y=0)P(y=0) = 0.018

P(y=1|x) = P(x|y=1)P(y=1) / P(x) = 0.167

Gdje griješim? Previse zaokruzujes. Ako odjednom upises sve u kalkulator dobit ces dobar broj.

7. *(N) Na skupu označenih primjera u ulaznome prostoru dimenzije n=3 treniramo G*

$$h(\mathbf{x})_j = -\frac{1}{2}\ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) + \ln P(y=j)$$
$$= -\frac{1}{2}\ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j\right) + \ln P(y=j)$$

aussov Bayesov klasifikator za klasifikaciju primjera u K=2 klase, uz pretpostavku dijeljene kovarijacijske matrice.

Rješava li se ovaj zadatak tako da doslovno uvrstim u formulu iznad (sumnjam jer mi ispada -6.06)?

Dijeljena kovarijacijska matrica dobiva se kao težinski prosjek danih kovarijacijskih matrica i dobije se $\boldsymbol{\Sigma}$
*= {{3,0,0},{0,1,0},{0,0,2}}. Zatim se dobije p(x|y=1) = 8.0719*10^(-3). h1(x)=ln(p(x|y)*P(y)) = -6.429. Točan odgovor je -6.885. Koju pogrešku radim? Vrijednosti kov matrice ne podijeliš s 2 na kraju jer je prosjek težinski. Prva matrica utječe na dijeljenu s faktorom 0.2, a druga s 0.8.*

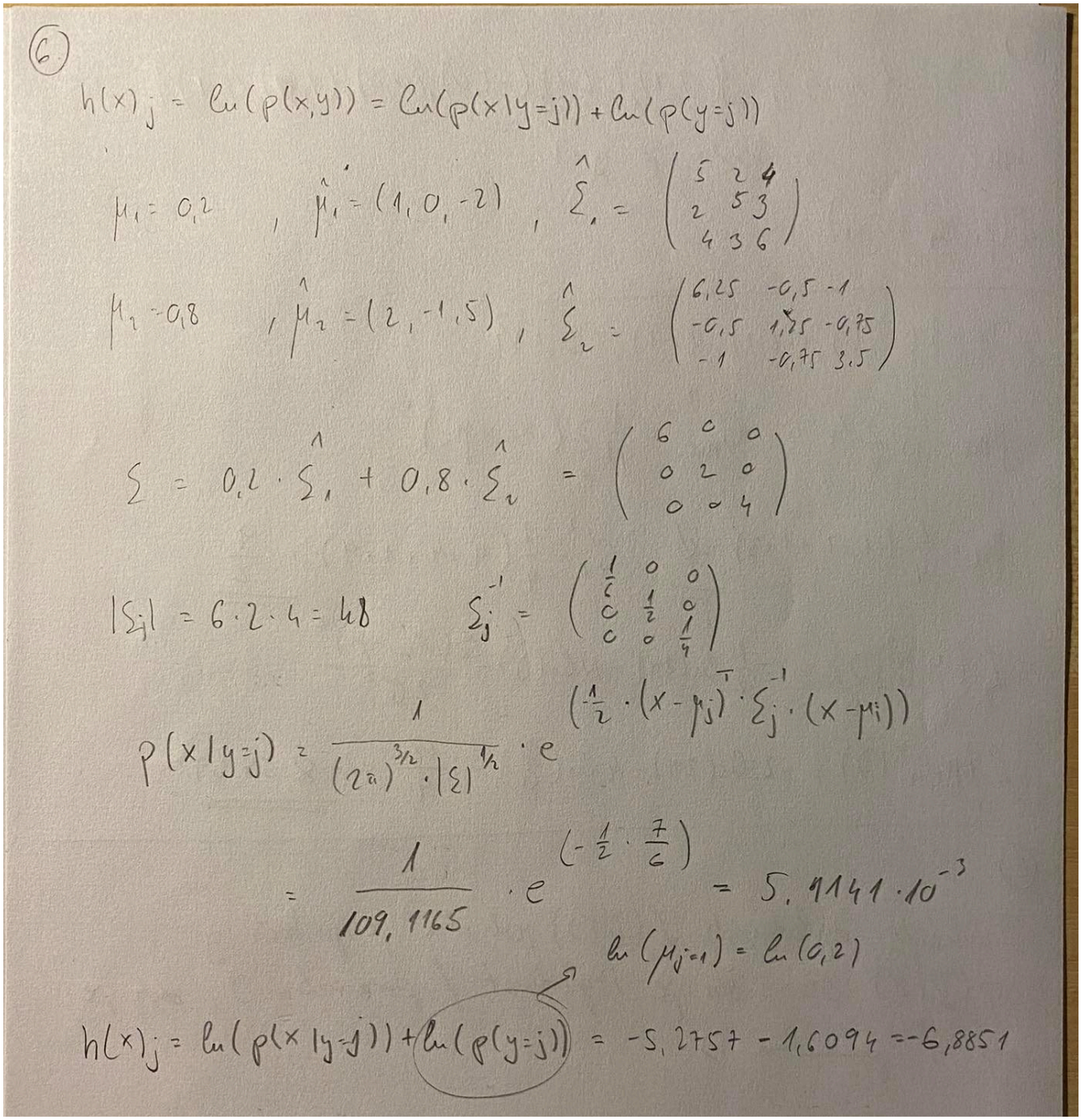
*Ja dobijem dijeljenu kov mat sa duplim vrijednostima, npr. Element 0,0 =0.2*5 + 0.8*6.25=6*
*Njena determinanta je umnozak po dijagonali=48 ,inverz su reciprocni brojevi po dijagonali, dobijem -7/12 kao eksponent od e i kad se sve uvrsti u ln(p(x|y=1)*p(y)) dobije se -6.885*
Sadrži li formula za p(x|y = 1) u sebi pi?

Kako ovaj eksponent dobiti? Kad idem računati mi se pojavi greška s dimenzijonalnosti.
Dimenzije : (1,3)x(3,3)x(3,1)=(1,1) :)

Otkuda dobijem P(y = j)?

Pretpostavljam da iz $\hat{\mu}_j = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}\{y^{(i)} = j\} = \frac{N_j}{N}$ ***mozemo zakljuciti da je p(y=j) = mi_j s obzirom da je K = 2, ako uvrstimo p(y=j) = 0.2 dobije se tocno rjesenje.***

Uspio sam riješiti zadatak tako da dobijem ispravni rezultat nakon skoro 2 sata mučenja skužio sam da sam predivido korijen na determinanti (Dabog da se ne ponovilo sutra :D, evo slike)



Šta nam je tu vektor mi, to nikako ne mogu shvatiti??
Ako je netko smotan kao ja, vektor mi(1) i mi(2) su zadani, kako nam se traži predikcija za y=1 koristimo vektor mi(1), da se tražila predikcija za y=2 onda bi se koristio drugi zadani.

Može netko staviti sliku postupka?

8. *(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.*

H1 -> značajke nisu korelirane, no imaju različite varijance unutar klase i između klasa
H2 -> značajke nisu korelirane, imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu
H3 -> između značajki postoje korelacije, ali se one ne razlikuju između klasa

Moj način razmišljanja (gledam šta nam sve treba):
- *H1: vjerojatnost klase, srednje vrijednosti, varijance za svaku klasu i za svaku značajku*
  - *K - 1 + nK + nK = 109*
- *H2: vj. klase, srednje vrijednosti, jedna varijanca za svaku klasu*
  - *K - 1 + nK + K = 69*
- *H3: vj. klase, srednje vrijednosti, jedna kov. matrica*
  - *K - 1 + nK + n/2 * (n+1) = 74*

Zato je H1 > H3 > H2. ***Može netko potvrdit ima li ovo smisla?***

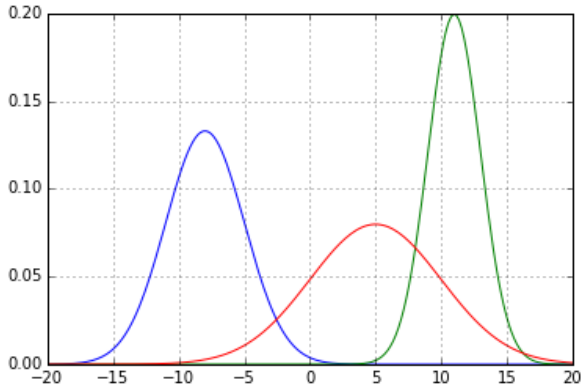
Kako ovaj? Po meni H1 ima 2KN parametara što je manje od H2 s K + Kn, ali je H1 očito složeniji od H2; nema nijedan odgovor s H2 > H1 i H1 složeniji od H2]
H1 - dijagonalna, H2 - dijeljena i dijagonalna, H3 - dijeljenja
^ Kod H2, zar "imaju jednaku varijancu unutar svake klase, no različitu za svaku klasu" ne znači izotropna + nije dijeljena? Da, u pravu si.
H3 je slozeniji od H2, H1 je slozeniji od H2
*Br parametara H1 = 50 (n*K), br parametara H2 = 5 (n), br parametara H3 = 15 (n/2 * (n+1))*
(samo gledamo kov matricu jer je ostali broj parametara isti za sva 3 modela)

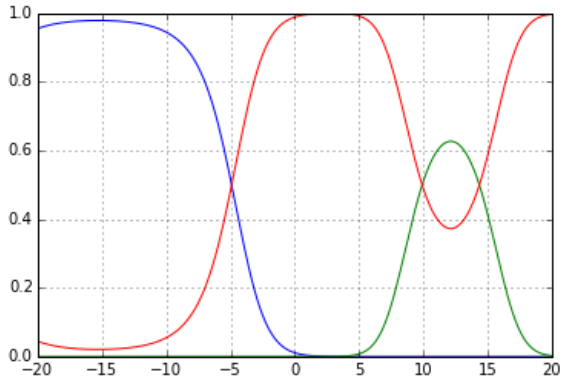
Zar ne bi H2 imao 10 parametara jer ima 10 klasa koje imaju međusobno različite varijance, ali unutar sebe jednake, tj. 10 različitih izotropnih matrica?
Da u pravu si.

H1 - dijagonalna, H2 - izotropna, H3 - dijeljena

Prema tome, rekao bih da broj parametara ide ovako:
> H1: (K - 1) + K*n = 59 (vjerojatnost svake klase, varijanca za svaku značajku i to puta broj klasa)
> H2: (K - 1) + K = 19 (vjerojatnost svake klase, varijanca značajki svih klasa jer je mat izotropna)
> H3: (K - 1) + (n / 2)*(n + 1) = 24 (vjerojatnost svake klase, jedna dijeljena kovarijacijska matrica)

9. ***(P) Koristimo Gaussov Bayesov klasifikator kako bismo riješili troklasni klasifikacijski problem. Procijenjene gustoće vjerojatnosti za izglednosti klasa su p(x|y=1)=N(−8,3), p(x|y=2)=N(5,5) i p(x|y=3)=N(11,2). Na slikama ispod prikazane su izglednosti klasa (lijeva slika) i aposteriorne vjerojatnosti dobivene Bayesovim pravilom (desna slika):***





S obzirom na ova dva grafikona, što su najizglednije vrijednosti za apriorne vjerojatnosti klasa?

$0.0807P(y = 1) = 0.0108P(y = 2)$
$P(y = 1) = 0.1338\ P(y = 2)$

$P(y = 3|x = 10) = P(y = 2|x = 10) = 0.5$

help. $P(y = 1|x = -5) = P(y = 2|x = -5) = 0.5$
$P(x = -5|y = 1)P(y = 1) = P(x = -5|y = 2)P(y = 2)$***f***

$P(x = 10|y = 3)P(y = 3) = P(x = 10|y = 2)P(y = 2)$
$0.1760P(y = 3) = 0.0484P(y = 2)$
$P(y = 3) = 0.2750P(y = 2)$

$$P(y = 1) + P(y = 2) + P(y = 3) = 1$$
$$0.1338P(y = 2) + P(y = 2) + 0.2750P(y = 2) = 1$$

$$P(y = 1) = 0.1$$
$$P(y = 2) = 0.7$$
$$P(y = 3) = 0.2$$

10. ***(N) Treniramo binarni klasifikator za analizu predsjedničke izborne kampanje. Svrha klasifikatora jest predvidjeti hoće li kandidat ili kandidatkinja skupiti dovoljno potpisa za kandidaturu. Model koristi pet značajki: x1 -- politička orijentacija (kategorička značajka s tri vrijednosti), x2,x3 -- dob kandidata i politički staž (dvije numeričke značajke), x4 -- populist (binarna značajka) i x5 -- kandidat/kinja velike političke stranke (binarna značajka). Primijetite da u istom modelu kombiniramo diskretne i kontinuirane značajke, što je sasvim legitimno. Razmatramo tri modela različite složenosti:***
H0: H1: H2: Bayesov klasifikator bez ikakvih pretpostavki o uvjetnoj nezavisnostiPolunaivan Bayesov klasifikatorNaivan Bayesov klasifikator

Polunaivan model H1 isti je kao i naivan model H2, s tom razlikom da smo u jedan faktor združili značajke x1 i x4, sluteći ipak da bi pokoji kandidat mogao dobro kapitalizirati populizam u kombinaciji s nekom etabliranom političkom orijentacijom. Kod naivnog Bayesovog klasifikatora naivnu pretpostavku uveli smo za sve varijable (i za diskretne i za kontinuirane). U sva tri modela za značajke x2 i x3 koristimo dijeljenu kovarijacijsku matricu.

Izračunajte broj parametara za svaki od ova tri modela. Koliko parametara sveukupno imaju ova tri Namodela?

$$H_0: P(y)P(x_1, x_4, x_5 \mid y)P(x_2, x_3 \mid x_1, x_4, x_5, y) = 1 + (11 \cdot 2) + (\frac{2}{2} \cdot (2 + 1) \cdot 12) + (2 \cdot 12 \cdot 2) =$$

$$H_1: P(y)P(x_1, x_4 \mid y)P(x_2 \mid y)P(x_3 \mid y)P(x_5 \mid y) = 1 + (5 \cdot 2) + (2 \cdot 1 + 1) + (2 \cdot 1 + 1) + (2 \cdot 1) =$$
$$H_2: P(y)P(x_1 \mid y)P(x_2 \mid y)P(x_3 \mid y)P(x_4 \mid y)P(x_5 \mid y) = 1 + (2 \cdot 2) + (2 \cdot 1 + 1) + (2 \cdot 1 + 1) + (2 \cdot 1)$$
$$Ukupno: 1 + 106 + 10 + 3 + 3 + 2 + 4 + 2 = 131$$***??***
Ne zbrajamo samo parametre pojedinih modela jer se neki parametri ponavljaju u vise modela. (tocno 64)

n = 5, K = 2
H0: parametara = n(n + 1)/2 + n*K = 15 + 10 = 25 (formula s 15. str. 15. skriptice) ne mozes bas ovo raditi za svih 5 znacajki, jer su ti samo 2 kontinuirane, dok su ostale diskretne

*H1: parametara = ((6 - 1) + (2*2 - 1) + (2 - 1))*K + K - 1 = (5 + 3 + 1)*2 + 1 = 19 (formula sa 7. str. 16. skriptice + umnožak jer su x2 i x3 združene)*

*H2: parametara= ((3 - 1) + (2*2 - 1) + (2 - 1) + (2 - 1))*K + K - 1 = (2 + 3 + 1 + 1)*2 + 1 = 15 (formula sa 7. str. 16. skriptice)*

Ukupno 25 + 19 + 15 = 59 parametara???

ODGOVOR:

H0 - za numeričke značajke x2 i x3 imamo kovarijacijsku matricu 2x2, dakle n=2 i to je 2/2(2+1) parametara, a s obzirom da je dijeljena matrica, nije potrebno množiti to s brojem klasa (ista matrica za y=0 i y=1). Za x2 i x3 imamo još i mi (jer je za x2 i x3 je dvodimenzionalan), a potreban nam je za obje klase, dakle još 2*2 parametara. S obzirom da su sve diskretne varijable zavisne, moramo gledati sve moguće kombinacije istih minus jedna koju možemo dobiti kao 1 - suma ostalih. Kombinacija ima 3*2*2 (jer x1 ima 3 mogućnosti, x4 2 mogućnosti, a x5 isto 2 mogućnosti). Oduzmemo od toga 1 i time smo dobili broj parametara za jednu klasu, no imamo y=0 i y=1 pa množimo dobiveni broj s 2. Još je potreban jedan parametar za P(y=0) (a P(y=1) računamo kao 1 - P(y=0)). Time za H0 dobivamo 2*(3*2*2-1)+(2/2*(2+1)+2*2)+1 = 30*

H1 - x1 i x4 su združene, dakle zavisne. Time imamo isti slučaj kao u H1 i zato imamo 2(3*2 -1) parametara za taj dio (prvih 2* zbog K=2, 3*2 jer x1 ima 3 mogućnosti i jer x4 ima 2 mogućnosti i -1 jer znamo da je suma = 1). Kod kontinuiranih značajki x2 i x3 imamo dijeljenu i dijagonalnu kovarijacijsku matricu - samim time samo 2 dijagonalna parametra u njoj i još za mi 2*2 parametra. Za x5 nam je još potrebno 1 parametar (za P(x5=0) jer P(x5=1) = 1 - P(x5=0)) i to 2 puta zbog K=2. I ponovno 1 parametar za P(y=0). Ukupno 19 parametara -> 2*(3*2-1) + 2 + 2*2 + 1*2 +1 = 19*

*H2 - Za numeričke varijable ponovno imamo dijagonalnu dijeljenu kovarijacijsku matricu - dakle 2 parametra i 2*2 parametra za mi. Za x1, x4 i x5 imamo za svaku K(k-1) parametra koji trebamo definirati (npr za x1 P(x1=0) i P(x1=1), a P(x1=2)=1-(P(x1=0)+P(x1=1)), dakle 2+1+1, no to moramo i za y=0 i za y=1, pa to ide *2. Još je potrebno i parametar za P(y=0). Ukupno je to: 2*(2+1+1)+2+2*2+1 = 15. Sveukupno 30+19+15=64.*

11. *Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar neke klase, to jest $x_j \perp x_k | y$. Međutim, u stvarnosti ta pretpostavka rijetko kada vrijedi. Kao primjer, razmotrite model za klasifikaciju novinskih članaka, čija je zadaća odrediti je li tema članka pandemija koronavirusa ili ne. Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u novinskom članku. Na primjer, izglednost P(stožer|y=1) jest vjerojatnost da se u članku koji je na temu pandemije koronavirusa pojavi riječ ``stožer''.*

    Razmotrite sljedeće četiri riječi koje se općenito mogu pojaviti u novinskim člancima: ``stožer'', ``pandemija'', ``koronavirus'' i ``general''.

Za koju od sljedećih jednakosti općenito očekujemo da ne vrijedi i da se time onda narušava pretpostavka naivnog Bayesovog klasifikatora?

P(general|y=0)=P(general|stožer,y=0) -> Kužim da ova ne vrijedi, al ne kužim zašto ostale jednakosti vrijede.
Gornja jednakost ne vrijedi jer govori: ako članak nije na temu koronavirusa, pojava riječi stožer ne utječe na pojavu riječi general. S obzirom da je stožer vojni pojam kao i general, naravno da utječe.
P(koronavirus|y=0)=P(koronavirus|general,y=0) < U člancima koji ne govore o pandemiji koronavirusa je općenito vjerojatnost da je spomenut "koronavirus" jednaka nuli.
P(pandemija|y=1)=P(stožer|y=1) < Ako članak u našim medijima govori o pandemiji koronavirusa, za očekivati je da sadrži riječi "pandemija" i "stožer" jednako često jer su oboje ključni pojmovi u situaciji.
Alternativno, ako članak govori o pandemiji koronavirusa, ne možemo tvrditi da gornje riječi nemaju jednaku vjerojatnost pojavljivanja. Jednostavno ne znamo te vjerojatnosti i moguće je da jednakost vrijedi pa ju u općenitom slučaju ne možemo odbaciti.
P(stožer|y=1)=P(stožer|pandemija,y=1) Za riječi u člancima o pandemiji (y = 1), izričito spominjanje pojma "pandemija" ne bi trebalo mijenjati vjerojatnosti nalaženja drugih riječi u članku jer ne donosi nikakvu novu informaciju; iz y = 1 već znamo da se radi o pandemiji.

11. Kako ide ovaj?

(N) Treniramo polunaivan Bayesov klasifikator sa $n = 3$ binarne varijable, x_1, x_2 i x_3. Zajednička vjerojatnost tih triju varijabli definirana je sljedećom tablicom:

	$x_3 = 0$		$x_3 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	0.2	0.1	0.1	0.0
$x_1 = 1$	0.3	0.0	0.2	0.1

Prije treniranja klasifikatora, koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Odlučili smo združiti onaj par varijabli koje imaju uzajamnu informaciju veću od 0.01. Ako to vrijedi za dva para varijabli, onda ćemo sve tri varijable združiti u jedan faktor.

Izračunajte uzajamne informacije između svih parova varijabli te odredite koje varijable ćemo združiti u zajedničke faktore prema gornjem pravilu.

Kako glasi faktorizacija zajedničke vjerojatnosti tog polunaivnog Bayesovog klasifikatora?

- $P(y)P(x_1, x_3|y)P(x_2|y)$ ✔
- $P(y)P(x_1|y)P(x_2|y)P(x_3|y)$
- $P(y)P(x_1, x_2|y)P(x_3|y)$
- $P(y)P(x_1, x_2, x_3|y)$

Uzajamna informacija:
I(x_i, x_j) = D_kl(P(x_i, x_j) || P(x_i) P(x_j)) = sum_x' sum_y' P(x', y') ln(P(x', y') / (P(x') * P(y'))
Formula u skripti Bayesov klasifikator 2 pri kraju negdje mozda cak i u onim natuknicama

Sad se racuna P(x1=1, x2=1) = sum_x3 P(x1=1, x2=1, x3) (Marginalizacija)
I tako za sve ostale kombinacije.
Isto se moze napraviti i za P(x=1) samo se onda jos marginalizira i po y odnosno
P(x=1) =sum_x3 sum_x2 P(x1=1, x2, x3)

Dobije se P(x=1) = 0.6 , P(x2=1)=0.8, P(x3=1) = 0.6
I onda se samo ovrstava u formulu. Npr.
I(x1, x2) = 0.1 * ln(0.1 / (0.6 * 0.2)) + 0.5 * ln(0.5 / (0.6 * 0.8)) + 0.1 * ln(0.1 / (0.4 * 0.2)) + 0.3 * ln(0.3 / (0.4*0.8)) = 0.005
Sada kako uzajamna informaicja predstavlja kb divergenciju izmedju P(x, y) i P(x) * P(y) govori nam kolko su ove dvije distribucije slicne. Veca vrijednost oznacava da su udaljenije odnosno

kada je I veci onda su ove dvije distribucije udaljenije te ih zato spajamo odnosno nisu nezavisne (tj. Nezavisne su samo kada je P(x, y) = P(x) * P(y), ali u nasem primjeru mi zelimo da budu sto manje zavisne pa smo zato postavili na 0.01 granicu gdje dopustamo nekumalu zavisnost, ali ne preveliku. U stvarnosti su primjeri skoro uvijek malo zavisni, ali ako ne stavimo tu neku marginu onda model postaje prekompleksan.. Objasnjeno na predavanju). Za I(x1, x3) = 0.03 uzajamna informacija je veca od 0.01 te su te znacajke zavisne. I(x2, x3) = 0.005 pa one takoder nisu zavisne. Na kraju zavrsavamo s prvom formulom odnosno zdruzujemo x1 i x3.

SideNote Posto D_kl racuna log(P / Q) onda ce te dvije distribucije biti najslicnije kada je D_kl 0 jer je log(1) = 0. Isto vrijedi i za I. Odnosno I >= 0.

Kako ovaj? Gore je naveden kao 6 Hvala!

(N) Treniramo naivan Bayesov model za binarnu klasifikaciju \emph{``Skupo ljetovanje na Jadranu''}. Skup primjera za učenje je sljedeći:

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

Procjene parametara radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija sljedećeg primjera:

$$\mathbf{x} = (\text{Istra}, \text{ne}, \text{kamp}, \text{bus})$$

Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?

- 0.1747
- 0.6856
- 0.0032
- 0.3144

Cjelina 4B: Probabilistički grafički modeli

Hoćemo li sva pitanja iz kviza tu staviti pa neka svatko objasni nešto pa da imamo kompletno? sounds fair :) Jesu li sad sva pitanja iz kviza 4B ovdje? Već je stigao i 6 kviz :D

1.

P(P=1) = P(S=1) = x ***P(P=0)=P(S=0) = (1-x)***

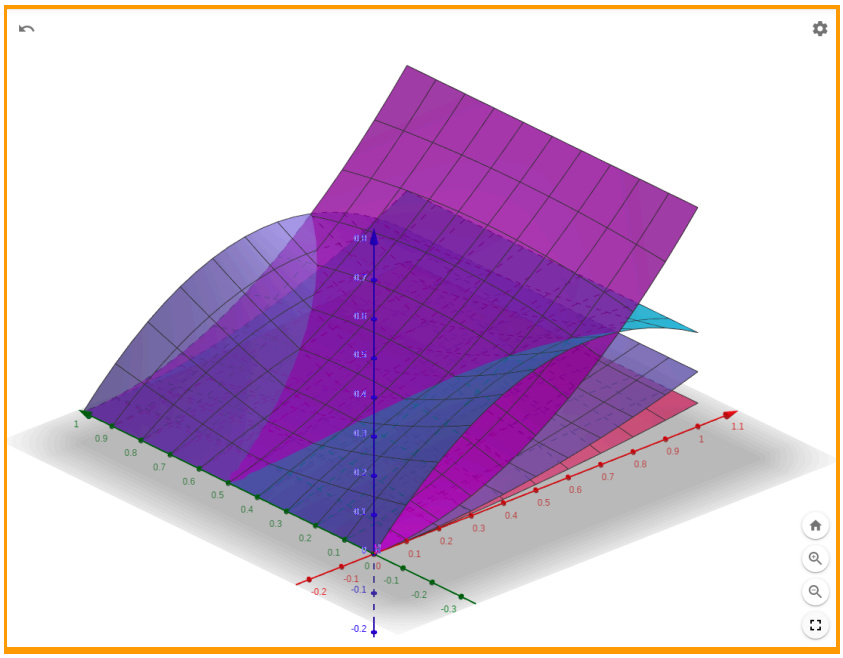
P(T=1|P=0,S=1) = P(D=1|P=0,S=1)= z ***P(T=0|P=0,S=1) = P(D=0|P=0,S=1)= (1-z)***
P(T=1|P=1,S=1) = 0.2 ***P(T=0|P=1,S=1) = 0.8***
P(D=1|P=1,S=1) = 0.7 ***P(D=0|P=1,S=1) = 0.3***

Argmax P(T,D|S=1) = ?
P(T,D,S,P) = P(P) P(S) P(D|P,S) P(T|P,S)
P(T=0,D=0|S=1) = ∑ P(T=0,D=0,P,S=1) / ∑ P(T,D,P,S=1)
P(T=0,D=1|S=1) = ∑ P(T=0,D=1,P,S=1) / ∑ P(T,D,P,S=1)
P(T=1,D=0|S=1) = ∑ P(T=1,D=0,P,S=1) / ∑ P(T,D,P,S=1)
P(T=1,D=1|S=1) = ∑ P(T=1,D=1,P,S=1) / ∑ P(T,D,P,S=1)

P(T=0,D=0,P,S=1) = (1-x)*x*(1-z)^2 + x^2*0.3*0.8 (0.24)
P(T=0,D=1,P,S=1) = (1-x)*x*z*(1-z) + x^2*0.7*0.8 (0.56)
P(T=1,D=0,P,S=1) = (1-x)*x*z*(1-z) + x^2*0.3*0.2 (0.06)
P(T=1,D=1,P,S=1) = (1-x)*x*z^2 + x^2*0.7*0.2 (0.14)

Zato sto je 0.56 najveci broj sam izabrala T=0 i D=1, ali zapravo ne znam koliki je 1. pribrojnik pa nije skroz tocan zadatak, samo mi se posrecilo.
Može li možda netko objasniti rješenje ovog zadatka (a da nije samo sreća u pitanju haha)?

Ubacio sam ove funkcije u Geogebru i cini se da MAP ovisi o vjerojatnostima x i z (slika dolje). Zadatak nije dobro postavljen? Mislim da je trebalo pisati da je z = 0.5, onda bi imalo smisla



(P) Razmotrite Bayesovu mrežu koja zajedničku vjerojatnost faktorizira na sljedeći način:

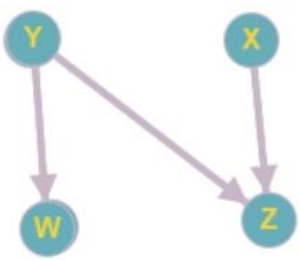
$$P(x, y, z, w) = P(x)P(y)P(z|x, y)P(w|y)$$

Odredite topološki uređaj varijabli. Ako postoji više mogućih topoloških uređaja, izaberite onaj koji po leksičkom poretku dolazi prvi (npr. x, y, z dolazi prije x, z, y). Zatim primijenite uređajno Markovljevo svojstvo te izvedite sve uvjetne nezavisnosti koje su kodirane u ovoj Bayesovoj mreži.

Koje sve uvjetne nezavisnosti vrijede u ovoj Bayesovoj mreži?

- $x \perp y,\ w \perp \{x, z\}|y$
- $y \perp x,\ w \perp x|\{y, z\}$
- $x \perp y|z,\ z \perp w|y$ ✖
- $x \perp y,\ z \perp w,\ x \perp w|\{z, y\}$

Kako se čitaju ove oznake u rješenjima? Recimo, za ovaj treći ponuđeni. Ja sam to ovako: "x je neovisan od y uz uvjet z. Z je neovisan od w uz uvjet y". To mi izgleda točno, ali nešto krivo radim. Ovo bi bila slika mreže:



Odgovor:

X i Y su nezavisni da, e sad w bi morao ovisiti o x,y i z da bi faktorizacija dala zajedničku vjerojatnost, ali mi smo pretpostavili da je w uvjetno

neovisan o x i z uz uvjet y i to je točan odgovor. Mora se voditi računa i o topološkom poretku, z je prije w prema tome w ne može utjecati na z jer nije predak od z

Iz kojeg razloga je z prije w? Jer je tim redom zapisano u formuli P(x,y,z,w)?
Topološki uređaj je XYZW, i onda samo gledamo po formuli za uređajno Markovljevo svojstvo: xk _|_ pred(xk) \ pa(xk) | pa(xk) i dobijemo 1. odgovor?

o

(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:

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

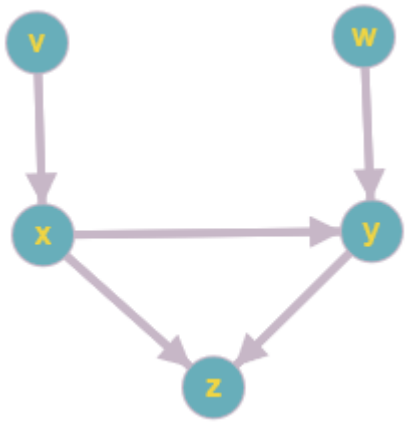
Izvedite faktorizaciju zajedničke distribucije koja odgovara ovoj Bayesovoj mreži. **Koliko parametara ima dotična Bayesova mreža?**

- 27
- 22
- 25
- 10

Obriši moj odabir

Kako procijeniti broj parametara iz faktorizacije?

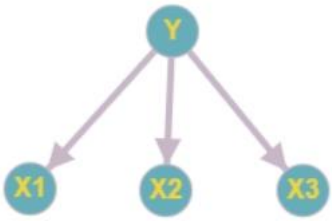
Broj parametara za jedan čvor u mreži je jednak umnošku vrijednosti varijabli oa kojima ovisi i broja svojih vrijednosti umanjenog za 1. Na primjer, ako je a binarna, a b i c ternarni, broj parametara za P(c | a, b) je 2 × 3 × (3 - 1) = 12.
Zadatak možete riješiti tako da nacrtate Bayesovu mrežu i primijenite pravilo iznad na izračun broja parametara.



Zajednicka distribucija glasi: P(x,y,z,v,w) = P(v) P(w) P(x|v) P(y|w,x) P(z|x,y)
*Broj parametara je onda: (2-1) + (2-1) + (3-1)*2 + (3-1)*2*3 + (2-1)*3*3 = 27*
Mislim da će ovdje pomoći ovaj alat => https://graphonline.ru/en/ ; zgodno je za crtati grafove pa možemo samo sliku staviti u objašnjenje za PGM-ove

P(v, w, x, y, z) = P(v) P(w | v) P(x | v, w) P(y | v, w, x) P(z | v, w, x, y) *(pravilo lanca)*
= P(v) P(w) P(x | v, w) P(y | v, w, x) P(z | v, w, x, y) *(marginalna nezavisnost v, w)*
= P(v) P(w) P(x | v) P(y | v, w, x) P(z | v, w, x, y) *(uvjetna nezavisnost w, x | v)*
= P(v) P(w) P(x | v) P(y | w, x) P(z | v, w, x, y) *(uvjetna nezavisnost v, y | {w, x})*
= P(v) P(w) P(x | v) P(y | w, x) P(z | x, y) *(uvjetna nezavisnost {v, w}, z | {x, y})*

Kako se računa broj parametara ovdje?
Ako dobro shvaćam, situacija za H1 je ovakva:

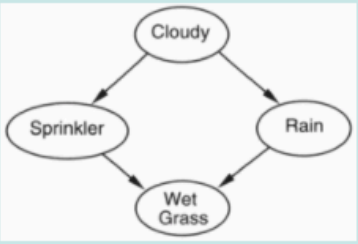


Sad je pitanje: Ako su sve 4 varijable binarne, koliko parametara ima ovaj model? Ovdje je rješavano ovom logikom:
*4 varijable, svaka ima max dvije vrijednosti, tj. Poveznica neke dvije je 4. I sad 3*4 = 12.*

Ja sam računala ovako:
*(2-1) + (2-1)*2 + (2-1)*2 + (2-1)*2 = 7*
*Svaki pribrojnik je jedan cvor, a broj 2 predstavlja vrijednost (binarna u ovom slucaju). Znaci mnozis za svaki cvor (vrijednost-1)*roditelji. Npr za x1 je (2-1)*2 jer ima samo jednog roditelja, y. Za H2 dobijem 9, a za H3 15. Pa mi je rjesenje 2 vise, 6 manje.*

*A za aciklički usmjereni graf… Ako bi na idućoj slici bila još usmjerena poveznica iz x1 u x3, bi li to bilo dobro? I kako se onda dobije 15 za H3? Ja sam ovako: (2-1) + (2-1)*2 + (2-1)*4 + (2-1)*4 = 13. Šta mi fali?*
*Odgovor: Y: 2-1, X1: (2-1)*2*2, X2: (2-1)*2, X3: (2-1)*2*2*2 => x3 ovisi onda o y, x1 i x2 što znači da se množi s 8*

(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.

Ima li netko postupak za ovaj :) ?

-> Raspises zajednicku vjerojatnost kao P(C, S, R, W) = P(C)*P(S|C)*P(R|C)*P(W|S,R). Trazimo vjerojatnost da pada kisa, dakle R=1, uz opazene varijable W=1 i C=0. Dakle P(R=1|C=0,W=1). Prema skripti to mozemo rapisati kao P(R=1|C=0,W=1) = P(R=1,C=0,W=1) / P(C=0,W=1) sto dalje raspisujemo kao ∑P(C=0,S,R=1,W=1) / ∑P(C=0,S,R,W=1). Kako bi izračunali ∑P(C=0,S,R=1,W=1) sumiramo po svim mogucim vrijednostima S, odnosno to je P(C=0,S=0,R=1,W=1) + P(C=0,S=1,R=1,W=1). Te vrijednosti uvrstavamo gore u izraz za zajedničku vjerojatnosti i samo citamo tablicu. Dakle za S=0 to bi bilo P(C=0,S=0,R=1,W=1) = P(C=0)*P(S=0|C=0)*P(R=1|C=0)*P(W=1|S=0,R=1). Citajuci iz tablica to je 0.5 * 0.5 * 0.2 * 0.9. Za S=1 bi islo identicno i kada to zbrojimo u brojniku imamo 0.0945. U nazivniku je malo vise posla jer ne idemo samo po S nego i po R, odnosno trebamo izracunati i sumirati P(C=0,S,R,W=1) za

svaki od parova S,R (0,0), (0,1), (1,0), (1,1), ali ide na identican nacin kao i gornji postupak. Ja sam u nazivniku dobio 0.1145 cisto da si znate provjeriti.
To je sve u 18. Skripti pod 1. Zakljucivanje, cak je i ista zajednicka vjerojatnost.

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

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

Parametre modela procjenjujemo MAP-procjeniteljem sa $\alpha = \beta = 2$ (za binarne varijable) odnosno $\alpha_k = 2$ (za ternarnu varijablu), što je istovjetno Laplaceovom zaglađivanju MLE procjene.

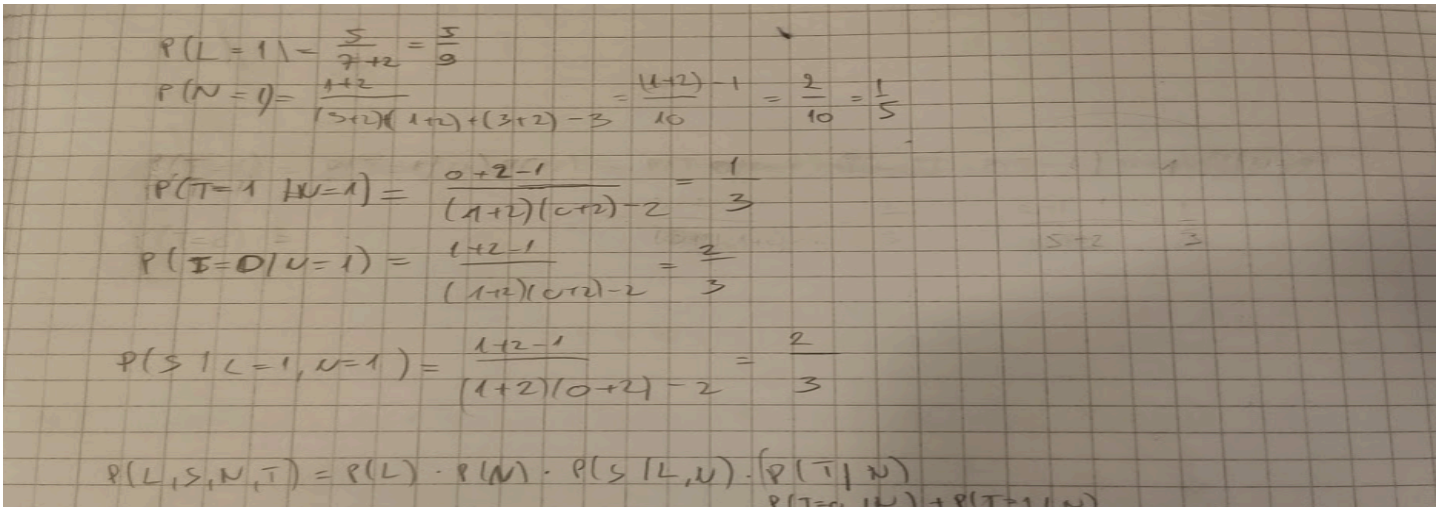
Na kraju nas, naravno, zanima koja je vjerojatnost života uz ljubav, sreću i malo novaca. Napravite potrebne MAP-procjene parametara.

Koliko iznosi zajednička vjerojatnost $P(L = 1, S = 1, N = 1)$?

- 0.023
- 0.143
- 0.833 ✖
- 0.074

/

ODGOVOR:
P(L)*P(N)*P(S|L,N)*P(T|N)
P(L) = (4+1)/(7+2) = 5/9
P(N) = (1+1)/(7+3) = 2/10
P(S=1|L=1,N=1) = (1+1)/(1+2) = 2/3
P(T=0|N=1)= (1+1)/(1+2) = 2/3
P(T=1|N=1)= (0+1)/(1+2) = 1/3



To sve izmnožiš i dobiješ 0.074. Ali ček, pa meni

ispada 0.01646 kad to izmnožim. Jesam jedina? :P

Dap, ovdje nešto ne štima. Može help?
=> štima sve jedino je ostalo ovo pitanje ispod
Na slici je točan rezultat ali zašto su zbrojene ove dvije vj?
Može se i bez toga jer gledamo sve vrijednosti varijable T i to se zbraja na 1. T i S su uvjetno nezavisne uz N.
Zasto se u racunu uopce uzima u obzir T ako trebamo tizracunati p(l,s,n) ?

(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 $P(y|x = 1, z = 0)$. Uzorkovanje smo ponovili ukupno $N = 1000$ puta.

Koja je očekivana veličina uzorka, odnosno koliko slučajnih vektora nećemo morati odbaciti?

- 54
- 200
- 739
- 124 ✔

Zna možda netko ovaj objasniti? Kako se dobije 124?
*N * P(x=1) * P(z=0)*
*N * (P(x=1) * (P(y=1)*P(z=0|x=1,y=1) + P(y=0)*P(z=0|x=1,y=0)) -- odakle ova formula*
*?1000 * (0.2 * (0.3 *0.9 + 0.7 *0.5)) = 124*

Nisam siguran jel ima nesto direkt iz skriptice, ali ja sam to common sense... pocnes s 1000 uzoraka. njih P(x=0) ce fejlat odmah jer ce iz cvora x krenuti kao false, a vidi da te pita za x=1,z=0.

Da bi dobio ovaj drugi dio z=0, marginaliziras po y (za svaku mogucnost y-a izvidis kolka je vjerojatnost da z bude 0) i samo pomnozis

Pitanje **4**
Nije još odgovoreno
Broj bodova od 1,00
Označi pitanje

(P) Bayesovom mrežom s pet binarnih varijabli modeliramo prometne prilike u gradu Zagrebu. U našoj mreži, jutarnje doba dana (J) i loše vrijeme (V) utječu na nastanak prometne gužve (G), u smislu da oba događaja povećavaju vjerojatnost nastanka prometne gužve. Loše vrijeme također utječe na nastupanje prometne nesreće (N), u smislu da povećava vjerojatnost prometne nesreće. Nadalje, nastupanje prometne nesreće utječe na nastanak prometne gužve, u smislu da povećava vjerojatnost nastanka prometne gužve. Loše vrijeme također utječe na zastoj tramvaja (T), u smislu da povećava vjerojatnost zastoja tramvaja. Međutim, nestanak struje (S) također uzorkuje zastoj tramvaja. Konačno, zastoj tramvaja uzrokuje masovno pješačenje putnika (P), što opet povećava vjerojatnost prometne nesreće.

U ovom kauzalnom modelu može nastupiti efekt objašnjavanja. **Kako bi se efekt objašnjavanja konkretno manifestirao?**

- $P(V=1|P=1,N=1) < P(V=1|N=1)$
- $P(T=1|V=1,P=1) > P(T=1|V=1)$
- $P(V=1|P=1,T=1) < P(V=1|T=1)$
- $P(G=1|J=1,V=1) > P(G=1|J=1)$

Je li točan odg d)?
Ja mislim da je a) jer se vjerojatnost smanji kada djeluje i uvjet P=1 pa je to onda efekt objasnjavanja. Vrijedi li za b) i c) jednakost (a ne nejednakost izraza) ?
d) je logicki tocno, ali sam shvatila (mozda krivo) da efektom objasnjavanja se smanji vjerojatnost varijable ako postoji jos jedan uvjet, a pod d) se vjerojatnost povecava dodatnim u vjetom

Pitanje **6**
Nije još odgovoreno
Broj bodova od 1,00
Označi pitanje

(T) Procjena parametara Bayesove mreže temelji se na maksimizaciji log-izglednosti parametara pod modelom. Procjena parametara može biti bitno drugačija za slučaj potpunih podataka, gdje su sve varijable opažene, u odnosu na slučaj nepotpunih podataka, gdje u model trebamo uključiti skrivene ili latentne varijable.

Što je prednost procjene parametara kod potpunih podataka (modela bez skrivenih varijabli) u odnosu na nepotpune podatke (modela sa skrivenim varijablama)?

- Kod potpunih podataka MLE procjena parametara ima rješenje u zatvorenoj formi, dok MAP procjena nema, za razliku od modela sa skrivenim varijablama kod kojeg je situacija obrnuta, a k tome taj model ima još više parametara od modela bez skrivenih varijabli
- Kod potpunih podataka log-izglednost se dekomponira po strukturi mreže, pa parametre svake uvjetne distribucije možemo procijeniti nezavisno od drugih čvorova i u zatvorenoj formi, međutim parametara može biti više nego kod modela sa skrivenim varijablama
- Kod potpunih podataka maksimizacija log-izglednosti ima rješenje u zatvorenoj formi, ali samo ako su opažene varijable na početku niza po topološkom uređaju čvorova, za razliku od modela sa skrivenim varijablama kod kojega MLE procjenitelj ne postoji u zatvorenoj formi
- Kod potpunih podataka minimizacija funkcije log-izglednosti ima rješenje u zatvorenoj formi, ali funkcija nije konkavna, pa može imati više lokalnih optimuma, za razliku od modela sa skrivenim varijablama koji ima više parametara, ali konkavnu funkciju log-izglednosti

Točan odg je: b ←------- Za koji zadatak? Gornji ili donji?

(T) Za Bayesovu mrežu kažemo da je generativni i parametarski model. **Zašto?**

- Generativni jer se može koristiti za generiranje skupa primjera na temelju zajedničke distribucije, a parametarski jer su broj čvorova mreže i njihovo povezivanje (dakle graf) definirani parametrima koji se mogu ugađati na skupu za učenje, čime se mogu dobiti različite strukture mreže
- Generativni jer svaki čvor odgovara uvjetnoj vjerojatnosti koja je, na temelju Markovljevog uređajnog svojstva, generirana distribucijama čvorova roditelja, a parametarski jer Bayesova mreža zapravo definira zajedničku distribuciju koja je opisana skupom parametara
- Generativni jer definira zajedničku vjerojatnost svih varijabli, i opaženih i skrivenih, a parametarski jer se parametri modela mogu dobiti MLE-procjenom za svaki čvor Bayesove mreže zasebno, budući da se log-izglednost dekomponira po strukturi mreže
- Generativni jer opisuje postupak kojim se mogu generirati podatci koji se pokoravaju određenoj zajedničkoj vjerojatnosnoj distribuciji, a parametarski jer svaki čvor Bayesove mreže definira uvjetnu vjerojatnost preko teorijske distribucije koja je opisana svojim parametrima

tocno: d)

Može objašnjenje za ovaj?

(P) Gaussov Bayesov klasifikator i logistička regresija su generativno-diskriminativni par modela, što znači da, uz prikladan odabir parametara, oba modela mogu ostvariti identičnu granicu u ulaznome prostoru. Međutim, Gaussov Bayesov klasifikator je generativni model, dok je logistička regresija diskriminativan model, pa ta dva modela općenito imaju različit broj parametara. U pravilu, logistička regresija imat će manje parametara od njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora.

Razmotrite slučaj binarne klasifikacije u ulaznome prostoru dimenzije $n = 100$ pomoću modela logističke regresije i njoj odgovarajućeg modela Gaussovog Bayesovog klasifikatora. **Koliko će model Gaussovog Bayesovog klasifikatora imati više parametara od modela logističke regresije?**

- 200
- 5049
- 5150 ✔
- 10200

Logistička regresija ima n+1 parametara -> 101, a Bayesov klasifikator n/2*(n+1)+2*n+1 parametara -> 50*101+200+1 = 5251. Razlika je 5251 - 101 = 5150 (skriptica 16, str. 3 na dnu).

Cjelina 6: Vrednovanje modela

(N) Na ispitnome skupu od $N = 10$ primjera evaluiramo binarnu logističku regresiju. Stvarne oznake primjera $y^{(i)}$ i predikcije klasifikatora $h(\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

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

Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y, birajući s jednakom vjerojatnošću između oznaka $y = 0$ i $y = 1$. Za evaluaciju klasifikatora koristimo F_2-mjeru umjesto F_1-mjere. Izračunajte F_2-mjeru logističke regresije i referentnog modela.

Koliko će očekivano logistička regresija biti bolja od referentnog modela po F_2-mjeri?

- 0.051
- 0.176 ✔
- 0.101
- 0.095

Izračunam F2 od logističke regresije s P = 3/7 i R = 3/4. Ispadne 0.6522. I okej, ali dalje mi nije jasno kako bi za referentni model trebala odredit TP,FP,FN. Npr. za TP je 7 primjera s oznakom y=1. Ako klasifikator pogađa nasumično s jednakom vjerojatnošću, onda bi dobila ili 3 ili 4 s oznakama h(xi) = 1 pa bi mi TP bila 3/7 ili 4/7. Tako slično za odziv. Al šta bi onda dobili/izabrali/kako dalje...?

Ja mislim da je ovako: imaju 4 y koji su 1 i 6 y koji su 0. Od ovih 4 će 2 bit TP i 2 FN i od ovih 6 će 3 biti TN i 3 FP i onda se dobije P=0.4 i R=0.5 što daje F2=0.4762 i onda kad se oduzme 0.6522-0.4762 dobije se 0.176. Možda ne treba ovako, ali poklopilo se s rješenjem.ya

Moze netko potvrdit ovaj postupak (prvi iznad)?

Aaa <u>žimku</u>, zamijenila sam h umjesto y za klasifikator. Okej, sve jasno. Hvala ti papi

(P) Evaluiramo model L_2-regularizirane logističke regresije. Za evaluaciju koristimo ugniježđenu unakrsnu provjeru u kojoj optimiramo regularizacijski faktor λ. Neka je λ_1 prosjek optimalnih vrijednosti regularizacijskog faktora, i neka je F_1^1 prosječna F_1-mjera na ispitnom skupu vanjske petlje. Međutim, naknadno smo ustanovili da nam se potkrala pogreška i da smo u unutarnjoj petlji model uvijek ispitivali na prvom preklopu. Kada to ispravimo, dobivamo λ_2 kao prosjek optimalnih vrijednosti regularizacijskog faktora i F_1^2 kao prosjek F_1-mjere na ispitnom skupu vanjske petlje. Nažalost, kasnije smo ustanovili da nam se potkrala još jedna pogreška: umjesto da u vanjskoj petlji optimalan model treniramo na cijelom skupu za treniranje, mi smo ga trenirali samo na skupu za treniranje zadnje iteracije unutarnje petlje. Kada i tu pogrešku ispravimo, dobivamo λ_3 odnosno F_1^3.

Što možemo očekivati o odnosima između procjena za optimalni λ i za F_1-mjeru na ispitnom skupu?

- $\lambda_1 < \lambda_2 = \lambda_3, F_1^1 > F_1^2, F_1^3 > F_1^2$ ✔
- $\lambda_1 < \lambda_3, F_1^1 < F_1^2 < F_1^3$
- $\lambda_1 = \lambda_3 < \lambda_2, F_1^2 < F_1^1, F_1^3 < F_1^2$
- $\lambda_1 > \lambda_2 > \lambda_3, F_1^1 < F_1^2, F_1^3 < F_1^2$

Prosječna vrijednost hiperparametra λ se određuje u unutarnjoj petlji. $\lambda_2 = \lambda_3$ ***jer se unutarnja petlja ne mijenja, ali zašto je*** F_1^1 ***veći od*** F_1^2***?***

Mislim da je F_1^1 ***veći od*** F_1^2 ***iz razloga što se u unutarnjoj petlji stalno ispituje na istom skupu, a između ostaloga se i trenira na tim primjerima pa će score biti veći. Dakle, primjeri za treniranje se mijenjaju dok primjeri za ispitivanje ostaju isti, ali u jednom trenu i ti ispitni dođu na red kada postaju primjeri za treniranje pa onda model može dobiti bolji score pošto ih je već vidio.***

a zašto je F_1^3 ***veći od*** F_1^2***?***

Je li zato što je F_1^3 ***treniran na većem skupu i time je dobiven bolji faktor lambda?***

Pitanje **2**

Nije još odgovoreno

Broj bodova od 1,00

Označi pitanje

(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):

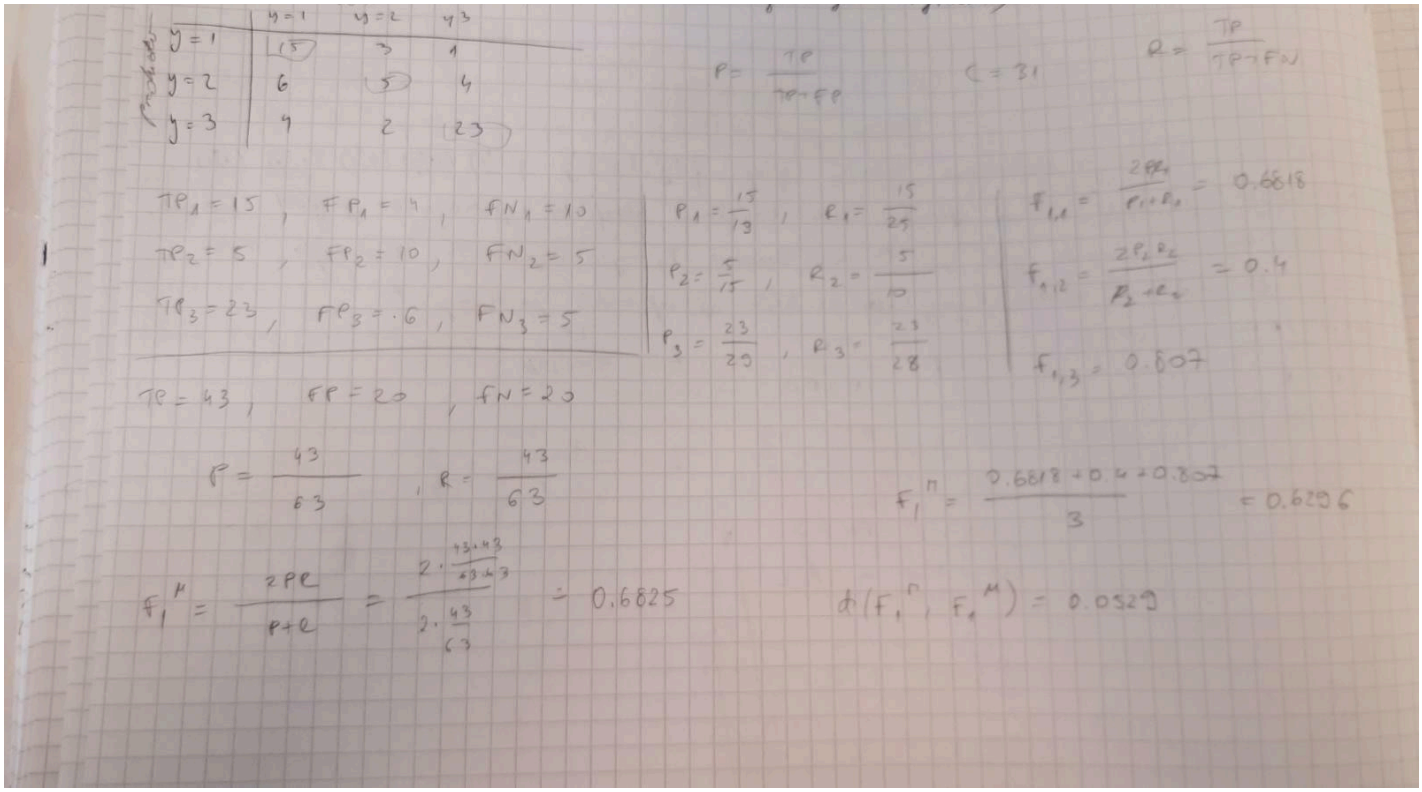
	$y=1$	$y=1$	$y=3$
$y=1$	15	3	1
$y=2$	6	5	4
$y=3$	4	2	23

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, $F_1^\mu - F_1^M$**?**

- 0.09
- 0.13
- 0.01
- 0.05

Ima u skriptici jedan slican primjer. Rjesen ispod.
Gdje u skriptici?
Natuknice, strana 3
Hvala papi



Pitanje **3**

Nije još odgovoreno

Broj bodova od 1,00

Označi pitanje

(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?

- 69201
- 35721
- 44640
- 49600

9*(992*4+1) … netko objasnjenje? Zasto je koristen br 9, a ne 10 u vanjskoj, te 4 a ne 5 u unutarnjoj petlji?

9 jer se svaki primjer koristi (10-1) puta za set za treniranje te 4 jer se unutar svakih 9/10 samo ⅘ koriste za treniranje te ⅕ za testiranje te jos +1 jer se koristi cijelih 9/10 za treniranje na kraju. Meni samo nije jasno zasto 992…

Rbf ima 961 kombinaciju, a linearna 31 pa je ukupno 992.
Hvala ti papi

DODATNO(ovdje se ne traži): svaki primjer će biti iskorišten za testiranje 9 puta. (provjeriti)
Kako dobijemo 961 i 31?
Linearna → optimiraš samo hiperparametar C koji može poprimiti 31 vrijednost
Rbf → optimiraš C i gamma, ALI moraš ih optimirati u parovima. C i gamma mogu imati svaki po 31 vrijednost, pa je onda ukupan broj parova 31*31 = 961

Pitanje **4**
Nije još odgovoreno
Broj bodova od 1,00
Označi pitanje

(P) Evaluirali smo model višeklasne logističke regresije i dobili da je vrijednost mjere makro-$F_{0.5}$ znatno manja od vrijednosti mjere makro-F_1, a da se vrijednosti mjere mikro-F_1 i mjere makro-F_1 znatno ne razlikuju.

Što možemo zaključiti o ovom modelu?

- Preciznost modela lošija je od odziva, ali model ne radi znatno lošije na primjerima iz klasa s manjim brojem primjera, pa su zbog toga mjere mikro-F_1 i makro-F_1 izjednačene
- Preciznost i odziv modela za pojedine klase znatno se ne razlikuju, neovisno o broju primjera u dotičnoj klasi, pa se onda znatno ne razlikuju niti mjere mikro-F_1 i makro-F_1
- Odziv modela lošiji je od preciznosti, osim na primjerima iz klasa s velikim brojem primjera, gdje se preciznost i odziv znatno ne razlikuju
- Model ima manji odziv na primjerima iz manjih klasa, ali veću preciznost na primjerima iz većih brojem primjera, pa su mjere mikro-F_1 i makro-F_1 izjednačene

Za beta<1 naglasak se daje na preciznost, dakle rezultat (Fbeta) bit će bliži vrijednosti koju ima preciznost. Kako je kod F0.5 vrijednost znatno manja nego kod F1, to znači da ju je preciznost privukla k sebi što znači da je preciznost modela lošija od odziva.
Ako model radi lošije na primjerima iz klase s manjim brojem primjera, onda će razlika između mikro i makro F1 mjere biti veća. To je zato što makro daje jednaku važnost svim klasama, dok kod mikro mjere klase s manjim brojem primjera imaju manju važnost.

< Cjelina 5. : GRUPIRANJE

(N) Algoritmom k-medoida (PAM) grupiramo $N = 5$ primjera. Za grupiranje koristimo mjeru različitosti, koja je za naših pet primjera definirana sljedećom matricom (matrica je simetrična, pa je donji trokut izostavljen):

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$
$\mathbf{x}^{(1)}$	0	0.2	0.9	0.7	0.5
$\mathbf{x}^{(2)}$		0	0.9	0.1	0.6
$\mathbf{x}^{(3)}$			0	0.7	0.3
$\mathbf{x}^{(4)}$				0	0.8
$\mathbf{x}^{(5)}$					0

Grupiramo u $K = 2$ grupe, s primjerima $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(5)}$ kao početnim medoidima.

Provedite prvu iteraciju algoritma k-medoida (PAM). **Koje medoide dobivamo nakon prve iteracije?**

- $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(2)}$
- $\mathbf{x}^{(1)}$ i $\mathbf{x}^{(3)}$
- $\mathbf{x}^{(3)}$ i $\mathbf{x}^{(4)}$
- $\mathbf{x}^{(2)}$ i $\mathbf{x}^{(5)}$ ✔

POSTUPAK:

Inicijalno grupiranje:

- ***x1 u grupi sa {x2,x4} -> GRUPA A***
- ***x5 u grupi sa {x3,x5} -> GRUPA B***

Primjeri su svrstani u grupu u kojoj su imali najmanju različitost sa početnim medioidom te grupe.

Primjer x4 smo stavili u grupu A, a razlika je 0.7 od medoida A, a da smo ga stavili u grupu B razlika bi bila 0.2 od medoida B. Zašto to nije krivo ako smo rekli da svrstavamo primjere prema najmanjoj različitosti?

Odg:Razlika bi bila 0.8, a ne 0.2

Pronalaženje novih medoida/reprezentativnih primjera

- ***GRUPA A***
  - ***Za reprezentativni primjer***
    - ***x1, J = 0.2 + 0.7 = 0.9***
    - ***x2, J = 0.2 + 0.1 = 0.3 => ovo je novi reprezentativni primjerak grupe A***
    - ***x4, J = 0.7 + 0.1 = 0.8***
- ***GRUPA B***
  - ***Za reprezentativni primjer***
    - ***x3, J = 0.3***
    - ***x5, J = 0.3 => ovo je novi/stari reprezentativni primjerak grupe B***

=> nakon prve iteracije, novi medoidi/reprezentativni primjeri grupa su x2 i x5

Kako se odreduju medoidi u sljedecoj iteraciji, koje sve primjere razmatramo kao potencijalne ?
Gledamo kombinacija od preostalih primjera koji trenutno nisu medioidi:
Izračunamo indikatore za svaki primjer na osnovu trenutnih medoida, tako da gledamo mjeru razlike u matrici između primjera i medoida svake klase, najmanja razlika određuje indikatorsku varijablu, npr za medoid x1 i primjer x2 razlika je 0.2 dok je za medoid x5 i primjer x2 razlika 0.6 prema tome b1 za x2 = 1, a b2 za x2 = 0, i tako za preostale primjere.

Sljedeće je izbor novih medoida, gdje su kandidati svi primjeri koji trenutno nisu medoidi, znači x2, x3, x4. Uzmemo x2 i razmatramo ga kao medoid za klasu 1, to rezultira sumom razlike x2 - ostali za sve primjere koji pripadaju klasi 1, njih prepoznajemo po b1 = 1, pa ponovno isti postupak za x3 i za x4. Kada su sve sume izračunate, biramo onaj x između x2, x3 i x4 koji je dao najmanju sumu.

E sad, ima jedan problem sa zadatkom, točno rješenje je x2 i x5 za nove medoide, međutim s obzirom na opisani algoritam, kako x5 može biti novi medoid, ako ne razmatramo trenutne medoide?

X5 zadrži najbolju vrijednost, pa se on ne mijenja, vidljivo iz donjeg opisa algoritma

PAM uses a greedy search which may not find the optimum solution, but it is faster than exhaustive search. It works as follows:

1. *(BUILD) Initialize: greedily select k of the n data points as the medoids to minimize the cost*
2. *Associate each data point to the closest medoid.*
3. *(SWAP) While the cost of the configuration decreases:*
    1. *For each medoid m, and for each non-medoid data point o:*
        1. *Consider the swap of m and o, and compute the cost change*
        2. *If the cost change is the current best, remember this m and o combination*
    2. *Perform the best swap of m best and o best, if it decreases the cost function. Otherwise, the algorithm terminates.*

(N) Želimo grupirati $N = 1000$ primjera, ali nemamo nikakvih saznanja o optimalnom broju grupa. Kako bismo odredili optimalan broj grupa, odlučili smo označiti uzorak primjera i na tom uzorku izračunati Randov indeks $\mathrm{RI}(K)$ za grupiranja dobivena s različitim brojem grupa K. Naposlijetku ćemo onda kao optimalan broj grupa odabrati onaj K koji maksimizira Randov indeks, $K^* = \mathrm{argmax}_K \mathrm{RI}(K)$. Budući da ne znamo koji je točan broj grupa, umjesto označavanja pojedinačnih primjera označavamo parove primjera. U tu svrhu smo iz skupa primjera uzorkovali 16 različitih primjera, uparili ih u 8 različitih parova primjera, te smo za svaki par primjera ručno označili trebaju li dotični primjeri pripadati istoj grupi ili ne. Rezultat označavanja je takav da tri para primjera trebaju pripadati istoj grupi (indeksi parova 1--3), a pet različitim grupama (indeksi parova 4--8). Nakon toga proveli smo grupiranje za $K \in \{3, 4, 5\}$ grupa. Za uzorak označenih primjera dobili smo ovakve grupe:

$$K = 3: \ \{1,1,2,4,8\}\ \{2,3,7\}\ \{4,5,3,5,6,6,7,8\}$$
$$K = 4: \ \{1,1,2\}\ \{4,8,4\}\ \{2,3,7,5,7\}\ \{3,5,6,6,8\}$$
$$K = 5: \ \{1,1\}\ \{3,4,8\}\ \{2,2,4\}\ \{7,5,7,3,5,6\}\ \{6,8\}$$

Brojke označavaju indeks para primjera. Na primjer, u grupiranju sa $K = 3$ grupe par primjera s indeksom 1 našao se u istoj grupi, a par primjera s indeksom 2 u različitim grupama.

Izračunajte Randov indeks $\mathrm{RI}(K)$ te optimalan broj grupa K^* prema Randovom indeksu, za $K \in \{3, 4, 5\}$.

Koliko iznosi Randov indeks za optimalan broj grupa, $\mathrm{RI}(K^*)$?

- 0.625
- 0.375
- 0.750
- 0.875

Mislim da je odgovor 0.625. Trebalo mi je previse vremena da shvatim sto se ustvari treba radit, al basically. R=(a+b)/(broj kombinacija). "a" se dobije tako da se pobroji u koliko slucajeva su brojevi 1,2 i 3 zajedno u grupi, a "b" tako da se pobroji koliko od 4-8 se nalazi u razlicitim grupama. R3=4/8, R4=⅜, R5=⅝ Najveci R je optimalan.
Je slazem se.

Nije li u nazivniku Randovog indeksa broj svih mogućih parova? O: piše ti da je 8 različitih parova primjera
Ovo je užasno napisan zadatak...

(T) Algoritmi grupiranja k-sredina i k-medoida razlikuju se, između ostaloga, i po vremenskoj računalnoj složenosti. Naime, algoritam k-medoida računalno je složeniji od algoritma k-sredina.

Zašto je algoritam k-medoida računalno složeniji od algoritma k-sredina?

- Kriterijska funkcija algoritma k-medoida jest mnogo složenija od one k-sredina, upravo zato što algoritam k-medoide koristi medoide, a ne centroide
- Za razliku od algoritma k-sredina koji se zasniva na euklidskoj udaljenosti, čiji je izračun računalno nezahtjevan, algoritam k-medoida koristi funkcije sličnosti čije računanje iziskuje mnogo računalnih operacija
- Za razliku od algoritma k-sredina, algoritam k-medoida je algoritam mekog grupiranja, što iziskuje provođenje dodatnih koraka unutar algoritma
- Budući da algoritam k-medoida ne koristi centroide, nego medoide, na kraju svake iteracije mora kombinatoričkom provjerom po primjerima pronaći medoide koje minimiziraju kriterijsku funkciju J

Koji je odgovor za ovo? Jel b) zato sto za svaki k moramo naci argmin za slicnosti izmedu svih primjera sto je vise neće imati nikakav značaj, ali je isto tako moguće da nam se otvori velika prilika za povratak u živooperacija nego samo udaljenost

Ja bih rekla da je d) s obzirom na diskusiju s laboratorijskih vježbi, sličnost nije toliko zahtjevnija nego udaljenost, već baš taj dio s kombinatoričkom provjerom

a) *Kriterijska funkcija nije složenija kod k-mediana.*
b) *Funkcija sličnosti može biti kompliciranija od euklidske udaljenosti, ali ne mora biti. Uz to, možemo spremiti f sličnosti za svaki par x-eva.*
c) *Ni jedan mi drugi algoritam ne grupiraju meko*
d) *točno*

(N) Raspolažemo sljedećim neoznačenim skupom primjera:

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

Primjere grupiramo algoritmom k-sredina sa $K = 2$ grupe. Za početna središta odabrali smo primjere $\mathbf{x}^{(2)} = (1,2)$ i $\mathbf{x}^{5} = (3,3)$. Provedite prvu iteraciju algoritma k-sredina.

Koliko iznosi vrijednost kriterijske funkcije J nakon ažuriranja centroida?

- 1.667
- 1.833
- 2.414
- 2.962

Izracunamo udaljenosti d(x2, xi) i d(x5, xi), i=1,3,4
Dobijemo da u klasu x2 pripadaju x1 i x3, a u klasu x5 pripada x4.
Sada racunamo nove centroide, za c1 racunamo na temelju x2, x1 i x3, a za c2 uzimamo x5 i x4 te se dobiju centroide c1(1.33, 1.67) i c2(2.5, 3). I sada jos treba izracunati kriterijsku fju tako da izracunamo euklidsku udaljenost od primjera iz prve klase do c1, i od primjera iz druge klase do c2, te sve to posumiramo. Na kraju dobijemo J = 1.833.

NOTE: Ako ste dobili rješenje J = 2.962, uzeli ste korijen pri izračunu udaljenosti dvije točke što je pogrešno (tj. nepotrebno jer se krati u kriterijskoj funkciji).
Može netko ovo objasniti? Zašto? https://en.wikipedia.org/wiki/Euclidean_distance Pa treba uzeti korijen?
Provjereno je računski, zbilja ispada točno bez korijena. E sad, gdje je greška? Da li možda default za k-sredina nije euklidska udaljenost? Da Euklidska udaljenost ima korijen, no u funkciji J ima ||a-b|| na kvadrat cime se efektivno ponistava taj korijen. Točno, hvala! Pretpostavljam da za udaljenost prilikom pridruživanja i dalje koristimo euklidsku a samo za J ovaj ||.||? Tako je, za J euklidska na kvadrat za svaku razliku vektora (kao što i formula kaze).

(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	2	1	0	0	2	1	2
$y_{true}^{(i)}$	1	1	0	2	0	0	1	1	1	2

Koliko iznosi Randov indeks grupiranja izračunat na ovom uzorku?

- 0.64
- 0.27
- 0.70
- 0.56

Ima li ovdje pametniji postupak ili se racuna kao na labosu usporedi (1-2,1-3,1-4...1-10, 2-3,2-4…)
Mislim da si to možeš nacrtat i dobiješ G1 sa 011, G2 sa 011 i G3 sa 0122 i onda gledaš parove za a i dobro odvojeno za b i dobi se a=3 (u g1 imaš dvije jedinice u g2 dvije jedinice i u g3 dvije dvojke pa je to 1+1+1) i b=22 (g1 i g2: ima u g1 nula dobro odvojena od dvije stvari u g2 pa je to 1*2 i imaju dvije jedinice u g1 dobro odvojene od jedne stvari u g2 pa je to 2*1 -> znači 1*2+2*1, onda tako i za g1g3 i g2g3) pa je to 25/45=0.56

Mislim da je na zadnjem? predavanju profesor rješavao zad s randovim indexom, ugl u toj zadaći s grupiranjem pa sam možete pogledat taj dio videa jer ne znam jel ovo dobro objašnjeno.
Računanje b:
Znači gledamo g1 i g2, g2 i g3 i zatim g1 i g3.
Gledamo jesu li primjeri iz g1 dobro odvojeni od onih iz g2 - imamo 011 u g1 i 001 u g2. Prvo gledamo nule: u g1 imamo jednu nulu koja je dobro odvojena od dvije jedinice u g2 pa je to 1*2, zatim gledamo jedinice i imamo dvije jedinice u g1 dobro odvojene od nule u g2 pa je to 2*1 i u g1 nemamo dvojke pa ništa.
Sada idemo g2 i g3. U g2 imamo 0 koja dobro odvojena od 122 u g3 znaci 1*3, imamo 11 dobro odvojeni od 022 pa je to 2*3 i nemamo dvojke u g2 pa smo gotovi.
Isto bi išlo za g1 i g3 al su g2 i g1 isti pa ista stvar kao i g2 i g3 - znači 1*3 za nulu, 2*3 za jedinice i 0 za dvojke.
Sada je b=(1*2+2*1+0)+(1*3+2*3+0)+(1*3+2*3+0)=22

ALTERNATIVNI POSTUPAK:
Rješenje se može dobiti preko matrice konfuzije. Randov index možemo izračunati kao: (TP+TN)/(TP+FP+FN+TN).
Matrica konfuzije izgleda:

	stvarno

		y = 0	y = 1	y = 2
model	y = 0	1	2	0
	y = 1	1	2	0
	y = 2	1	1	2

TP + FN = 3C2 + 5C2 + 2C2 = 14 (suma vrijednosti po stupcu povrh 2 za svaki stupac)
TP + FP = 3C2 + 3C2 + 4C2 = 12 (suma vrijednosti po retku povrh 2 za svaki redak)
TP = 2C2 + 2C2 + 2C2 = 3 (svaka celija povrh 2, uzimamo 0C2=0, 1C2 = 0 za svaku celiju)

Nisi li krivo racunao TP+FP (potrebno sumiranje po retcima) i TP+FN (potrebno sumiranje po stupcima)? U konacnici se dobije isti rez, ali ako se matrica konfuzije podijeli prema klasama, onda se lijepo vide odnosi pozicija:
TP | FP
FN | TN
Točno, u izračunu su zamjenjene vrijednosti FN i FP. ISPRAVLJENO, sada bi trebalo biti uredu.

Iz toga dobijemo TP = 3, FP = 9, FN = 11.

Još moramo odrediti TN.
TN = 10C2 - TP - FP - FN = 22 (10 je suma elemenata matrice konfuzije)

Konačno,
R = (TP+TN)/(TP+FP+FN+TN) = (3+22)/(3+9+11+22) = 0.555555555556

Ali zašto TP nije 5? Zar TP nije ako je stvarno = model? A po dijagonali u tablici vidimo da ih je 5 takvih?
Mislim da ne možeš tako gledati jer ovdje radimo sa parovima, vidiš npr. da u tablici ima ukupno 10 vrijednosti, a TN = 22, ne mozeš vrijednosti direktno isčitati iz tablice.

(P) Algoritam GMM koristimo za grupiranje $N = 10$ primjera u dvodimenzijskome ulaznom prostoru. Skup primjera koje grupiramo je sljedeći:

$$\mathcal{D} = \{(0,0), (1,1), (1,2), (2,2), (2,3), (5,0), (5,1), (6,0), (6,6), (7,7)\}$$

Razmatramo tri modela GMM:

$$\begin{aligned} \mathcal{H}_1: &\quad K = 2 \text{ grupa, puna kovarijacijska matrica} \\ \mathcal{H}_2: &\quad K = 2 \text{ grupa, izotropna kovarijacijska matrica} \\ \mathcal{H}_3: &\quad K = 3 \text{ grupe, izotropna kovarijacijska matrica} \end{aligned}$$

Za sva tri modela kovarijacijska matrica je nedijeljena, dakle svaka komponenta ima svoju kovarijacijsku matricu. Za početne centroide odabiremo nasumično dva odnos tri primjera iz $\mathcal{D}$, ovisno o broju grupa K. Za svaki model grupiranje ponavljamo 100 puta te kao konačno grupiranje uzimamo ono s najvećom log-izglednošću na sku $\mathcal{D}$.

Zanima nas kojoj grupi najvjerojatnije pripada primjer $\mathbf{x}^{(5)} = (2,3)$, to jest zanima nas k koji maksimizira odgovornost $h_k^{(5)} = P(y = k|\mathbf{x}^{(5)})$. Ta vrijednost će biti različita za ova tri modela. Označimo sa h_α maksimalnu odgovornost za primjer $\mathbf{x}^{(5)}$ u modelu $\mathcal{H}_\alpha$, to jest vjerojatnost pripadanja tog primjera najvjerojatnijoj grupi dobivenoj grupiranjem pomoću modela $\mathcal{H}_\alpha$.

Što možemo zaključiti o odgovornostima h_α za ova tri modela?

- $h_{\alpha_2} < h_{\alpha_1} < h_{\alpha_3}$
- $h_{\alpha_1} < h_{\alpha_2} < h_{\alpha_3}$
- $h_{\alpha_2} > h_{\alpha_1} > h_{\alpha_3}$
- $h_{\alpha_1} > h_{\alpha_2} > h_{\alpha_3}$

Ovaj odgovor mi je otprilike jasan zašto je točan (i njega sam odabrao), samo bih molio provjeru zaključivanja:
H1 > H2 jer je puna cov matrica (za isti K) daje složeniji model pa je izglednost grupe k veća, a koeficijenti mješavine (Pl k) su isti (i centroidi su isti) -> odgovornost hk je veća za H1 (model veće složenosti)?
H2 > H3 jer je za H3 K = 3 veći od K = 2 za H2 pa je Pl k manji za H3 (jer je masa vjerojatnosti raspoređena na više grupa), a izglednosti grupe k su isti jer je ista kovarijacijska matrica (i centroidi su isti) -> odgovornost hk je veća za H2 (model s manje komponenti K)? Kako znamo da su izglednosti iste, također zar nebi centroid u slučaju K=3 trebao biti bliže primjeru x5 od onog za K=2?

Moje razmisljanje je bilo da je h3>h1>h2 jer kad imamo 3 grupe onda je ova treca grupa jako gusta i zbijena pa je zato odgovornost h3 ful velika. Izmedu h2 i h1 uz pretpostavku da im je srediste negdje na sredini (tj. Na istom mjestu) odlucio sam se da je h1>h2 jer su izohipse elipticne za punu kov. Matricu pa je onda h1 blize sredistu.
dmin
Igra li ovdje ulogu to što se razmatra primjer (2, 3) ili bi bio isti rezultat za bilo koji drugi primjer?

(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):

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$
$\mathbf{x}^{(1)}$	0	0.4	0.5	0.7	0.5
$\mathbf{x}^{(2)}$		0	0.9	0.3	0.6
$\mathbf{x}^{(3)}$			0	0.7	0.1
$\mathbf{x}^{(4)}$				0	0.8
$\mathbf{x}^{(5)}$					0

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

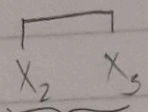
Kojem binarnom stablu odgovara dobiveni dendrogram?

- $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(4)}), (\mathbf{x}^{(5)}, \mathbf{x}^{(1)}))$
- $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), ((\mathbf{x}^{(4)}, \mathbf{x}^{(1)}), \mathbf{x}^{(5)}))$
- $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), \mathbf{x}^{(1)}), (\mathbf{x}^{(4)}, \mathbf{x}^{(5)}))$
- $((\mathbf{x}^{(2)}, \mathbf{x}^{(3)}), ((\mathbf{x}^{(4)}, \mathbf{x}^{(5)}), \mathbf{x}^{(1)}))$ ✔

Može objašnjenje kako se radi nova matrica nakon što smo povezali prva dva člana u stablu? Stavit ću svoj postupak, ne znam da li je ispravan ali se nadam da će netko ispraviti što sam pogriješio. (matricu sličnosti sam pisao pomnoženo s 10, za jednostavnost pisanja)

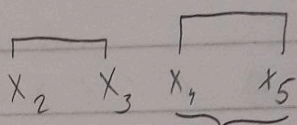
	x_1	x_2	x_3	x_4	x_5
x_1	0	4	5	7	5
x_2		0	9	3	6
x_3			0	7	1
x_4				0	8
x_5					0

x_2 I x_3 IMAJU NAJVEĆU SLIČNOST



	x_1	G_1	x_4	x_5
x_1	0	4	7	5
G_1		0	7	1
x_4			0	8
x_5				0

ANALOGNO ODABIR X_4, X_5



	x_1	G_1	G_2
x_1	0	4	7
G_1		0	7
G_2			0

- SPAJAMO LI X_1 S G_2 ILI G_1 S G_2?
- ZAD - "POTPUNO POVEZIVANJE"

$$\text{distance}(G_i, G_j) = \max_{\substack{x_i \in G_i \\ x_j \in G_j}} d(x_i, x_j)$$

- MAKSIMALNA UDALJENOST 2 TOČKE JE ANALOGNO KAD SU JAKO RAZLIČITE; NISKA SLIČNOST

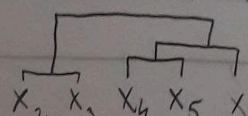
		G_1		G_2	
	x_1	x_2	x_3	x_4	x_5
x_1	0	4	5	7	5
G_1 x_2		0	9	3	6
G_1 x_3			0	7	1
G_2 x_4				0	8
G_2 x_5					0

$$\max d(x_1, G_2) = \max d(x_1, x_5) = 5$$

VS

$$\max d(G_1, G_2) = \max d(x_3, x_5) = 1$$

- POŠTO JE $5 > 1$ TJ. X_1 JE SLIČNIJI ZA NAJUDALJENIJI ČL. G_2 DOBIVAMO



Ovdje bi dobro došlo objašnjenje :)

(P) Primjere grupiramo algoritmom GMM. Koristimo nekoliko varijanti modela, koje se međusobno razlikuju po (1) broju grupa K, (2) načinu inicijalizacije središta (1 = slučajno, 2 = k-sredina, 3 = k-means++) te (3) po pretpostavci na kovarijacijsku matricu (1 = puna, 2 = dijagonalna, 3 = izotropna). Označimo sa $\mathcal{H}(K,i,k)$ model sa K grupa, inicijalizacijom i, pretpostavkom na kovarijacijsku matricu k. Na primjer, $\mathcal{H}(10,1,2)$ je model sa 10 grupa, sa slučajno inicijaliziranim središtima i s izotropnom kovarijacijskom matricom. Sa svakim modelom grupiranje ponavljamo 1000 puta i zatim za svaki model crtamo graf funkcije log-izglednosti kroz iteracije EM-algoritma, uprosječen kroz svih 1000 ponavljanja. Neka je $LL^0(K,i,k)$ prosječna log-izglednost za model $\mathcal{H}(K,i,k)$ na početku izvođenja EM-algoritma, a neka je $LL^*(K,i,k)$ prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma.

Što možemo unaprijed zaključiti o ovim log-izglednostima?

- $LL^*(K,i,k) \geq LL^0(K,i,k)$, $LL^0(K+1,i,k) > LL^*(K,i,k)$, $LL^*(K,3,k) \geq LL^*(K,2,k) \geq LL^*(K,1,k)$, $LL^0(K,i,1) \geq LL^0(K,i,2) \geq LL^0(K,i,3)$
- $LL^*(K,i,k) \leq LL^0(K,i,k)$, $LL^*(K,i,k) > LL^*(K+1,i,k)$, $LL^*(K,3,k) \geq LL^*(K,2,k) \geq LL^*(K,1,k)$, $LL^*(K,i,1) \leq LL^*(K,i,2) \leq LL^*(K,i,3)$
- $LL^*(K,i,k) \geq LL^0(K,i,k)$, $LL^*(K+1,i,k) > LL^*(K,i,k)$, $LL^*(K,3,k) \geq LL^*(K,2,k) \geq LL^*(K,1,k)$, $LL^*(K,i,1) \geq LL^*(K,i,2) \geq LL^*(K,i,3)$ ✔
- $LL^*(K,i,k) \geq LL^0(K,i,k)$, $LL^*(K,i,k) > LL^*(K+1,i,k)$, $LL^*(K,3,k) \geq LL^*(K,2,k) \geq LL^*(K,1,k)$, $LL^*(K,i,1) \leq LL^*(K,i,3) \leq LL^*(K,i,2)$

Složeniji modeli su točniji, a složenost raste:
Porastom K
Odabir početne sredine je K-means++ (pa zatim k-sredina, pa najlošiji za nasumičan odabir)
Kada je kovarijacijska matrica puna (pa zatim dijagonalna, pa izotropna)
Također je model točnij na kraju (LL*) nego na početku (LL0)

S obzirom na te zaključke dobivaš točan odgovor:
Točniji je na kraju (LL* > LL0); Točniji je za veći broj grupa (K+1 > K); itd.