

Procjena parametara II

1. (T) Funkcija izglednosti $\mathcal{L}(\theta|\mathcal{D})$ nije isto što i vjerojatnost. **Po čemu se izglednost razlikuje od vjerojatnosti?**

- A Funkcija izglednosti $\mathcal{L}(\theta|\mathcal{D})$ jednaka je gustoći vjerojatnosti $p(\mathcal{D}|\theta)$, samo što je izglednost funkcija parametara θ , dok je $p(\mathcal{D}|\theta)$ funkcija uzorka \mathcal{D}
- B Funkcija izglednosti $\mathcal{L}(\theta|\mathcal{D})$ jednaka je gustoći vjerojatnosti $p(\theta|\mathcal{D})$, ali, za razliku od gustoće vjerojatnosti, nije odozgo ograničena sa 1
- C Ako su podaci diskretni (kategoričke značajke), onda je funkcija izglednosti parametara θ isto što i zajednička vjerojatnost uzorka \mathcal{D} i parametara θ
- D Za razliku od vjerojatnosti, funkcija izglednosti $\mathcal{L}(\theta|\mathcal{D})$ je simetrična, u smislu da vrijedi $\mathcal{L}(\theta|\mathcal{D}) = p(\mathcal{D}|\theta)$

FJKA IZGLEDNOSTI

→ gustoća vjerojatnosti naših podataka; uz fiksirane

parametre θ

→ \mathcal{D} fiksiran

zato: $\mathcal{L}(\theta|\mathcal{D})$

→ parametru θ dodjele je vjerojatnost da iz populacije s param. θ izvučen uzorak \mathcal{D}

$$\mathcal{L} : \theta \mapsto p(\mathcal{D}|\theta)$$

→ vj. slupci pod. \mathcal{D} za dist. sparcm. θ

$$\mathcal{L}(\theta|\mathcal{D}) = \prod_{i=1}^N p(x^{(i)}|\theta) = p(\mathcal{D}|\theta) \quad \text{i.i.d.}$$

suprotno od:

$p(\mathcal{D}|\theta) \rightarrow$ ovisi samo o parametru θ

2. (N) Raspolažemo sljedećim skupom označenih primjera:

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

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

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

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

- A -22.60 B -8.68 C -8.76 D +0.48

$y=1$

$$X_1 = \{-2, -2, 1, 3\}$$

$$\hat{\mu}_1 = \frac{1}{N} \sum_{x \in X_1} x^{(i)} = \frac{-2 - 2 + 1 + 3}{4} = \emptyset \quad \text{↳ samo podskup}$$

$$\hat{\sigma}_1^2 = \frac{1}{N} \sum_{x \in X_1} (x^{(i)} - \hat{\mu}_1)^2 = \frac{4 + 4 + 1 + 9}{4} = \frac{18}{4} = 4,5$$

pretpostavka i.i.d.

$$\theta = \{\mu_1, \sigma_1^2\} \quad \text{log-izglednost}$$

$$\mathcal{L}(\mu_1, \sigma_1^2 | \mathcal{D}_{y=1}) = p(X_1 | \mu_1, \sigma_1^2) = \ln \prod_{i=1}^N p(x^{(i)} | \theta)$$

$$= \ln \prod_{x \in X_1} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2} \frac{(x - \mu_1)^2}{\sigma_1^2}\right\}$$

$$= \ln \prod_{x \in X_1} \frac{1}{\sqrt{2\pi}\frac{9}{2}} \exp\left\{-\frac{1}{2} \frac{\frac{x^{(i)2}}{9}}{\frac{9}{2}}\right\} - \frac{18}{9}$$

$$= \sum_{x \in X_1} \left(\ln \frac{1}{3\sqrt{\pi}} - \frac{x^{(i)2}}{9} \right) = \ln \frac{1}{3\sqrt{\pi}} - \frac{4}{9}$$

$$+ \ln \frac{1}{3\sqrt{\pi}} - \frac{4}{9} + \ln \frac{1}{3\sqrt{\pi}} - \frac{1}{9} + \ln \frac{1}{3\sqrt{\pi}} - \frac{1}{9} = -8,68$$

3. (P) Neka je $\mathcal{L}(\mu, \sigma^2 | \mathcal{D})$ log-izglednost parametara μ i σ^2 normalne distribucije izračunata nad uzorkom \mathcal{D} koji sadrži ukupno N opažanja normalne varijable x . Nadalje, neka su $(\mu_{MLE}, \sigma_{MLE}^2)$ parametri distribucije razdiobe procijenjeni MLE-om nad uzorkom \mathcal{D} , te neka je σ_{UB}^2 nepristrana procjena varijance, izračunata kao $\sigma_{UB}^2 = \frac{N}{N-1} \sigma_{MLE}^2$. Konačno, neka je \mathcal{D}' slučajno uzorkovan podskup uzorka \mathcal{D} , tj. $\mathcal{D}' \subset \mathcal{D}$, pri čemu je poduzorkovanje načinjeno nakon procjene parametara. Razmotrite sljedeće četiri vrijednosti funkcije log-izglednosti $\mathcal{L}(\mu, \sigma^2 | \mathcal{D})$:

$$\begin{aligned}\mathcal{L}_0 &= \mathcal{L}(\mu_{MLE}, \sigma_{MLE}^2 | \mathcal{D}) && \text{podjednje} \\ \mathcal{L}_1 &= \mathcal{L}(0, 1 | \mathcal{D}) \\ \mathcal{L}_2 &= \mathcal{L}(\mu_{MLE}, \sigma_{UB}^2 | \mathcal{D}) && \text{nepristrana procjena} \\ \mathcal{L}_3 &= \mathcal{L}(\mu_{MLE}, \sigma_{UB}^2 | \mathcal{D}') && \text{varijance}\end{aligned}$$

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

- A $\underline{\mathcal{L}_0 > \mathcal{L}_1}$, $\mathcal{L}_2 \geq \mathcal{L}_3$
- B $\underline{\mathcal{L}_1 \geq \mathcal{L}_0}$, $\mathcal{L}_2 \geq \mathcal{L}_3$
- C $\mathcal{L}_0 < \mathcal{L}_3$, $\underline{\mathcal{L}_0 \leq \mathcal{L}_2}$
- D $\underline{\mathcal{L}_0 \geq \mathcal{L}_1}$, $\underline{\mathcal{L}_0 > \mathcal{L}_2} > \mathcal{L}_3$

OK OK

$\mathcal{L}_0 \rightarrow$ ne znamo jesu

li to parav. logi

0,1 → mogu, ali NE MORAJU BITI MLE

nivo mališinštva izglednost : $\mathcal{L}_0 \leq \mathcal{L}_1$

MLE → po def. nivo maks. izglednost

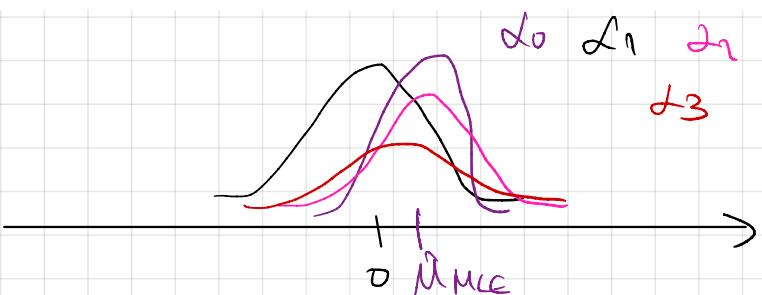
$\Rightarrow \mathcal{L}_2$: logl. za parav. logi NISU MLE : $\mathcal{L}_2 < \mathcal{L}_0$

\mathcal{L}_2 : izglednost na \mathcal{D}

\mathcal{L}_3 : podskup od \mathcal{D} (nismo mogli) ; isti parav.

→ moguće gesticirati vi. koje se moguće ; Ali NISMOSIG.

JESU I MANJE OD 1 (obično jesu) → mogući slup → vrednost



4. (T) MAP-procjenitelj definiramo kao $\hat{\theta}_{\text{MAP}} = \operatorname{argmax}_{\theta} p(\mathcal{D}|\theta)p(\theta)$. Pri odabiru apriorne distribucije $p(\theta)$, nastojimo da je to neka standardna teorijska distribucija i da je konjugatna distribucija za izglednost $\mathcal{L}(\theta|\mathcal{D})$. **Što to znači i zašto to želimo?**

- A** To znači da će umnožak izglednosti i apriorne distribucije dati distribuciju koja je iste vrste kao i apriorna distribucija, a ako je riječ o standardnoj teorijskoj distribuciji iz eksponencijalne familije, njezin **mod (maksimizator)** postoji u zatvorenoj formi, što nam omogućava da procjenitelj izračunamo analitički
- B** To znači da je apriorna distribucija ista vrsta distribucije kao i vjerojatnost podataka uz dane parametre, tj. izglednost parametara, pa će njihov umnožak biti distribucija koja je proporcionalna aposteriornoj distribuciji i čiji ćemo maksimum moći izračunati Bayesovim pravilom
- C** To znači da je apriorna distribucija upravlјana hiperparametrima kojima možemo ugoditi vjerojatnost parametara slučajne varijable koju procjenujemo, tj. parametri apriorne distribucije i parametri izglednosti su identični, što nam omogućava da te dvije distribucije pomnožimo i zatim nađemo maksimizator
- D** To znači da je aposteriorna distribucija parametara ista kao izglednost parametara, pa primjenom Bayesovog pravila možemo izračunati apriornu vjerojatnost parametara te, nakon zanemarivanja nazivnika koji je za fiksiran skup podataka konstantan, pronaći parametre koji maksimiziraju aposterionu vjerojatnost

$$p(\theta|D) \propto p(D|\theta)p(\theta)$$

↑ ↓ ↓
 apst. dist. izgledost apriorna dist.
 Konjugatne

5. (T) Kod MAP-procjenitelja, apriorna distribucija parametra $p(\theta)$ tipično se odabire tako da bude konjugatna za funkciju izglednosti $p(\mathcal{D}|\theta)$. Pretpostavimo da MAP-procjenitelj izračunavamo heurističkom metodom (npr., gradijentnim usponom). Što se događa ako za apriornu distribuciju parametra upotrijebimo distribuciju koja nije konjugatna funkciji izglednosti?

- \rightarrow ne koristimo kod MAP-a ??
- A Zajedničku distribuciju $p(\mathcal{D}, \theta)$ ne možemo izvesti u zatvorenoj formi, pa MAP nije definiran
 - B Aposteriornu distribuciju $p(\theta|\mathcal{D})$ ne možemo izvesti u zatvorenoj formi, ali MAP možemo izračunati heurističkom optimizacijom *
 - C Ako je apriorna distribucija $p(\theta)$ iz eksponencijalne familije, onda je apostериorna distribucija $p(\theta|\mathcal{D})$ u zatvorenoj formi i MAP je izračunljiv
 - * smatra biti konjugatna & garancija
 - D Neovisno o apriornoj distribuciji parametra $p(\theta)$, MAP je izračunljiv optimizacijom drugog reda (npr., Newtonovim postupkom) \rightarrow nema veze s pitagorinim
 \hookrightarrow ne možemo tvrditi da je neovisno

* samo čemo iskoristiti neki postupak kako bi smo dobili;

APPROKSIMIRATI aposteriornu dist.

6. (N) U beta-Bernoullijevom modelu, apriornu vjerojatnost parametra μ modeliramo beta-distribucijom čija je gustoća vjerojatnosti definirana kao:

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

Mod (maksimizator) te distribucije jest:

$$\mu^* = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Aposteriorna distribucija parametra definirana je kao:

$$p(\mu|\mathcal{D}, \alpha, \beta) = \mu^{m+\alpha-1} (1-\mu)^{N-m+\beta-1} \frac{1}{B(\alpha, \beta)p(\mathcal{D})}$$

Neka $\alpha = \beta = 2$. Računamo MAP-procjenu za parametar μ Bernouljeve varijable. To radimo na dva uzorka, $\mathcal{D}_1 = (N_1, m_1)$ i $\mathcal{D}_2 = (N_2, m_2)$, 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 $N_1 = 50$, Bernoullijeva varijabla realizirana je s vrijednošću $y = 1$ ukupno $m_1 = 42$ puta. U drugom uzorku, veličine $N_2 = 15$, Bernoullijeva varijabla realizirana je s vrijednošću $y = 1$ ukupno $m_2 = 3$ puta. Izračunajte MAP-procjene za parametar μ na temelju ova dva uzorka. **Koliko iznosi promjena u procjeni za μ između prve i druge procjene?**

- A -0.59 B +0.45 C -0.14 D -0.64

$$N_1 = 50$$

$$m_1 = 42$$

$$N_2 = 15$$

$$m_2 = 3$$

$$\mu_1 = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{43}{52}$$

$$\mu_2 = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{46}{67}$$

$$\alpha' = m_1 + \alpha = 44$$

$$\alpha'' = m_2 + \alpha' = 3 + 44 = 47$$

$$\beta' = N_1 - m_1 + \beta = 10$$

$$\beta'' = N_2 - m_2 + \beta' = 15 - 3 + 10 = 22$$

$$\Delta = \frac{46}{67} - \frac{43}{52} = -0,14$$

7. (P) Koristimo MAP-procjenitelj kako bismo procijenili parametre distribucije kategoričke (multinuličeve) 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 definišali Dirichletovu apriornu gustoću vjerojatnosti, $p(\mu_1, \mu_2, \mu_3 | \alpha_1, \alpha_2, \alpha_3)$, gdje je $(\alpha_1, \alpha_2, \alpha_3)$ vektor hiperparametara (parametri Dirichletove distribucije). Konkretno, te smo hiperparametre definišali kao $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 1)$. Međutim, skup podataka \mathcal{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) ?**

- A ~~$\mu_1 = 0$~~ , $\frac{1}{2} < \mu_2 < 1$, $0 < \mu_2 < \mu_3 < 1$
- B $0 < \mu_1 < \frac{1}{3}$, $\frac{1}{2} < \mu_2 < 1$, ~~$\mu_3 = 0$~~
- C $0 < \mu_1 < \mu_3 < 1$, $\frac{1}{3} < \mu_2 < \frac{2}{3}$
- D $0 < \mu_1 < \frac{1}{3}$, $\frac{1}{3} < \mu_2 < 1$, $0 < \mu_3 < \mu_2 < 1$

MAP za Dirichlet kat. model:

↗ broj realizacija za k-tu var.

$$\hat{\mu}_k = \frac{N_k + \alpha_k - 1}{\sum_k (N_k + \alpha_k) - K}$$

↗ broj dimenzija, ovdje 3

$$\hat{\mu}_1 = \frac{0 + 2 - 1}{N + 5 - 3} = \frac{1}{N + 2}$$

$$\hat{\mu}_2 = \frac{\frac{N}{2} + 2 - 1}{N + 5 - 3} = \frac{\frac{N}{2} + 1}{N + 2} = \frac{\frac{N+2}{2}}{N+2} = \frac{1}{2}$$

$$\hat{\mu}_3 = \frac{\frac{N}{2} + 1 - 1}{N + 5 - 3} = \frac{\frac{N}{2}}{N + 2} = \frac{N}{2N + 4}$$

N - barem 2 da bi zad. imao smisla

$$\frac{2a}{\hat{\mu}_1} = \frac{1}{4}$$

ne može

8. (P) Bacanje igraće kocke modeliramo kategoričkom varijablu \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 $\boldsymbol{\alpha} = (1, 1, 1, 1, 1, 1)$ → ne unosimo nikakvu dodatnu informaciju → jedn. vj. po cijeloj
- B $\boldsymbol{\alpha} = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ → gustoća se dist. u ekstremne vrijednosti
→ nemojmo fokus na neku konkretnu ishodu
- C $\boldsymbol{\alpha} = (2, 2, 2, 2, 2, 2)$ → pravednija kocke
- D $\boldsymbol{\alpha} = (1, 1, 1, 1, 3, 1)$
↳ izdvojimo češću 6, a ne 5

površini trokuta

DIRICHLET-KAP. MODEL

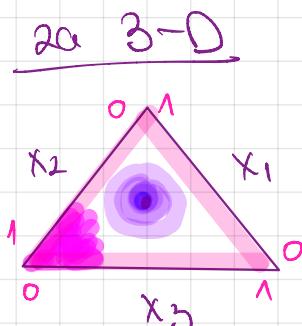
$$P(\boldsymbol{\mu} | \mathbf{z}) = P(\mu_1, \dots, \mu_k | z_1, \dots, z_k) = \frac{1}{B(\mathbf{z})} \prod_{i=1}^k \mu_i^{z_{ki}}$$

podređuje kaku je vj. dist.
ajdu VEKTOR \mathbf{z} odr. vj.

↳ normalizacijski faktor

param $\boldsymbol{\mu}$: INDIREKTNO

$\sum_{k=1}^K \mu_k = 1$: μ_k se nalazi u $(K-1)$ -dim.
standardnom simpleksu



0 → 1 vrijednost

• jedn. vj. dc

dc x_1, x_2 i x_3

primitivni vj. 0,5

• $x_2 = 3$

→ vjerovatnije vj. bliže jedinicama za x_2

Bayesov klasifikator

1. (T) Bayesov klasifikator definirali smo na sljedeći način:

$$h_j(\mathbf{x}; \boldsymbol{\theta}) = P(y = j | \mathbf{x}) = \frac{p(\mathbf{x}|y)P(y)}{\sum_{y'} p(\mathbf{x}|y')P(y')}$$

Neka je broj klasa veći od dva, $K > 2$, a značajke neka su realni brojevi, $\mathbf{x} \in \mathbb{R}^n$. **Koje teorijske distribucije ćemo koristiti za $P(y)$ i $P(\mathbf{x}|y)$?**

- A Kategoričku distribuciju za $P(y)$ i Gaussovnu distribuciju za $P(\mathbf{x}|y)$
- B Bernoullijevu distribuciju za $P(y)$ i Gaussovnu distribuciju za $P(\mathbf{x}|y)$
- C Kategoričku distribuciju za $P(y)$ i za $P(\mathbf{x}|y)$
- D Gaussovnu distribuciju za $P(y)$ i multinulijevu distribuciju za $P(\mathbf{x}|y)$

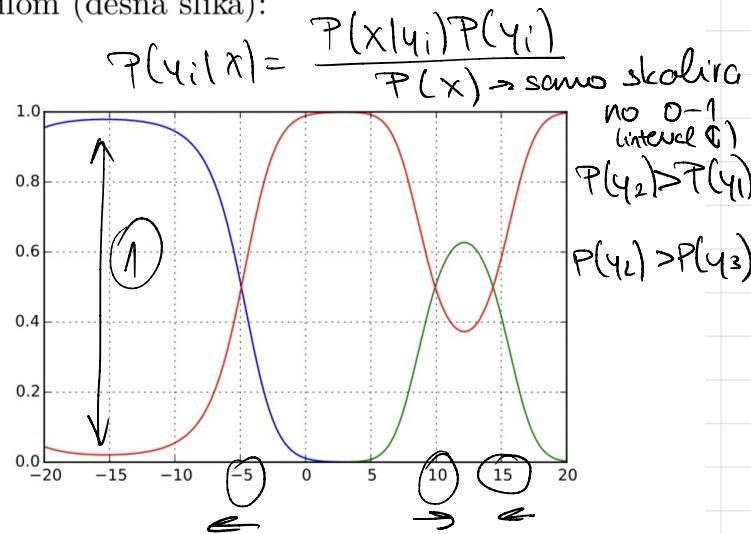
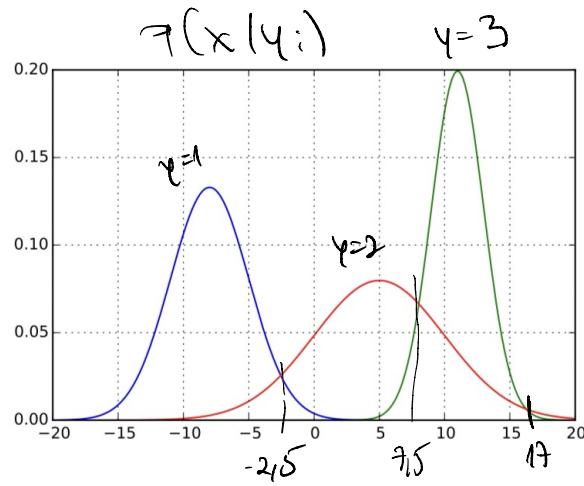
2. (T) Bayesov klasifikator definiran je kao

$$h(\mathbf{x}; \boldsymbol{\theta}) = \operatorname{argmax}_y p(\mathbf{x}|y)P(y)$$

Po čemu se vidi da je ovo generativan, a ne diskriminativan model?

- A Modelira vjerojatnost primjera i oznaka, budući da je, na temelju pravila umnoška, umnožak $p(\mathbf{x}|y)P(y)$ jednak zajedničkoj vjerojatnosti $p(\mathbf{x}, y)$
- B Zajedničku vjerojatnost primjera i oznaka, $p(\mathbf{x}|y)P(y)$, faktorizira u dva faktora te zanemaruje nazivnik $p(\mathbf{x})$, koji je ionako konstantan za svaku klasu y
- C Parametre distribucija $p(\mathbf{x}|y)$ i $P(y)$, a time indirektno i parametre aposteriorne distribucije $P(y|\mathbf{x})$, računa MAP-procjeniteljem, čime sprječava prenaučenost
- D Primjer \mathbf{x} klasificira prema MAP-hipotezi, dakle u klasu koja maksimizira aposteriornu vjerojatnost oznake, $p(y|\mathbf{x})$, koja je proporcionalna zajedničkoj vjerojatnosti primjera i oznaka, $p(\mathbf{x}, y)$

3. (P) Koristimo Gaussov Bayesov klasifikator kako bismo riješili trokласни klasifikacijski problem. Procijenjene gustoće vjerojatnosti za izglednosti klase su $p(x|y=1) = \mathcal{N}(-8, 3)$, $p(x|y=2) = \mathcal{N}(5, 5)$ i $p(x|y=3) = \mathcal{N}(11, 2)$. Na slikama ispod prikazane su izglednosti klase (lijeva slika) i aposteriorne vjerojatnosti dobivene Bayesovim pravilom (desna slika):



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

- A $P(y=1) = 0.1, P(y=2) = 0.7, P(y=3) = 0.2$
- B $P(y=1) = P(y=2) = P(y=3) = \frac{1}{3}$ \rightarrow uniformne (jeck. skolice) za svu x
- C $P(y=1) = P(y=2) = 0.4, P(y=3) = 0.2$
- D ~~$P(y=1) = P(y=2) = 0.1, P(y=3) = 0.8$~~ \times

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

- neki prep. može smaziti sl. modela (kapacitet reprezentacije), ali povećati br. param.

\mathcal{H}_1 : misl. kov. \rightarrow dijagonala matrica, ne modeliramo zavisnost između različitih značajki u modelu
 različ. var. između klase \rightarrow ne koristimo dijag. unutar \rightarrow modeliramo se svaku znač. \rightarrow NEMATOS DODTRUPNU

DIJAGONALNA NEĐ. (različ. el. ne dijag.)

$$\boxed{2Kn} = 2 \cdot 5 \cdot 10 = 100 \text{ param.}$$

↓
st. vr.

\mathcal{H}_2 : dijag.

jeck. unutar : DODTRUPNA (1 sigma unutar svake klase)

$$\boxed{K + Kn} = 60$$

\hookrightarrow std dev

\mathcal{H}_3 : postoji korelacija \rightarrow puna matrica

ne različ. sc \rightarrow DIJELJENA

$$\frac{n(n+1)}{2} + Kn = 65$$

5. (N) Na skupu označenih primjera u ulaznom prostoru dimenzije $n = 3$ treniramo Gaussov Bayesov klasifikator za klasifikaciju primjera u $K = 2$ klase, uz pretpostavku dijeljene kovarijacijske matrice. Model je definiran kao

$$h_j(\mathbf{x}) = \ln p(\mathbf{x}, y)$$

Prisjetimo se da je izglednost klase s oznakom $y = j$ kod Gaussovog Bayesovog klasifikatora definirana multivarijantnom Gaussovom gustoćom vjerojatnosti:

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

gdje je Σ_j matrica kovarijacije za klasu j . Treniranjem modela dobili smo sljedeće procjene za parametre:

$$\hat{\mu}_1 = 0.2$$

$$\hat{\boldsymbol{\mu}}_1 = (1, 0, -2)$$

$$\hat{\Sigma}_1 = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{pmatrix}$$

$$\hat{\mu}_2 = 0.8$$

$$\hat{\boldsymbol{\mu}}_2 = (2, -1, 5)$$

$$\hat{\Sigma}_2 = \begin{pmatrix} 6.25 & -0.5 & -1 \\ -0.5 & 1.25 & -0.75 \\ -1 & -0.75 & 3.5 \end{pmatrix}$$

Iz ovoga smo zatim procijenili dijeljenu kovarijacijsku matricu $\hat{\Sigma}$ definiranu kao težinski prosjek kovarijacijskih matrica $\hat{\Sigma}_j$, $j = 1, 2$. Zanima nas klasifikacija modela za primjer $\mathbf{x} = (0, 0, 0)$. **Koliko iznosi predikcija modela za klasu $y = 1$ za taj primjer, $h_1(\mathbf{x})$?**

- A -6.885 B +0.002 C -4.819 D -6.429

$$h_1(\mathbf{x}) = \ln p(\mathbf{x}|y=1) + \ln(p(y=1))$$

$$\hat{\Sigma} = \hat{\mu}_1 \cdot \hat{\Sigma}_1 + \hat{\mu}_2 \cdot \hat{\Sigma}_2$$

$$= 0.12 \begin{bmatrix} 5 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{bmatrix} + 0.8 \begin{bmatrix} 6.25 & -0.5 & -1 \\ -0.5 & 1.25 & -0.75 \\ -1 & -0.75 & 3.5 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$$p(\mathbf{x}|y=1) = \ln \frac{1}{(2\pi)^{3/2} \sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right\}$$

$$= \ln \left(\frac{1}{\sqrt{6.24}} \right) - \frac{1}{2} \cdot \frac{7}{6} = -5,276$$

$$p(y=1) = \ln(0.12) = -1.61$$

$$h_1(\mathbf{x}) = -5,276 - 1,61 = -6,885$$

$$[-1 \ 0 \ 2] \begin{bmatrix} 1/6 & & \\ & 1/2 & \\ & & 1/4 \end{bmatrix} = [-1/6 \ 0 \ 1/2] [-1 \ 0 \ 2]$$

$$= \frac{1}{6} + 1 = \frac{7}{6}$$

M

Bayesov klasifikator II

2. [Svrha: Isprobati na konkretnom primjeru izračun parametara naivnog Bayesovog klasifikatora.]
 Naivan Bayesov model želimo upotrijebiti za binarnu klasifikaciju "Skupo ljetovanje na Jadranu".
 Skup primjera za učenje je sljedeći:

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) Izračunajte MLE procjene svih parametara modela te klasificirajte primjere (Istra, ne, kamp, bus) i (Dalmacija, da, hotel, bus).
 (b) Izračunajte Laplaceove (zaglađene) procjene za sve parametre modela te klasificirajte nanovo iste primjere.

a) ① apriorne vj.

MLE

$$P(y=\text{da}) = \frac{4}{7}$$

$$P(y=\text{ne}) = 1 - P(y=\text{da}) = \frac{3}{7}$$

$m = \text{Istra}$

$$0 = \text{NE} \rightarrow P(o=\text{ne} | y=\text{da}) = \frac{1}{4}$$

$$P(o=\text{ne} | y=\text{ne}) = \frac{3}{3} = 1$$

$$s = \text{kamp} \rightarrow P(s=\text{kamp} | y=\text{da}) = \emptyset$$

$$P(s=\text{kamp} | y=\text{ne}) = \frac{2}{3}$$

$$p = \text{bus} \rightarrow P(p=\text{bus} | y=\text{da}) = \emptyset$$

$$P(p=\text{bus} | y=\text{ne}) = \frac{2}{4}$$

$$P(y_j | x) = \prod_k P(x_k | y_j) P(y_j) = \frac{2}{4} \cdot \frac{1}{4} \cdot \emptyset \cdot \emptyset = \frac{4}{7} = \emptyset$$

$y_i = \text{da}$

② Zaglađene vj.

$$P(m=\text{Istra} | y=\text{da}) = \frac{2}{4}$$

$$\begin{aligned} * \frac{LP}{\sum} \\ = \frac{2+1}{4+3} \end{aligned}$$

$$P(m=\text{Kvarner} | y=\text{da}) = \frac{\emptyset}{4} = 0$$

$$\begin{aligned} P(m=\text{Dalmacija} | y=\text{da}) &= 1 - \frac{2}{4} - \emptyset \\ &= \frac{2}{4} \end{aligned}$$

$$P(m=\text{Istra} | y=\text{ne}) = \frac{2}{3}$$

$$P(m=\text{Kvarner} | y=\text{ne}) = \frac{1}{3}$$

$$P(m=\text{Dalmacija} | y=\text{ne}) = 1 - \frac{2}{3} = \frac{1}{3}$$

b) Laplace : dodajemo $\Rightarrow 1$ primjer svakoj klasi
 +1 u brojniku

+uk. br. anzapački u nazivniku

1. (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?**

- A 200 B 5049 C 5150 D 10200

- $n = 100$
- binarna klas.: $K = 2$

LOG. REG.

$$\text{za drug. } 1 - \hat{y}_j \text{. prve} : \underline{\underline{n+1}} =$$

GBC

$$\text{s djeljenom kov. matricom} : \frac{n}{2}(n+1) + nK + K-1 = 5050$$

2. (N) Treniramo naivan Bayesov model za binarnu klasifikaciju "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 $P(y=1|\mathbf{x})$?

- A 0.1747 B 0.0032 C 0.6856 D 0.3144

① apriorne y_j

$$P(y=\text{da}) = \frac{3}{7}$$

$$P(y=\text{ne}) = \frac{4}{7}$$

$$y=\text{ne}$$

$$\frac{2+1}{4+3} = \frac{3}{7}$$

② izgledaju y_j (izglednost)

$$P(m=\text{Istra} | y=\text{da}) = \frac{0+1}{3+3} = \frac{1}{6}$$

$$\frac{3}{4+2} = \frac{3}{6} = \frac{1}{2}$$

$$P(o=\text{ne} | y=\text{da}) = \frac{1+1}{3+2} = \frac{2}{5}$$

$$\frac{3}{4+3} = \frac{3}{7}$$

$$P(s=\text{kamp} | y=\text{da}) = \frac{1+1}{3+3} = \frac{2}{6}$$

$$\frac{2}{4+3} = \frac{2}{7}$$

$$P(p=\text{bus} | y=\text{da}) = \frac{1+1}{3+3} = \frac{2}{6}$$

$$P(y=1 | \mathbf{x}) = \frac{P(x | y=1) P(y=1)}{\sum_j P(x | y=j) P(y=j)} = \frac{\frac{1}{6} \cdot \frac{2}{5} \cdot \frac{2}{6} \cdot \frac{2}{7} \cdot \frac{3}{7}}{\frac{1}{6} + \underbrace{\frac{9}{343} \cdot \frac{4}{7}}_{\frac{36}{2401}}} = \frac{\frac{1}{315}}{\frac{1}{315} + \frac{36}{2401}} = \frac{1}{315} + \frac{36}{2401} = 0,1747$$

3. (P) Naivan Bayesov klasifikator prepostavlja 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 ($y = 1$) ili ne ($y = 0$). Model koristi binarne značajke koje indiciraju pojavljivanje određene riječi u novinskom članku. Na primjer, izglednost $P(\text{stožer}|y = 1)$ 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?

- A $P(\text{stožer}|y = 1) = P(\text{stožer}|\text{pandemija}, y = 1)$ → nema dodatne info
 - B $P(\text{general}|y = 0) = P(\text{general}|\text{stožer}, y = 0)$ → ako nije očl. o koroni, može se pojaviti "koronavirus"
 - C $P(\text{koronavirus}|y = 0) = P(\text{koronavirus}|\text{general}, y = 0)$ → nema "general" neće svakako pojaviti
 - D $P(\text{pandemija}|y = 1) = P(\text{stožer}|y = 1)$
- moglo bi biti
dosta desto sljepo

$y=0 \nsubseteq$ tene koje možemo dobiti
generali kao stanovi stožer → vjerno
→ veća šansa da će i general
biti unutrašnji
(← manji broj člancova)

4. (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: x_1 – politička orijentacija (kategorička značajka s tri vrijednosti), x_2, x_3 – dob kandidata i politički staž (dvije numeričke značajke), x_4 – populist (binarna značajka) i x_5 – 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:

\mathcal{H}_0 : Bayesov klasifikator bez ikakvih pretpostavki o uvjetnoj nezavisnosti

\mathcal{H}_1 : Polunaivan Bayesov klasifikator

\mathcal{H}_2 : Naivan Bayesov klasifikator

Polunaivan model \mathcal{H}_1 isti je kao i naivan model \mathcal{H}_2 , s tom razlikom da smo u jedan faktor združili značajke x_1 i x_4 , 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 x_2 i x_3 koristimo dijeljenu kovarijacijsku matricu. Izračunajte broj parametara za svaki od ova tri modela. **Koliko parametara sveukupno imaju ova tri modela?**

- A 52 B 61 C 62 D 64

$$\mathcal{H}_1 : x_1 \text{ i } x_4$$

$$\text{dijeljenje: } x_2, x_3$$

$$x_1 : K = 3$$

$$x_2, x_3 \Rightarrow \text{num.}$$

$$x_4 : K = 2$$

$$x_5 : K = 2$$

$$\underline{\mathcal{H}_2} \quad x_2, x_3 : 2 + 2^*2 = 6$$

$$x_1, x_4, x_5 :$$

$$K=2 : 2^*(2+1+1) = 8$$

$$P(y=0) \oplus 1$$

$$\approx 15$$

$$\text{UK: } 30 + 19 + 15$$

$$= 64$$

\mathcal{H}_0

$$- x_2, x_3 : \text{kov. matrica } 2 \times 2 : \frac{2}{2}(2+1) = 3$$

$\xrightarrow{\text{dijeljenje: ne treba množiti sbroj klase}}$
(ista za $y=0$ i $y=1$)

$$\mu : 2 \times 2 = 4 \quad (2D ; 2 \text{ klase})$$

- diskretne var. : sve zavisne

$$\Rightarrow \text{sve komb.} - 1 \quad (1 - \Sigma)$$

$$y=0 : \underbrace{3 \times 2 \times 2 - 1}_{\text{ }} = 11$$

$$y=1 : \quad x_2 \quad \Rightarrow 22$$

$$- \text{ dodatni param. za } P(y=\emptyset) \quad \frac{1}{30}$$

\mathcal{H}_1

x_1 i x_4 zavisne (zavisne)

$$x_5 : 2$$

$$K=2 \quad \begin{cases} y=0 : \underbrace{3 \times 2 - 1}_{\text{ }} = 5 \\ y=1 : \quad x_2 \end{cases} = 10$$

$$\text{per } \frac{K=2}{\Sigma=1} \quad \frac{1}{1 \times 2}$$

$$y : 1$$

UK:

19

x_2, x_3 : dijeljena i dijag. kov. m. : 2 dij. p.

$$\mu : 2 \times 2$$

$$- 6$$

5. (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:

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

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?**

- A $P(y)P(x_1, x_2|y)P(x_3|y)$
- B $P(y)P(x_1, x_2, x_3|y)$
- C $P(y)P(x_1, x_3|y)P(x_2|y)$
- D $P(y)P(x_1|y)P(x_2|y)P(x_3|y)$

$$I(X_1, Y) > 0,01 \text{ združimo}$$

$$X = \{X_1, X_2, X_3\}$$

$$I(X_1, Y) = \sum_{X_1, Y} P(X_1, Y) \cdot \ln \frac{P(X_1, Y)}{P(X)P(Y)}$$

$$P(X_1 = \emptyset) = 0,2 + 0,1 + 0,1 = 0,4$$

$$P(X_1 = 1) = 1 - 0,4 = 0,6$$

$$P(X_3 = 0) = 0,2 + 0,1 + 0,3 = 0,6$$

$$P(X_3 = 1) = 0,4$$

$$P(X_2 = \emptyset) = P(X_2 = 0 | X_3 = 0) \cdot P(X_3 = 0)$$

$$+ P(X_2 = \emptyset | X_3 = 1) \cdot P(X_3 = 1) \\ = 0,5 \cdot 0,6 + 0,3 \cdot 0,4 \\ = 0,42$$

$$P(X_2 = 1) = 0,58$$

$$P(X_2 = 0) = \sum_{X_1, X_3} P(X_1, X_2 = 0, X_3) = 0,8$$

$$P(X_2 = 1) = 0,2$$

$$\begin{aligned} P(X_1 = 1, X_2 = 1) &= \sum_{X_3} P(X_1 = 1, X_2 = 1, X_3) \\ &= P(X_1 = 1, X_2 = 1, X_3 = 1) + \\ &\quad P(X_1 = 1, X_2 = 1, X_3 = \emptyset) \\ &= 0,1 \cdot 0 = 0,1 \end{aligned}$$

$P(X_1 = 0, X_2 = 0) = 0,3 \quad P(X_1 = 0, X_2 = 1) = 0,1 \quad P(X_1 = 1, X_2 = 0) = 0,5$

$$\begin{aligned} I(X_1, X_2) &= 0,1 \cdot h \frac{0,1}{0,6 \cdot 0,2} + 0,3 \cdot h \frac{0,3}{0,4 \cdot 0,8} \\ &\quad + 0,1 \cdot h \frac{0,1}{0,4 \cdot 0,2} + 0,5 \cdot h \frac{0,5}{0,6 \cdot 0,8} = 5,13 \cdot 10^{-3} \end{aligned}$$

X_2, X_3

$$\begin{array}{ll} P(X_2=0) = 0,8 & P(X_3=0) = 0,6 \\ P(X_2=1) = 0,2 & P(X_3=1) = 0,4 \end{array}$$

$$P(X_2=1, X_3=1) = 0,1$$

$$P(X_2=1, X_3=0) = 0,1$$

$$P(X_2=0, X_3=1) = 0,3$$

$$P(X_2=0, X_3=0) = 0,5 \quad \Sigma = 1 \quad \checkmark$$

$$I(X_2, X_3) = \sum_{X_2, X_3} P(X_2, X_3) \ln \frac{P(X_2, X_3)}{P(X_2)P(X_3)}$$

$$= 0,1 \ln \frac{0,1}{0,2 \cdot 0,6} + 0,1 \ln \frac{0,1}{0,2 \cdot 0,6} + 0,3 \ln \frac{0,3}{0,8 \cdot 0,4} + 0,5 \ln \frac{0,5}{0,8 \cdot 0,6}$$

$$= 5,13 \cdot 10^{-3}$$

X_1, X_3

$$\begin{array}{ll} P(X_1=0) = 0,4 & P(X_3=0) = 0,6 \\ P(X_1=1) = 0,6 & P(X_3=1) = 0,4 \end{array}$$

$$P(X_1=1, X_3=1) = 0,3$$

$$P(X_1=1, X_3=0) = 0,3$$

$$P(X_1=0, X_3=1) = 0,1$$

$$P(X_1=1, X_3=1) = 0,3 \quad \# \Sigma = 1$$

$$I(X_1, X_3) = \sum_{X_1, X_3} P(X_1, X_3) \ln \frac{P(X_1, X_3)}{P(X_1)P(X_3)}$$

$$= 0,032 \rightarrow \text{zdroženo}$$

Probabilistički grafički modeli

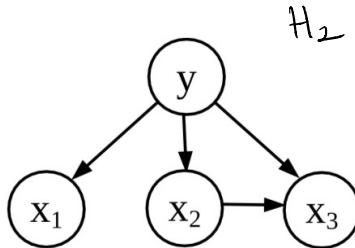
1. (T) Za Bayesovu mrežu kažemo da je generativni i parametarski model. **Zašto?**

- A 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
- B 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
- C 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
- D Generativni jer opisuje postupak kojim se mogu generirati podaci 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

2. (T) Bayesove mreže na sažet način prikazuju zajedničku distribuciju te kodiraju uvjetne stohastičke nezavisnosti između varijabli. No, kao i svaki model strojnog učenja, tako se i Bayesove mreže mogu prenaučiti. **Koja je veza između uvjetnih nezavisnosti varijabli u Bayesovoj mreži i opasnosti od prenaučenosti?**

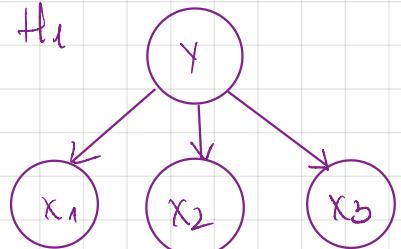
- A Uvođenje pretpostavki o uvjetnoj nezavisnosti pojednostavljuje strukturu Bayesove mreže i smanjuje broj parametara, čime se smanjuje i mogućnost prenaučenosti
- B Uvođenjem pretpostavki o uvjetnoj nezavisnosti povećava se broj čvorova mreže, a time i broj parametara, što model čini složenijim i time sklonijim prenaučenosti
- C Uvjetne nezavisnosti određuju strukturu mreže na način da definiraju koji su čvorovi mreže međusobno povezani, međutim to nema utjecaja na složenost modela niti na sklonost prenaučenosti
- D Pretpostavke o uvjetnoj nezavisnosti čine induktivnu pristranost modela, pa što je više uvjetnih nezavisnosti, to je veća pristranost i model je lako prenaučiti

3. (P) Na slici ispod prikazana je Bayesova mreža koja odgovara polunaivnom Bayesovom klasifikatoru. Pretpostavite da su značajke x_1 , x_2 i x_3 binarne varijable te da je oznaka klase y također binarna varijabla. Označimo ovaj model sa \mathcal{H}_2 . Model \mathcal{H}_2 može se pojednostaviti ako se ukloni brid između varijabli x_2 i x_3 . Označimo takav model sa \mathcal{H}_1 . S druge strane, od modela \mathcal{H}_2 može se napraviti još složeniji model koji odgovara potpuno povezanom acikličkom grafu. Označimo takav model sa \mathcal{H}_3 .



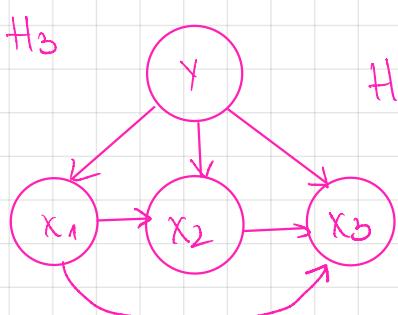
Razmotrite koliko parametara imaju modeli \mathcal{H}_1 , \mathcal{H}_2 i \mathcal{H}_3 . **Koliko model \mathcal{H}_2 ima više parametara od modela \mathcal{H}_1 , a koliko manje parametara od modela \mathcal{H}_3 ?**

- A 2 više, 3 manje B 2 više, 6 manje C 4 više, 4 manje D 4 više, 8 manje



$$\mathcal{H}_1: P(\dots) = P(y) \cdot P(x_1|y) \cdot P(x_2|y) \cdot P(x_3|y)$$

1 2 2
 (Bernoulli) $2 * (2-1) = 2$
 ↓ ↓ ↓
 broj vrijednosti sumu mora
 od y biti 1
 => 7



$$\mathcal{H}_3: P(\dots) = P(y) \cdot P(x_1|y) \cdot P(x_2|y, x_1) \cdot P(x_3|y, x_1, x_2)$$

1 2 1 * (2 * 2) 1 * (2 * 2 * 2)
 ↓ ↓ = 4 = 8
 => 15

$$\mathcal{H}_2: P(\dots) = P(y) \cdot P(x_1|y) \cdot P(x_2|y) \cdot P(x_3|y, x_2)$$

↓ ↓ ↓ ↓
 1 2 2 2
 2 * (2-1) = 4
 => 9

var. upite: 2 varj: $2-1=1$
 ↑ 2 moguće varj
 q ↓ 2 moguće varj. } $2 \times 2 = 4$

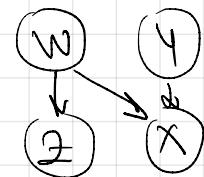
4. (P) Razmotrite Bayesovu mrežu koja zajedničku vjerojatnost faktorizira na sljedeći način:

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

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

- A $w \perp y, z \perp \{x, y\} | w$ B $x \perp y | z, z \perp w | y$ C $w \perp y, z \perp x, x \perp w | \{z, y\}$ D $y \perp w, y \perp x | \{w, z\}$

$$\begin{aligned} P(w, y, x, z) &= P(w)P(y)P(x|w, y)P(z|w) \\ w \perp y &= P(w, y)P(x|w, y)P(z|w) \\ &= P(w, y, x)P(z|w) \\ &= P(w, y, x)P(z|x, y, w) \\ z \perp \{x, y\} | w &= P(w, y, x, z) \end{aligned}$$



5. (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:

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

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

- A 10 B 22 C 25 D 27

- v, w, z : binarne

sveka Bayesova mreža kodira v, w, y .

- x, y : ternarne

v, w, x, y, z

potpuno povezana mreža

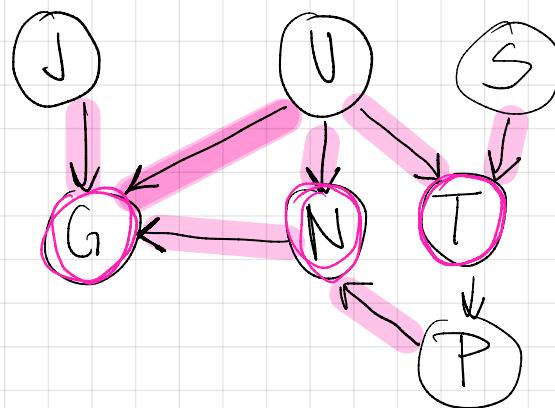
$$\begin{aligned} P(v, w, x, y, z) &= P(v)P(w|v)P(x|v, w)P(y|v, w, x)P(z|v, w, x, y) \\ &= P(v)P(w)P(x|v)P(y|w, x)P(z|x, y) \\ &\quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\ &\quad 1 \quad 1 \quad \underbrace{(3-1)^* 2}_{4} \quad \underbrace{(3-1)^*(2^* 3)}_{12} \quad \underbrace{(2-1)^*(3^* 3)}_9 \end{aligned}$$

$\Rightarrow 27$ param.

6. (P) Bayesovom mrežom s pet binarnih varijabli modeliramo prometne prilike u gradu Zagrebu. U našoj mreži, jutarnje doba dana (J) i loše vrijeme (V) utječe 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 uzrokuje 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?**

sraz

- A $P(V = 1 | P = 1, T = 1) < P(V = 1 | T = 1)$
- B $P(G = 1 | J = 1, V = 1) > P(G = 1 | J = 1)$
- C $P(V = 1 | P = 1, N = 1) < P(V = 1 | N = 1)$ *uvj. mora biti s obje strane*
- D $P(T = 1 | V = 1, P = 1) > P(T = 1 | V = 1)$



J ↗ V | G

$$P(J | V, G) \neq P(J | G)$$

$$P(V = 1 | P = 1, T = 1) \neq P(V = 1 | T = 1)$$

V i P nisu u srazu \Rightarrow
takođe odbaceno

$$P(V = 1 | P = 1, N = 1) \neq P(V = 1 | N = 1) \rightarrow V \not\perp P | N \quad (\checkmark)$$

smjer u vj.

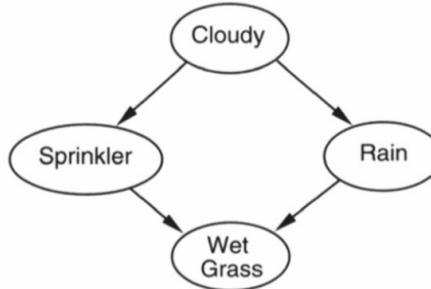
za losije vrijene \Rightarrow

→ dovoljan uzrok;

određena vj. od V

Probabilistički grafički modeli II.

1. (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.



			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 0.491 C 0.709 D 0.825

$$P(S, R, W, C) = P(C) P(S|C) P(R|C) P(W|S, R)$$

↗ var.
↓ smetnje

$$P(R=1 | W=1, C=\emptyset) = \frac{\sum_S P(S, R=1, W=1, C=\emptyset)}{\sum_{S, R} P(S, R, W=1, C=\emptyset)} = 0,825$$

var. opte ↪

BR. : $P(C=\emptyset) P(S=0 | C=\emptyset) P(R=1 | C=\emptyset) P(W=1 | S=0, R=1)$

$$+ P(C=\emptyset) P(S=1 | C=\emptyset) P(R=1 | C=\emptyset) P(W=1 | S=1, R=1)$$

$$= 0,5 \cdot 0,5 \cdot 0,2 \cdot 0,9 + 0,5 \cdot 0,5 \cdot 0,2 \cdot 0,99$$

$$= 0,0945$$

NAPOMENI: $P(C=\emptyset) P(S=\emptyset | C=\emptyset) P(R=\emptyset | C=\emptyset) P(W=1 | S=\emptyset, R=\emptyset) = 0,5 \cdot 0,5 \cdot 0,8 \cdot 0 = 0$

$$+ 0,1 P(C=\emptyset) P(S=\emptyset | C=\emptyset) P(R=1 | C=\emptyset) P(W=1 | S=\emptyset, R=1) = 0,5 \cdot 0,5 \cdot 0,2 \cdot 0,9 = \frac{9}{200}$$

$$+ 0,1 P(C=\emptyset) P(S=1 | C=\emptyset) P(R=\emptyset | C=\emptyset) P(W=1 | S=1, R=\emptyset) = 0,5 \cdot 0,5 \cdot 0,8 \cdot 0,1 = \frac{1}{50}$$

$$+ 0,1 P(C=\emptyset) P(S=1 | C=\emptyset) P(R=1 | C=\emptyset) P(W=1 | S=1, R=1) = 0,5 \cdot 0,5 \cdot 0,2 \cdot 0,99 = \frac{99}{2000}$$

$$\leq 0,1145$$

2. (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 modelira 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 $P(L = 1, S = 1, N = 1)$?**

- [A] 0.023 [B] 0.074 [C] 0.143 [D] 0.833

$$P(L, N, S, T) = P(L)P(N)P(S|L, N)P(T|N)$$

$$P(L=1, S=1, N=1, T) = P(L=1)P(N=1)P(S=1|L=1, N=1)P(T|N=1)$$

MAP

$$P(L=1) = \frac{k+1}{7+2} = \frac{5}{9}$$

$$P(N=1) = \frac{1+1}{7+3} = \frac{2}{10} = \frac{1}{5}$$

$$P(S=1|L=1, N=1) = \frac{1+1}{1+2} = \frac{2}{3}$$

Laplace

→ dodjeljeno 2 (S je bin.)

$$\underbrace{P(T|N=1)}$$

$$P(T=0|N=1) + P(T=1|N=1)$$

$$= 1 \quad (T \text{ je bin. var.})$$

$$= \frac{5}{9} \cdot \frac{1}{8} \cdot \frac{2}{3} \cdot 1 = \frac{2}{27} = 0.074$$

3. (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?**

- A 54 B 124 C 200 D 739

$$\begin{aligned}
 P(y|x=1, z=0) &= P(x=1)P(y=0)P(z=0|x=1, y=0) \\
 &\quad + P(x=1)P(y=1)P(z=0|x=1, y=1) \\
 &= 0.2 \cdot 0.7 \cdot 0.5 + 0.2 \cdot 0.3 \cdot 0.9 \\
 &= 0.124
 \end{aligned}$$

↙
marginalizirano
po y

ostaje nam $1000 \cdot 0.124 = 124$ vektora

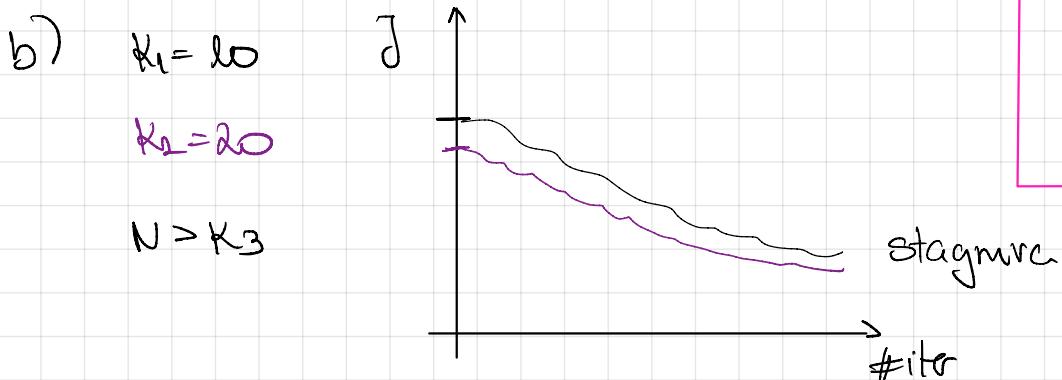
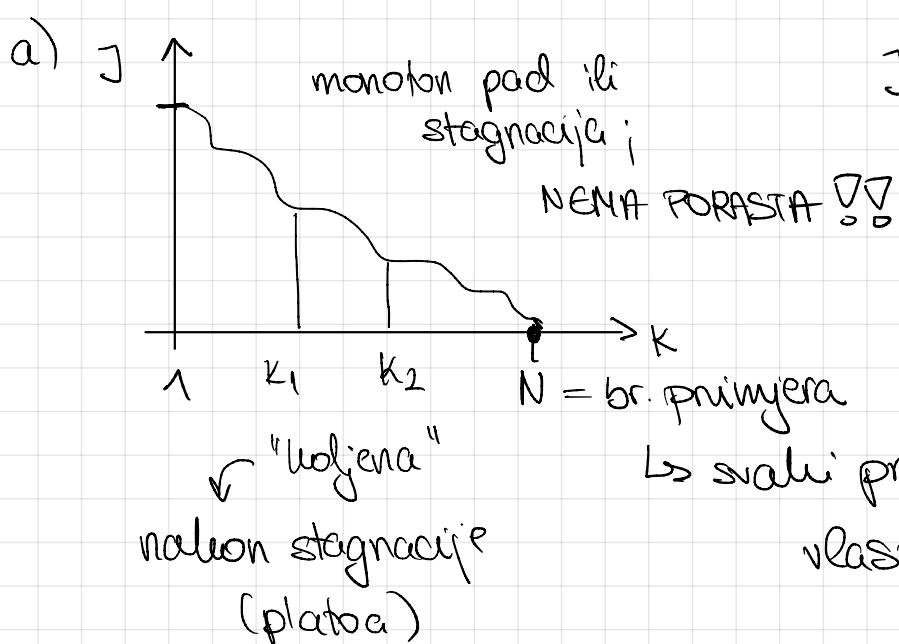
4. (T) Procjena parametara Bayesove mreže temelji se na maksimizaciji log-izglednosti parametara pod modelom. Procjena parametara može biti bitno drugačija za slučaj potpunih podataka, gdje su sve varijable opažene, u odnosu na slučaj nepotpunih podataka, gdje u model trebamo uključiti skrivene ili latentne varijable. **Što je prednost procjene parametara kod potpunih podataka (modela bez skrivenih varijabli) u odnosu na nepotpune podatke (modela sa skrivenim varijablama)?**

- A 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
- B 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
- C 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
- D 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

Grupiranje

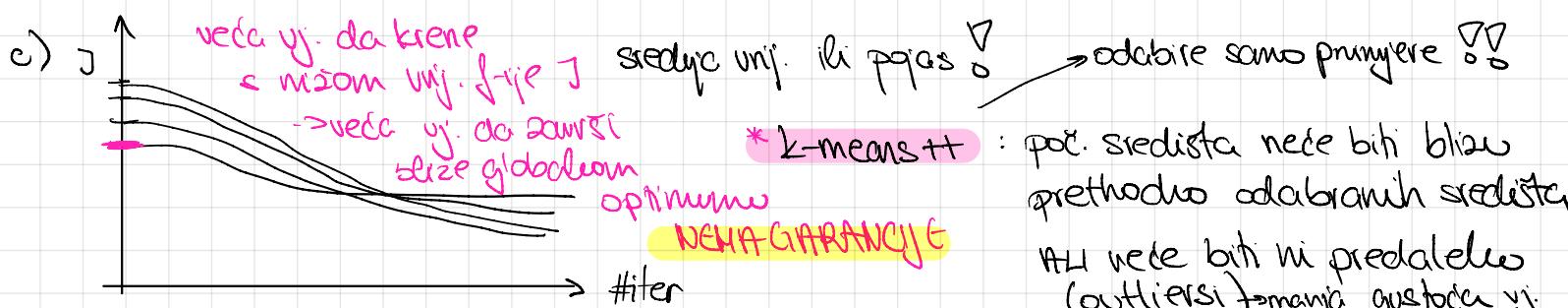
1. [Svrha: Razumjeti rad algoritma k-sredina u smislu minimizacije kriterija pogreške. Razumjeti kako rad algoritma ovisi o broju grupa K i odabiru početnih središta.] \rightarrow (o središnje (centroide))
 Algoritam k-sredina minimizira kriterij pogreške $J(\mu_1, \dots, \mu_K | \mathcal{D})$. Vrijednost tog kriterija ovisi o broju grupa K , koji je unaprijed postavljen, te o položajima središta, koja se mijenjaju kroz iteracije.

- (a) Nacrtajte skicu vrijednosti kriterija pogreške J kao funkcije broja grupa K . Koja je minimalna vrijednost funkcije J i zašto?
- (b) Izaberite na skici iz zadatka (a) tri vrijednosti za K i skicirajte na jednom grafikonu vrijednost kriterija pogreške J kao funkcije broja iteracija (tri krivulje).
- (c) Izaberite na skici iz zadatka (a) jednu vrijednost za K . Skicirajte na jednom grafikonu vrijednosti kriterija pogreške J kao funkcije broja iteracija, ali ovaj put uvezvi u obzir stohastičnost uslijed slučajnog odabira početnih središta (nacrtajte nekoliko mogućih krivulja na istom grafikonu). Koje od tih krivulja su izglednije za algoritam k-means++?



k-središnja: $O(T n N K)$

k-medoida: $O(T K(N+K)^2)$
svaki sa svakim



3. [Svrha: Isprobati izračun Randovog indeksa na konkretnom primjeru. Razumjeti primjenjivost Randovog indeksa.] Nedostatak svih algoritama grupiranja koje smo razmotrili jest što se broj grupa K mora zadati unaprijed. Osim u rijetkim slučajevima kada nam je taj broj unaprijed poznat, to predstavlja problem.

- (a) Kada su primjeri ili podskup primjera označeni, kvaliteta grupiranja (uključivo i broj grupa K) može se procijeniti Randovim indeksom. Randov indeks zapravo izračunava točnost s kojom ćemo par jednako označenih primjera smjestiti u istu grupu, odnosno par različito označenih primjera u različitu grupu. Izračunajte Randov indeks za sljedeću particiju označenih primjera (podskupovi su grupe dobivene grupiranje, a brojke su oznake klase primjera):

$$\{\{0, 0, 1, 2\}, \{1, 1\}, \{2, 2, 2, 1, 0\}\}.$$

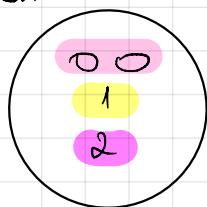
- (b) Skicirajte vrijednost Randovog indeksa kao funkcije broja grupa K .
(c) Randov indeks možemo koristiti samo ako su podaci označeni ili je podskup podataka označen. Međutim, čini se da to onda ujedno podrazumijeva da je unaprijed poznat broj grupa K . Imamo li koristi od Randovog indeksa čak i onda kada unaprijed znamo broj grupa? Možemo li ikako upotrijebiti Randov indeks, a da nam broj grupa nije unaprijed poznat?

Randov indeks: mjeri točnosti

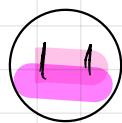
$\text{Re } [0, 1]$

$$N=11$$

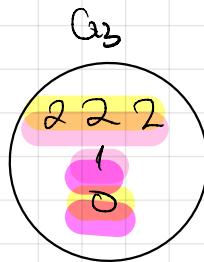
$$G_1$$



$$G_2$$



$$G_3$$



$$R = \frac{TP + TN}{\binom{N}{2}}$$

a: TP: broj ispravno zadruženih parova

$$G_1 : \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)$$

$\rightarrow \emptyset$ parova

$$G_2 : \left(\frac{1}{2}\right) + \left(\frac{2}{2}\right) + \left(\frac{1}{2}\right)$$

$$G_3 : \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) + \left(\frac{3}{2}\right)$$

b: TN: ispravno razdruženi parovi

$$\begin{aligned} \emptyset : & 2 \cdot 2 + 2 \cdot 3 + 2 \cdot 1 + \\ 1 : & 1 \cdot 3 + 1 \cdot 1 + \\ 2 : & 1 \cdot 2 + 1 \cdot 1 + 1 \cdot 1 + \end{aligned} \quad \left. \begin{array}{l} G_1 - G_2 \\ G_1 - G_3 \end{array} \right\} = 1 + 1 + 3 = 5$$

$$\begin{aligned} 1 : & 2 \cdot 3 + 2 \cdot 1 \\ & = 28 \end{aligned} \quad \left. \begin{array}{l} G_2 - G_3 \end{array} \right\}$$

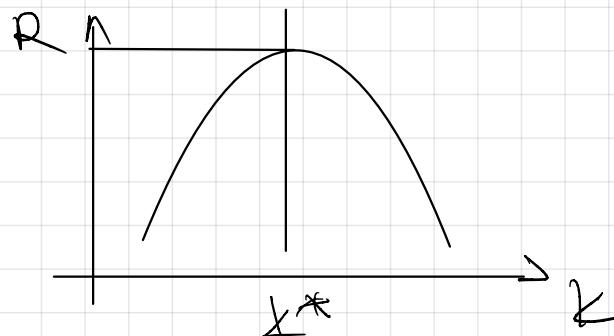
$$R = \frac{5 + 28}{\binom{11}{2}} = \frac{33}{55} = 0.6$$

b	$\gamma = \emptyset$	$\gamma = 1$	$\gamma = 2$
$G_1 G_2$	$2 \cdot 2$		$1 \cdot 2$
$G_1 G_3$	$2 \cdot 4$	$1 \cdot 4$	$1 \cdot 2$
$G_2 G_3$		$2 \cdot 4$	

$$\Sigma = 28$$

b) unimodalan

→ postoji vrijednost br. grupa za koju će biti najbolji

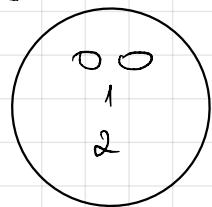


c) dovoljne nam je lista parova

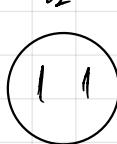
Randou indeks II. nacin:

$N=11$

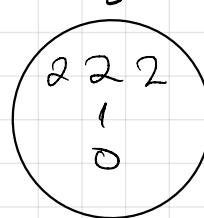
G_1



G_2



G_3

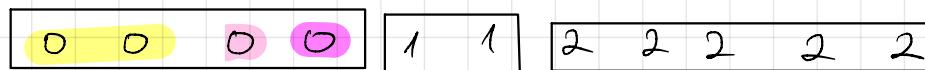


$$\{ \underbrace{\{0, 0, 1, 2\}}_{G_1}, \underbrace{\{1, 1\}}_{G_2}, \underbrace{\{2, 2, 2, 1, 0\}}_{G_3} \}$$

i | 1 2 3 4 5 6 7 8 9 10 11

br. el. v grupama

y_{pred}



$$TP+FP = \sum \binom{N_i}{2}$$

y_{true}



$$TP+FN = \sum \binom{N_i}{2}$$

0 y_{true} 1 2

4	\leftarrow	0	2	1	1
2	\leftarrow	pred	1	2	0
5	\leftarrow	2	4	1	3

$TP+FP$

\downarrow \downarrow \downarrow
 3 4 4
 $TP+FN$

$$\binom{4}{2} + \binom{2}{2} + \binom{5}{2} = 6 + 1 + 10 = 17 = TP+FP$$

$$TP+FN = \binom{3}{2} + \binom{4}{1} + \binom{4}{2} = 3 + 6 + 6 = 15$$

TP : svih el. povrh 2 (ima smisla da min 2)

$$TP = \binom{2}{2} + \binom{2}{2} + \binom{3}{2} = 5$$

$$TP+FP = 17$$

$$\binom{11}{2} = 55$$

$$FP = 17 - 5 = 12$$

$$TN = N - FP - TP - FN = 55 - 12 - 5 - 10$$

$$= 28 \approx$$

$$TP+FN = 15$$

$$FN = 15 - 5 = 10$$

$$\{ \{0, 0, 1, 1, 2\}, \{1, 1\}, \{2, 2, 2, 1, 0\} \}$$

$$(\textcolor{pink}{N}) = \binom{11}{2} = 5$$

$$\text{TP} = \binom{2}{2} + \binom{2}{2} + \binom{3}{2}$$

$$\text{TP} + \text{FP} = \binom{4}{2} + \binom{2}{2} + \binom{5}{2}$$

$$\left. \begin{array}{l} 0: 3 \text{ puts} \\ 1: 4 \text{ puts} \\ 2: 4 \text{ puts} \end{array} \right\}$$

$$\text{TP} + \text{FN} = \binom{3}{2} + \binom{4}{2} + \binom{4}{2}$$



1. (T) Konvergencija je poželjno svojstvo algoritma grupiranja. **Je li točno da algoritam k-sredina uvijek konvergira?**

- A Da, algoritam uvijek konvergira zato što je broj particija N primjera u K skupova ograničen, a optimizacijski postupak definiran je tako da se J u svakoj iteraciji smanjuje
- B Algoritam konvergira samo ako su početna središta dobro odabранa, inače se može dogoditi da algoritam oscilira između dva rješenja
- C Kako se radi o algoritmu koji grupira primjere u vektorskom prostoru, broj rješenja je neograničen, stoga algoritam ne mora konvergirati
- D Algoritam uvijek konvergira zato što je broj primjera N uvijek veći ili jednak broju grupa K , a kao mjeru udaljenosti koristi se euklidska udaljenost, koja je nužno nenegativna ??

2. (T) Algoritmi grupiranja k-sredina i k-medoida razlikuju se, između ostalog, 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?**

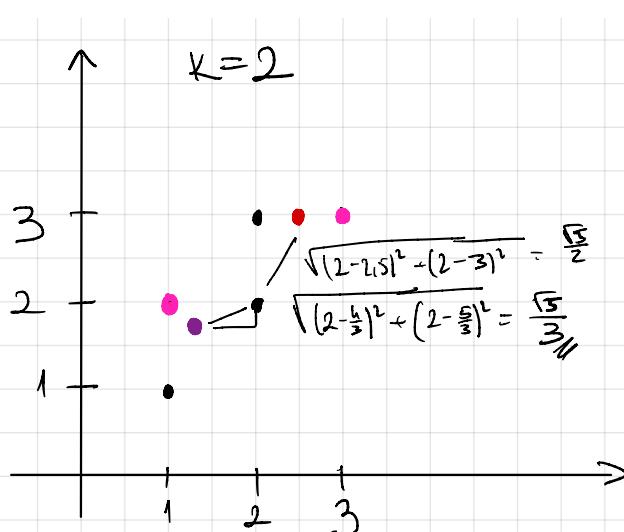
- A Za razliku od algoritma k-sredina, algoritam k-medoida je algoritam mekog grupiranja, što iziskuje provođenje dodatnih koraka unutar algoritma
- B Budući da algoritam k-medoida ne koristi centroide, nego medoide, na kraju svake iteracije mora kombinatoričkom provjerom po primjerima pronaći medoide koje minimiziraju kriterijsku funkciju J
- C 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
- D Kriterijska funkcija algoritma k-medoida jest mnogo složenija od one k-sredina, upravo zato što algoritam k-medoida koristi medoide, a ne centroide

3. (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?**

- A 2.962 B 1.833 C 1.667 D 2.414



		<u>#1 K</u>	<u>#2 K</u>
x_1	(1,1)	1	1
x_2	(1,2)	1	1
x_3	(2,2)	1	1
x_4	(2,3)	2	2
x_5	(3,3)	2	2

$$\mu_k = \frac{\sum_i b_k^i x^i}{\sum_i b_k^i}$$

$$\mu_1 : x : \frac{1+1+2}{3} = \frac{4}{3}$$

$$y : \frac{1+2+2}{3} = \frac{5}{3}$$

$$\left(\frac{4}{3}, \frac{5}{3} \right)$$

$$\mu_2 : x : \frac{2+3}{2} = \frac{5}{2}$$

$$y : \frac{3+3}{2} = 3$$

$$(2.5, 3)$$

$$J = \sum_{k=1}^K \sum_{i=1}^N b_k^{(i)} \|x^{(i)} - \mu_k\|^2$$

$$K=0 : = \left(\frac{4}{3} - 1 \right)^2 + \left(\frac{5}{3} - 1 \right)^2 + \left(\frac{4}{3} - 1 \right)^2 + \left(\frac{5}{3} - 2 \right)^2 + \left(\frac{4}{3} - 2 \right)^2 + \left(\frac{5}{2} - 2 \right)^2$$

$$K=1 : + (2.5 - 2)^2 + (3 - 3)^2 + (2.5 - 3)^2 + (3 - 3)^2$$

$$= \left(\frac{1}{3} \right)^2 + \left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 + \left(\frac{1}{3} \right)^2 + \left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 + \left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2$$

$$= 2 \cdot \frac{1}{9} + 2 \cdot \frac{4}{9} + 2 \cdot \frac{1}{9} = \frac{4+8}{9} + \frac{1}{2} = \frac{12}{9} + \frac{1}{2} = \frac{4}{3} + \frac{1}{2} = \frac{8+3}{6} = \frac{11}{6} = 1.833$$

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

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

$$\frac{K1}{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(4)}}$$

$$\frac{K2}{\mathbf{x}^{(3)}, \mathbf{x}^{(5)}}$$

$$\text{NOVI MEDOIDI}$$

$\underline{\mathbf{x}1}$	$\underline{\mathbf{x}2}$	$\underline{\mathbf{x}4}$
---------------------------	---------------------------	---------------------------

$$\mathbf{x}1: \underline{0} + 0,2 + 0,7 = 0,9$$

$$\mathbf{x}3: \underline{0} + 0,3 = 0,3$$

$$(\mathbf{x}2) 0,2 + \underline{0} + 0,1 = 0,3 \Rightarrow \text{novi medoid}$$

$$(\mathbf{x}5) 0,3 + \underline{0} = 0,3 \rightarrow \text{ostaje isti}$$

$$\mathbf{x}4: 0,7 + 0,1 + \underline{0} = 0,8$$

5. (N) Particijskim algoritmom 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?

- A 0.27 B 0.56 C 0.64 D 0.70

$$\binom{N}{2} = \binom{10}{2} = \frac{10 \cdot 9}{2} = 45$$

$$TP = \binom{2}{2} + \binom{2}{2} + \binom{2}{2} = 3 \rightarrow a$$

$$TP + FP = \binom{3}{2} + \binom{3}{2} + \binom{4}{2} = 3 + 3 + 6 = 12$$

$$TN + FN: \quad \emptyset: 3 \quad \binom{3}{2} + \binom{5}{2} + \binom{2}{2} = 3 + 10 + 1 = 14$$

$$1: 5$$

$$2: \underline{\frac{2}{10}} \quad \cancel{w}$$

$$FN = 14 - 3 = 11$$

$$TN = N - \boxed{TP + FP} - FN = 45 - 12 - 11 = 22 \rightarrow b$$

$$R = \frac{a+b}{\binom{N}{2}} = \frac{3+22}{45} = \frac{25}{45} = 0,56 \quad m$$

6. (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 $RI(K)$ za grupiranja dobivena s različitim brojem grupa K . Naposljetku ćemo onda kao optimalan broj grupa odabrati onaj K koji maksimizira Randov indeks, $K^* = \operatorname{argmax}_K RI(K)$. 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 $RI(K)$ te optimalan broj grupa K^* prema Randovom indeksu, za $K \in \{3, 4, 5\}$. **Koliko iznosi Randov indeks za optimalan broj grupa, $RI(K^*)$?**

- A 0.375 B 0.625 C 0.750 D 0.875

$N = 1000$ $1 \rightarrow$ trebali su završiti zajedno

		Y	$K=3$	$K=4$	$K=5$
1	1 - 1	1	1	1	1
2	2 - 2	1	0	0	1
3	3 - 3	1	0	0	0
4	4 - 4	0	0	1	0
5	5 - 5	0	1	0	1
6	6 - 6	0	1	1	0
7	7 - 7	0	0	1	1
8	8 - 8	0	0	0	0

↓

$\text{TP} = 1$ $\text{TP} = 1$ $\text{TP} = 2$
 $\text{TN} = 3$ $\text{TN} = 2$ $\text{TN} = 3$

$R = \frac{1+3}{8} = 0.5$ $R = \frac{1+2}{8} = 0.1$ $R = \frac{2+3}{8} = 0.625$ $\rightarrow 5$ grupa

$(\binom{8}{2}) = 8$

Grupiranje II.

1. [Svrha: Razumjeti model miješane gustoće i razlog zašto maksimizacija log-izglednosti nije analitički rješiva. Razumjeti kako uvođenje latentnih varijabli rješava taj problem. Razumjeti, na općenitoj razini, E-korak i M-korak. Razumjeti rad algoritma kao maksimizacije log-izglednosti i razumjeti kako ishod ovisi o broju grupa i početnoj inicijalizaciji.] Algoritam maksimizacije očekivanja (EM-algoritam), kada se koristi za grupiranje, zapravo je poopćenje algoritma k-sredina.
- Što je prednost, a što nedostatak, algoritma maksimizacije očekivanja u odnosu na algoritam k-sredina?
 - Napišite izraz za gustoću $p(\mathbf{x})$ za model miješane gustoće (bez latentnih varijabli) i izraz za pripadnu (nepotpunu) log-izglednost.
 - Napišite izraz za mješavinu s latentnim varijablama i izvedite izraz za (potpunu) log-izglednost tog modela. Možemo li dalje raditi izravno s tom log-izglednošću? Zašto?
 - Definirajte E-korak i M-korak algoritma maksimizacije očekivanja primjenjenog na Gaussovou mješavinu.
 - Skicirajte vrijednost log-izglednosti $\ln \mathcal{L}(\boldsymbol{\theta} | \mathcal{D})$ modela Gaussove mješavine kao funkcije broja iteracija, i to za tri različite vrijednosti parametra K (broj grupa): $K = 1$, $K = 10$ i $K = 100$. Na istom grafikonu skicirajte krivulju za $K = 10$ kada se za inicijalizaciju središta koristi algoritam k-sredina.

a) GMM:

- prenosti: 1. među grupiranje (možemo modelirati nesigurnost)

odgovornost da spriječi i pripada grupi j , zapravo aposteriorna vj. :

$$h_j^i = P(z_i=j | \bar{x}^i)$$

- 2. modeliranje nesferičnosti grupe (pun, nedjeljena kov. m.)

\Rightarrow veća ekspresivnost modela

- nedostatci: - više param., moguća prenaučenošt

- optimizacija je složenija
- sponje konvergira

GMM vs K-sredina

\rightarrow umjesto binarnih indikatora (kod K-sredina) konstimmo odgovornosti u intervalu $[0, 1]$

\rightarrow umjesto prognoze samo sredina (K-s.), moramo i kov. matrice

- parametri: inicijalizirati GMM alg. s k -sredinama

npr. K-means ++

→ prvočimo k -sredinu dok ne konvergira

→ nabavimo samo srediste

→ vršimo GMM (možda malo pomiješani sredine, jer u obzir dolaze i nov.)

⇒ brza konvergenca

c) GMM s latentnim var.

ne možemo da je raditi pravno s log-izgleđivošću

- dve sume sa log izmedju (ne možemo analitički minimizirati u zatvorenoj formi)

- posljedica konceptualnog problema:

ne znamo koji primjer pripada kojoj grupi

e) log-izgleđost $\ln L(\theta | D)$

(minimizirano $-\ln L(\theta | D)$)

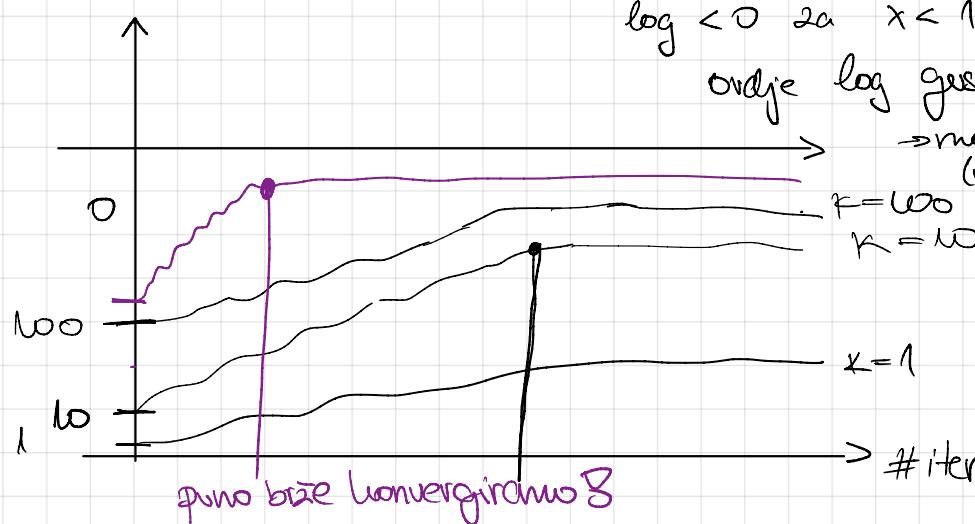
$K=1$

$K=10$

$K=100$

STOFASTIČNO!

(poči sredista)



GMM s k -sredinama (inic. pomoću K-means++)

↳ pronalazi centrole koji minimiziraju kvadratna odstupanja

\Leftrightarrow maksimizirano log-izgl. uz pretp. izvrsnih kov. m. uz pretp. sredinosti podataka

2. [Svrha: Isprobati rad algoritma hijerarhijskog aglomerativnog grupiranja (HAC) na konkretnom primjeru, za slučaj kada primjeri nisu vektori. Uočiti razliku između udaljenosti i sličnosti te razliku između jednostrukog i potpunog povezivanja.] Jednako kao i algoritam k-medoida, algoritam hijerarhijskog algomerativnog grupiranja može se primijeniti u slučajevima kada primjeri nisu prikazani kao vektori značajki te kada umjesto mjere udaljenosti između vektora raspolažemo općenitjom mjerom sličnosti (ili različitosti). Neka je *sličnost* primjera iz \mathcal{D} definirana sljedećom matricom sličnosti:

$$S = \begin{matrix} & a & b & c & d & e \\ a & 1.00 & 0.26 & 0.15 & 0.20 & 0.17 \\ b & 0.26 & 1.00 & 0.24 & 0.31 & 0.31 \\ c & 0.15 & 0.24 & 1.00 & 0.20 & 0.50 \\ d & 0.20 & 0.31 & 0.20 & 1.00 & 0.24 \\ e & 0.17 & 0.31 & 0.50 & 0.24 & 1.00 \end{matrix} \rightarrow \text{sigurno sličnost}$$

- (a) Izgradite dendrogram uporabom jednostrukog povezivanja. Kada bi bilo potrebno napraviti particiju grupa, na kojoj biste razini presjekli taj dendrogram?
- (b) Izgradite dendrogram uporabom potpunog povezivanja. Kada bi bilo potrebno napraviti particiju grupa, na kojoj biste razini presljekli taj dendrogram?

HAC, kao i k-medoida | snaga: mogu raditi s primjenom logi nisu u vektorskom prostoru

a) jednostruko povezivanje

$$S_1 = \begin{matrix} & a & b & c & d & e \\ a & 1 & 0.26 & \boxed{0.15} & 0.2 & \boxed{0.17} \\ b & 1 & \boxed{0.24} & 0.31 & 0.31 & \\ c & 1 & \boxed{0.12} & \boxed{0.15} & 0.24 & \\ d & 1 & & 1 & 0.24 & \\ e & 1 & & & 1 & \end{matrix} \quad 5 \times 5$$

vremeno max(0,15, 0,17)
= 0,17

ne bi smjeloni biti razmaka

ima smisla presjecanje
→ najveća realika

optimalan br. grupe: 4

1. korak: grupiranje noj sličnije

0,5 : (c, e) - bršemo lemb. red-stupac $\rightarrow 4 \times 4$

$$S_2 = \begin{matrix} & a & b & ce & d \\ a & 1 & \boxed{0.26} & \boxed{0.17} & 0.2 \\ b & 1 & \boxed{0.31} & 0.31 & \\ ce & 1 & & 0.24 & \\ d & 1 & & 1 & \end{matrix} \quad 4 \times 4$$

želimo max sličnost \leftarrow JEDNOSTRUKO Pov. - nailiberajuće

ROTUNDO Povezivanje
za udaljenost ?

moramo odabrati: b & ce

$$S_3 = bce \begin{pmatrix} a & bce & d \\ 1 & 1 & 1 \\ & \boxed{0,26} & \boxed{0,2} \\ & & \boxed{0,31} \end{pmatrix}_{3 \times 3}$$

$$S_1 = bcd e \begin{pmatrix} a & bcd e \\ 1 & 1 \\ & \boxed{0,26} \end{pmatrix}_{2 \times 2}$$

b) potpuno povezivanje

$$S_1 = \begin{array}{ccccc} a & b & c & d & e \\ \hline a & 1 & 0,26 & 0,15 & 0,2 & 0,17 \\ b & & 1 & 0,24 & 0,31 & 0,31 \\ c & & & 1 & 0,2 & 0,5 \\ d & & & & 1 & 0,21 \\ e & & & & & 1 \end{array}$$

1. tražimo max : ISTO ↑
- spajamo ce
→ Ači uzmimo
najveće sl. vrij.
⇒ min (a, e) →
min (0,15, 0,17)
0,15

	jednostavno	potpuno
udaljenost	min	max
sličnost	max	min

KAD A STAPAMO : UNIJEK MAX

3. [Svrha: Razumjeti kako se unutarnji kriterij algoritma grupiranja može (pokušati) upotrijebiti za provjeru grupiranja (odabir optimalnog broja grupa). Razumjeti da Akaikeov kriterij u stvari oponaša regulariziranu funkciju pogreške, koja pak aproksimira pogrešku generalizacije.]

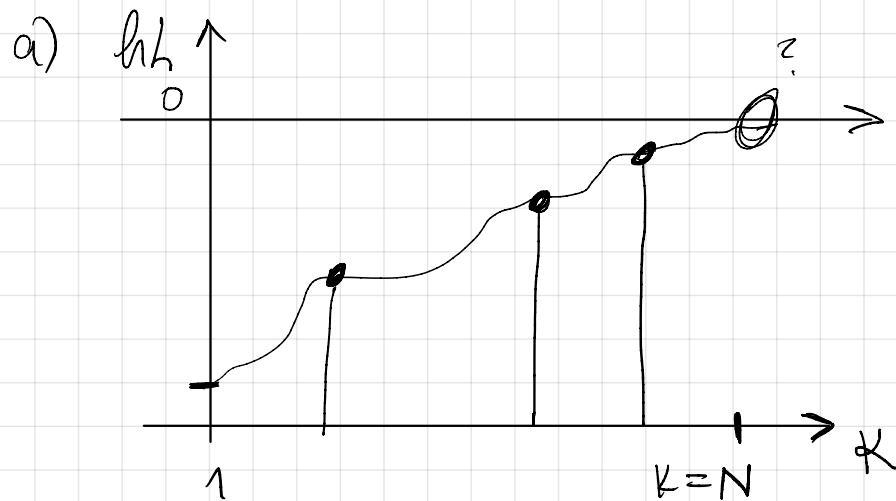
- (a) Skicirajte krivulju log-izglednosti kod EM-algoritma kao funkciju broja grupa K . Obrazložite izgled krivulje. Možete li temeljem ove krivulje odrediti optimalan broj grupa? Kako?
- (b) Optimizacija broja grupa K može se provesti nekim kriterijem koji kombinira funkciju pogreške (odnosno log-izglednost) i složenost modela. Takav kriterij odgovara strukturnom riziku modela, koji je minimalan za optimalan broj grupa. Jedan takav kriterij jest Akaikeov informacijski kriterij (AIC):

$$K^* = \underset{K}{\operatorname{argmin}} (-2 \ln \mathcal{L}(K) + 2q(K))$$

gdje je $-\ln \mathcal{L}(K)$ negativna log-izglednost podataka za K grupa, a $q(K)$ je broj parametara modela s K grupa.

\Rightarrow s real. kov. matricama

Pretpostavite da podatci \mathcal{D} u stvarnosti dolaze iz $K = 5$ grupa. Podatke grupiramo dvjema varijantama EM-algoritma: standardni algoritam i preinačeni algoritam s dijeljenom kovarijacijskom matricom (zajednička kovarijacijska matrica procijenjena nad čitavim skupom primjera \mathcal{D} na početku izvođenja algoritma). Skicirajte za ta dva algoritma funkciju koju minimizira Akaikeov minimizacijski kriterij.



• ne može direktno maksimizirati

\hookrightarrow svi primjeri
je svoj centroid

b) br. param. GMM puna kov. m. Σ , nedjelyenc :

$$\text{K grupe} \quad \frac{n \cdot (n+1)}{2} + nK + \underbrace{\cancel{K}-1}_{\substack{\text{takut +} \\ \text{dijag. kov.m}}} \quad \text{centroidi kof. mjesavine}$$

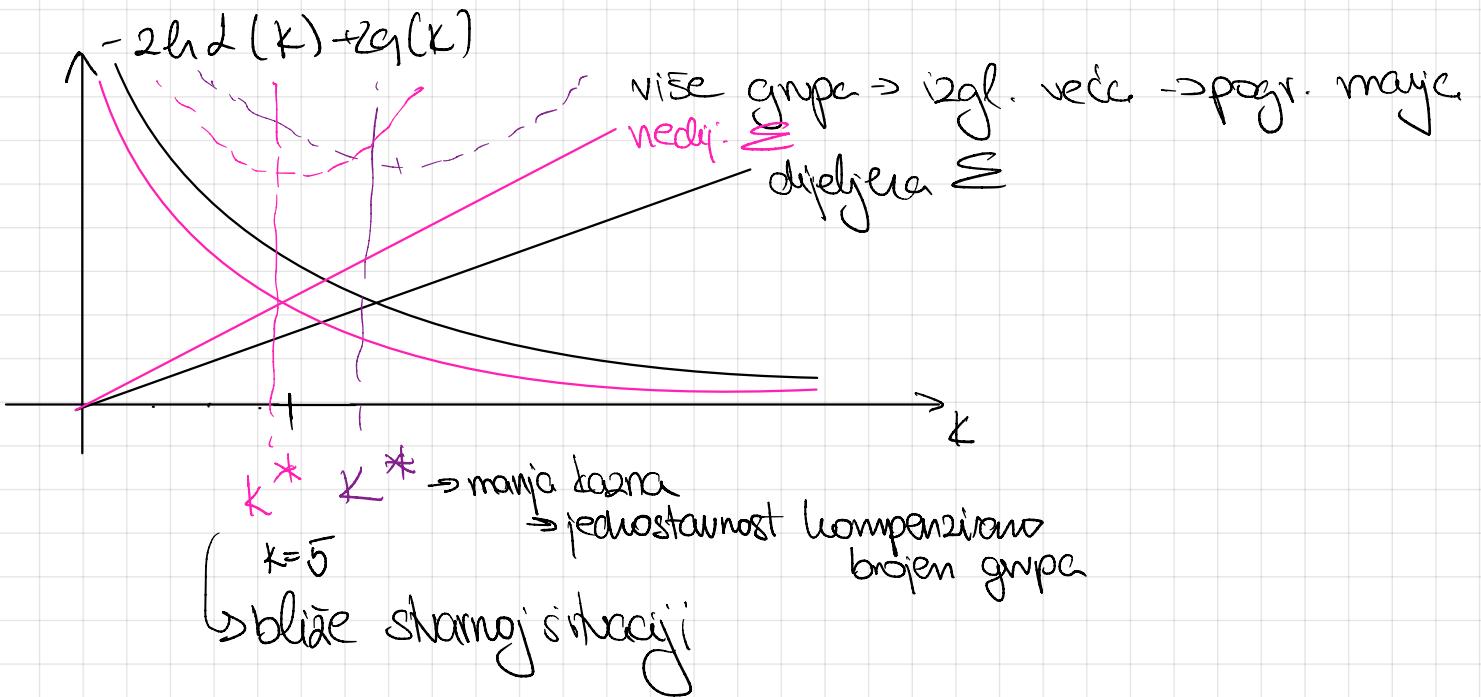
$q_1(K)$ \rightarrow linearna funkcija od K

\hookrightarrow prava većeg radij + krenuti će s vredom uj.

zagl. kov. m.

$$\frac{n \cdot (n-1)}{2} + nK + \cancel{K}-1$$

$$q_2(K)$$



1. (T) Algoritam GMM, odnosno model Gaussove mješavine s algoritmom maksimizacije očekivanja kao optimizacijskim postupkom, poopćenje je algoritma k-sredina. **Uz koje uvjete algoritam GMM degenerira u algoritam k-sredina?**

- A Umjesto maksimizacije log-izglednosti, minimizira se negativna log-izglednost, a početna središta se odabiru algoritmom k-sredina
- B Koeficijenti mješavine su jednaki za sve komponente Gaussove mješavine, a kovarijacijske matrice su dijagonalne
- C Kovarijacijska matrica komponenti Gaussove mješavine je jedinična matrica, a maksimizira se negativna log-izglednost
- D Kovarijacijska matrica komponenti Gaussove mješavine je dijeljena i izotropna, a odgovornosti su zaokružene na cijeli broj

2. (T) Algoritam maksimizacije očekivanja (EM-algoritam) maksimizira očekivanje potpune log-izglednosti, što se pokazuje da dovodi i do maksimizacije nepotpune log-izglednosti. **Koja je razlika između potpune i nepotpune log-izglednosti, i zašto maksimiziramo očekivanje potpune log-izglednosti umjesto izravno log-izglednost?**

- A Potpuna log-izglednost je izglednost izračunata na svim primjerima iz neoznačenog skupa primjera, dok je nepotpuna log-izglednost izračunata samo za označene primjere koji se koriste za evaluaciju modela, a očekivanje računamo zato jer je postupak grupiranja stohastičan
- B Potpuna log-izglednost je log-izglednost s neopaženim varijablama, a u slučaju GMM-a to su centroidi i kovarijacijske matrice komponenata, koje procjenjujemo metodom MLE, koja maksimizira očekivanje log-izglednosti
- C Potpuna log-izglednost je log-izglednost modela GMM s latentnim varijablama, koje definiraju koji primjer pripada kojoj grupi, međutim kako to zapravo ne znamo, moramo računati s očekivanjem tih varijabli
- D Potpuna log-izglednost računa se za označene primjere a nepotpuna log-izglednost za neoznačene primjere, a u oba slučaja kod modela GMM računamo očekivanje log-izglednosti jer postupak za različite početne centroide može dati različite log-izglednosti

3. (T) 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 Algoritam uvijek konvergira, i to do točke u prostoru parametara koja maksimizira log-izglednost parametara, no brzina konvergencije ovisi o toma kako su inicializirani parametri
- C 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
- 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

4. (T) Optimizaciju parametara modela Gaussove mješavine (GMM) ne provodimo u zatvorenoj formi. S druge strane, parametre Gaussovog Bayesovog klasifikatora, koji je sličan modelu GMM, optimiramo u zatvorenoj formi. **Zašto parametre GMM-a ne optimiramo u zatvorenoj formi, dok kod Gaussovog Bayesovog klasifikatora to radimo?**

- A Za razliku od Gaussovog Bayesovog klasifikatora, GMM je nenadizirani algoritam, pa log-izglednost podataka nije definirana i nije moguća maksimizacija u zatvorenoj formi
- B Kod GMM-a, pored koeficijenata mješavine i vektora sredina, trebamo procijeniti i kovarijacijske matrice, za što ne postoji procjenitelj u zatvorenoj formi
- C Kod GMM-a ne znamo koji primjer pripada kojoj grupi, pa je gustoća primjera jednaka zbroju gustoći komponenti, za što ne postoji maksimizator u zatvorenoj formi
- D Parametri oba modela mogu se optimirati u zatvorenoj formi, međutim kod modela GMM računalno je jednostavnije koristiti EM-algoritam

5. (P) Algoritam GMM koristimo za grupiranje $N = 10$ primjera u dvodimenziskome 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:

- \mathcal{H}_1 : $K = 2$ grupa, puna kovarijacijska matrica
- \mathcal{H}_2 : $K = 2$ grupa, izotropna kovarijacijska matrica
- \mathcal{H}_3 : $K = 3$ grupe, izotropna kovarijacijska matrica

Za sva tri modela kovarijacijska matrica je nedijeljena, dakle svaka komponenta ima svoju kovarijacijsku matricu. Za početne centroide odabiremo nasumično dva odnosno 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 skupu \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}_α , to jest vjerojatnost pripadanja tog primjera najvjerojatnijej grupi dobivenoj grupiranjem pomoću modela \mathcal{H}_α . **Što možemo zaključiti o odgovornostima h_α za ova tri modela?**

- A $h_{\alpha_1} > h_{\alpha_2} > h_{\alpha_3}$ B $h_{\alpha_1} < h_{\alpha_2} < h_{\alpha_3}$ C $h_{\alpha_2} > h_{\alpha_1} > h_{\alpha_3}$ D $h_{\alpha_2} < h_{\alpha_1} < h_{\alpha_3}$

$\boxed{h_{\alpha_1} > h_{\alpha_2}}$: isti K , ali \mathcal{H}_1 imao punu kov. matrice
 → složeniji model
 \Rightarrow izgl. grupe veća

h_2 = odgovornost

→ isto : - kof. mješavine
 - centroidi

$\boxed{h_{\alpha_2} > h_{\alpha_3}}$: - kof. mješavine manji za veći K
 (masa viši raspoređene na više grupe)

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

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

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

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

\mathcal{H}_4 : Model sa $K = 10$ središta inicijaliziranim algoritmom k-sredina i dijeljenom kov. 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_{α}^0 prosječna log-izglednost za model \mathcal{H}_{α} na početku izvođenja EM-algoritma, a neka je LL_{α}^* prosječna log-izglednost za taj model na kraju izvođenja EM-algoritma. **Što možemo unaprijed zaključiti o ovim log-izglednostima?**

A $LL_2^0 \geq LL_4^0$, $LL_1^* \geq (LL_2^* \geq LL_3^*)$ sig. tačno

B $LL_3^0 \geq LL_4^0$, $LL_1^* \geq LL_3^* \geq LL_4^*$ \hookrightarrow ne znam možemo

C $LL_2^0 \geq LL_4^0 \geq LL_3^0$, $LL_1^* \geq LL_2^*$ li ovo sa sigurnošću

D $LL_2^0 \leq LL_4^0$, $LL_2^* \leq LL_1^* \geq LL_3^*$ tvrditi

K: veći $K \rightarrow$ veća izglednost : H_4 najlošiji

k-sredina i generalno bolje od nasumičnog odabira : H_3 najgori

pure kov. matrica bolja od dijeljene : H_1 najbolja

7. (T) Broj grupa K hiperparametar je mnogih algoritama grupiranja, pa tako i algoritma GMM. Optimalan broj grupa može se odrediti na razine načine, a jedan od njih je Akaikeov kriterij. **Na kojem se principu temelji odabir broja grupa Akaikeovim kriterijem?**

- A Model s optimalnim brojem grupa je onaj koji podatke čini najvjerojatnijima, ali to čini sa što manje parametara
- B Optimalan broj grupa je onaj kod kojeg, nakon dalnjeg povećanja broja grupa, vrijednost log-izglednosti stagnira ili blago raste
- C Model s optimalnim brojem grupa je onaj koji minimizira log-izglednost nepotpunih podataka, a maksimizira log-izglednost potpunih podataka
- D Optimalan broj grupa je onaj koji maksimizira očekivanje log-izglednost modela, uz pretpostavku izotropne kovarijacijske matrice

8. (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):

$$S_1 = \begin{pmatrix} & a & b & c & d & e \\ a & x^{(1)} & & & & \\ b & x^{(2)} & 1 & 0.4 & & \\ c & x^{(3)} & & 1 & 0.5 & \\ d & x^{(4)} & & & 1 & 0.7 \\ e & x^{(5)} & & & & 1 \end{pmatrix}_{5 \times 5}$$

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

$$S_2 = \begin{pmatrix} & a & bc & d & e \\ a & 1 & 0.4 & 0.7 & 0.5 \\ bc & & 1 & 0.3 & 0.1 \\ d & & & 1 & 0.8 \\ e & & & & 1 \end{pmatrix}_{4 \times 4}$$

$$S_3 = \begin{pmatrix} & a & bc & de \\ a & 1 & 0.4 & 0.5 \\ bc & & 1 & 0.1 \\ de & & & 1 \end{pmatrix}_{3 \times 3}$$

$$S_4 = \begin{pmatrix} & bc & ade \\ bc & 1 & 0.1 \\ ade & & 1 \end{pmatrix}_{2 \times 2}$$

Vrednovanje modela

1. [Svrha: Izvježbati izračun mjera uspješnosti modela na konkretnom primjeru.]

Raspolažemo skupom od 11 ispitnih primjera koje želimo klasificirati u tri klase. Oznaka $y^{(i)}$ i izlaz modela $h(\mathbf{x}^{(i)})$ za svaki od 11 primjera su sljedeći:

$$\{(y^{(i)}, h(\mathbf{x}^{(i)}))\}_{i=1}^{11} = \{(1, 1), (0, 2), (2, 2), (1, 2), (1, 1), (0, 0), (1, 1), (2, 1), (0, 1), (2, 0), (2, 1)\}.$$

(a) Izračunajte točnost klasifikatora.

(b) Izračunajte preciznost, odziv i mjeru F_1 , i to mikro i makro varijante.

a)

true - $y^{(i)}$					
			0	1	2
pred $h(\mathbf{x}^{(i)})$	0	1	0	1	
	1	1	3	2	
	2	1	1	1	

3x3

$y=0$; spada u 2. klasu

b)

$y=0$			$y=1$			$y=2$		
1	0	0	1	3	3	0	1	2
1	1	1	3	3	3	0	1	2
0	2	7	0	1	4	0	3	5

MAKRO : za svaku klasu & prosjek

$$11 - 1 - 2 - 3 = 5$$

✓ trag matrice

$$Acc = \frac{1+3+1}{11} = \frac{5}{11} = 0,455$$

1	TP	FP
0	FN	TN
	(2 broj v stupcu)	

2broj po redku

svi ostaci

PRECIZNOST: $P = \frac{TP}{TP + FP}$

ODZIV: $R = \frac{TP}{TP + FN}$

HARM. SREDINA: $F_1 = \frac{2PR}{P+R}$

$$\underline{Y=0}$$

$$P_0 = \frac{TP}{TP+FP} = \frac{1}{1+1} = \frac{1}{2} \quad R_0 = \frac{TP}{TP+FN} = \frac{1}{1+2} = \frac{1}{3}$$

$$F_{1,0} = \frac{2P_0R_0}{P_0+R_0} = 0,4$$

$$\underline{Y=1}$$

$$P_1 = \frac{TP}{TP+FP} = \frac{3}{3+3} = \frac{1}{2} \quad R_1 = \frac{TP}{TP+FN} = \frac{3}{3+1} = \frac{3}{4}$$

$$F_{1,1} = \frac{2P_1R_1}{P_1+R_1} = 0,6$$

$$\underline{Y=2}$$

$$P_2 = \frac{TP}{TP+FP} = \frac{1}{1+2} = \frac{1}{3} \quad R_2 = \frac{TP}{TP+FN} = \frac{1}{1+3} = \frac{1}{4}$$

$$F_{1,2} = \frac{2P_2R_2}{P_2+R_2} = 0,286$$

→ same klasse → obično lošije performance

MAKRO

$$P^M = \frac{1}{3} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{3} \right) = 0,44$$

$$R^M = \frac{1}{3} \left(\frac{1}{3} + \frac{3}{4} + \frac{1}{4} \right) = 0,44$$

$$F_1^M = \frac{1}{3} (0,4 + 0,6 + 0,286) = 0,429 \leftarrow \text{manje od Acc}$$

MIKRO zbrajamo matrice (može i težinski) : Σ istih elem.

	*
1	5
0	6

$$P^M = R^M = F_1^M = ACC \quad *$$

dovoljno npr. $P^M = \frac{TP}{TP+FP} = \frac{5}{5+6} = 0,455$

• mora biti isto

* ne vrijedi za multilabel class;

2. [Svrha: Izvježbati izračun mjere F_1 na temelju parcijalno zadane matrice zabune.] Od $N = 1000$ primjera, klasifikator je za prvu, drugu i treću klasu ispravno pozitivno klasificirao njih 620, 146 odnosno 134. Od preostalih 100 neispravno klasificiranih primjera, 50 ih je klasificirano u drugu klasu umjesto u prvu, 20 u drugu umjesto u treću, a 30 u treću umjesto u drugu klasu. Izračunajte makro- F_1 .

true			
	$\gamma=1$	$\gamma=2$	$\gamma=3$
$\gamma=1$	620	\emptyset	\emptyset
$\gamma=2$	50	146	20
$\gamma=3$	\emptyset	30	134

3×3

F_1^M -----

3. [Svrha: Znati kako na temelju probabilističkog izlaza klasifikatora skicirati krivulju ROC. Znati mjerom AUC možemo usporediti klasifikator s nasumičnim klasifikatorom. Znati kako pomoću krivulje ROC uspoređivati klasifikatore međusobno.] Na ispitnome skupu od $N = 10$ primjera evaluiramo tri binarna klasifikatora: logističku regresiju (h_{LR}), naivan Bayesov klasifikator (h_{NB}) i stroj potpornih vektora s probabilističkim izlazom dobivenim metodom Plattove kalibracije (h_{SVM}). Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije triju klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1 | \mathbf{x}^{(i)})$ na tom skupu su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y^{(i)}$	1	1	0	0	1	1	1	0	0	1
$h_{LR}(\mathbf{x})$	0.8	0.6	0.8	0.6	0.8	0.8	0.8	0.2	0.2	0.2
$h_{NB}(\mathbf{x})$	0.3	0.8	0.3	0.5	0.8	0.3	0.8	0.5	0.3	0.5
$h_{SVM}(\mathbf{x})$	0.6	0.1	0.7	0.6	0.1	0.7	0.7	0.6	0.1	0.7

Na temelju ovog uzorka želimo procijeniti krivulju ROC te mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma).

- (a) Skicirajte krivulje ROC za ova tri klasifikatora, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora.
- (b) Skicirajte krivulje ROC za ova tri klasifikatora.
- (c) Izračunajte mjere AUC za sva tri klasifikatora.
- (d) Kako izgleda krivulja ROC za nasumični klasifikator. Zašto?
- (e) Koji je od navedenih klasifikatora lošiji od nasumičnog klasifikatora, a koji biste klasifikator odabrali kao najbolji?

• prag prilagođavamo ovisno o tome je li nam bitnja preciznost ili odziv

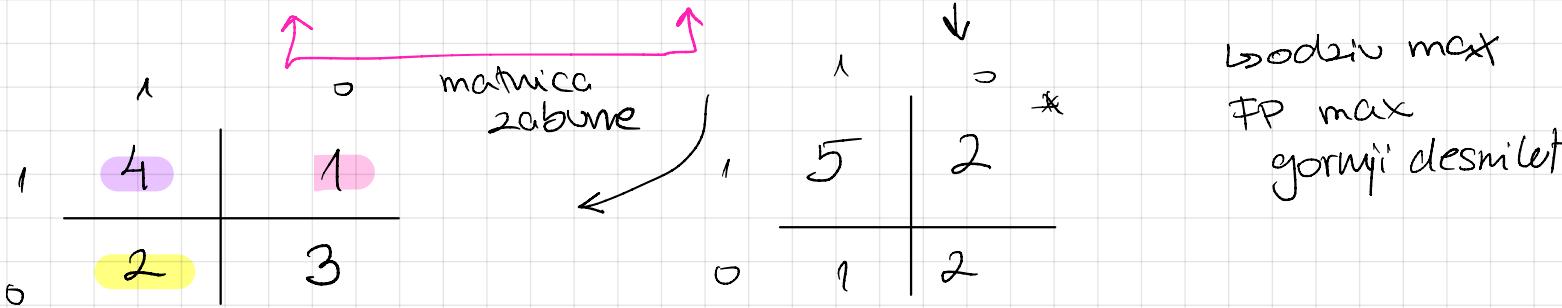
- uz pretp.: ako klasifikator već nije siguran u klas. da je barem dobar u procjeni svoje nesigurnosti
 \Rightarrow dobro kalibriran

ROC: TPR kao funkcija od FPR
 \downarrow
odziv

\hookrightarrow linearno interpoliramo za vrij. koje nemamo

veći prag - ljevo na ROC krivici

a)	i	y^i	LR	$LR \geq 0,8$	$LR \geq 0,6$	$LR \geq 0,2$	nije potrebno
1	1	1	0,8	1	1	1	
2	2	1	0,6	0	1	1	
3	3	0	0,8	1	1	1	
4	4	0	0,6	0	1	1	
5	5	1	0,8	1	1	1	
6	6	1	0,8	1	1	1	
7	7	1	0,8	1	1	1	
8	8	0	0,2	0	0	1	
9	9	0	0,2	0	0	1	
10	10	1	0,2	0	0	1	



$$FPR = \frac{FP}{FP+TN} = \frac{1}{1+3} = 0,25 \quad TPR = \frac{TP}{TP+FN} \rightarrow \text{svi poz. prim}$$

odsiv ↴

$$= \frac{4}{4+2} = 0,67$$

stopa lažnog alarmu

alarmu

= koliko puta smo nešto proglasili pozitivním, a zapravo je neg.

→ v odnosu na uk. br.

negativních průměrů

Nasumiční klasifikator → pravac

$$* FPR = \frac{FP}{FP+TN} = \frac{2}{2+2} = 0,5$$

$$TPR = \frac{TP}{TP+FN} = \frac{5}{5+1} = 0,83$$

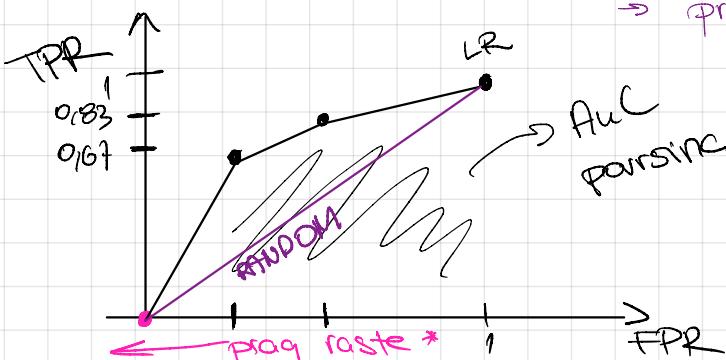
* max prag: místa neče být poz.

→ odsiv \emptyset (TPR)

⇒ nemáme v řadě FP → $FPR = \emptyset$

• prag \emptyset : $FPR = 100\% \quad \frac{\emptyset}{N+\emptyset} = 1$

• • •



4. [Svrha: Razumjeti na koji se način provodi ugniježđena unakrsna provjera, kako se razdjeljuju primjeri kroz iteracije petlji te kako ugraditi dodatne predobradbe značajki, a pritom ne kompromitirati podjelu na skup za učenje i skup za ispitivanje.] Raspolažemo sa 1000 označenih primjera. Za vrednovanje SVM-a s hiperparametrima C i γ koristimo ugniježđenu unakrsnu provjeru sa po 5 ponavljanja u obje petlje. Hiperparametre optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-5}, 2^{-4}, \dots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^3\}$.

- Koliko ćemo ukupno puta trenirati model?
- Koliko ćemo primjera u svakoj od iteracija koristiti za treniranje, koliko za provjeru, a koliko za ispitivanje?
- Kako glase odgovori na prethodna dva pitanja, ako bismo u vanjskoj petlji umjesto petrostrukih unakrsnih provjera koristili unakrsnu provjeru *izdvoji jednoga* (engl. *leave one out, LOOCV*)?
- Klasifikator SVM posebno je osjetljiv na razlike u rasponima između značajki (zašto?), pa se preporuča standardizirati značajke. Što to točno znači i kako biste standardizaciju značajki ugradili u ugniježđenu unakrsnu provjeru?
- Gdje biste u ugniježđenu unakrsnu provjeru ugradili odabir značajki modela i optimizaciju praga po mjeri AUC?

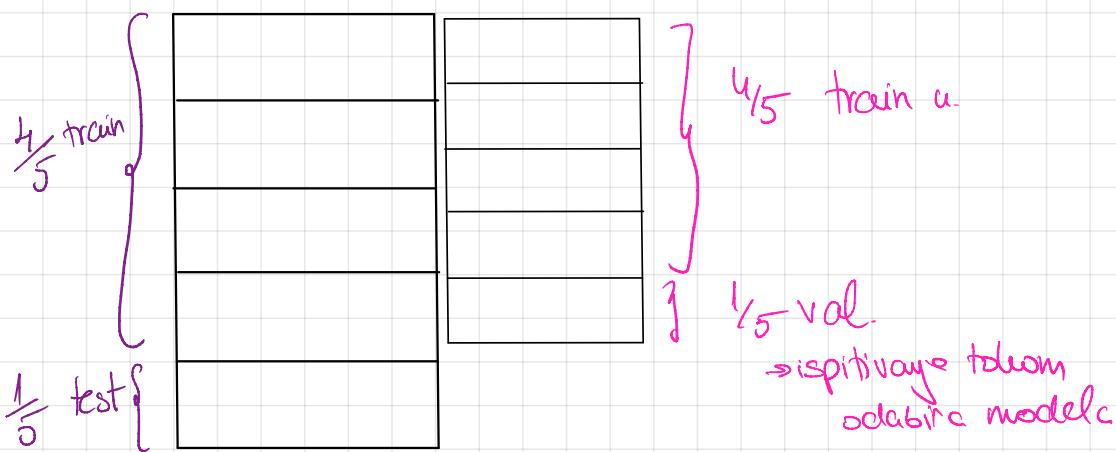
$$N = 1000$$

nested c.v. 5×5

$$C = \{2^{-5}, 2^{-4}, \dots, 2^{15}\} \rightarrow 5 + 1 + 15 = 21$$

$$\gamma = \{2^{-15}, \dots, 2^3\} \rightarrow 15 + 1 + 3 = 19$$

$$\Rightarrow 21 \cdot 19$$



a)

$5 \times$ vanjska

odabir: $21 \cdot 19$

$5 \times$ unutarnja

$$5 \cdot (21 \cdot 19 \cdot 5 + 1) = 9980$$

freniramo ne svih 800
i testiramo ne 200

+ 1 train + na svih $\frac{4}{5}$ (njibolji)

train vanjska:

$$\frac{4}{5} \cdot 1000 = 800$$

→ t. unutarnja!

$$\frac{4}{5} \cdot \frac{4}{5} \cdot 1000 = 640$$

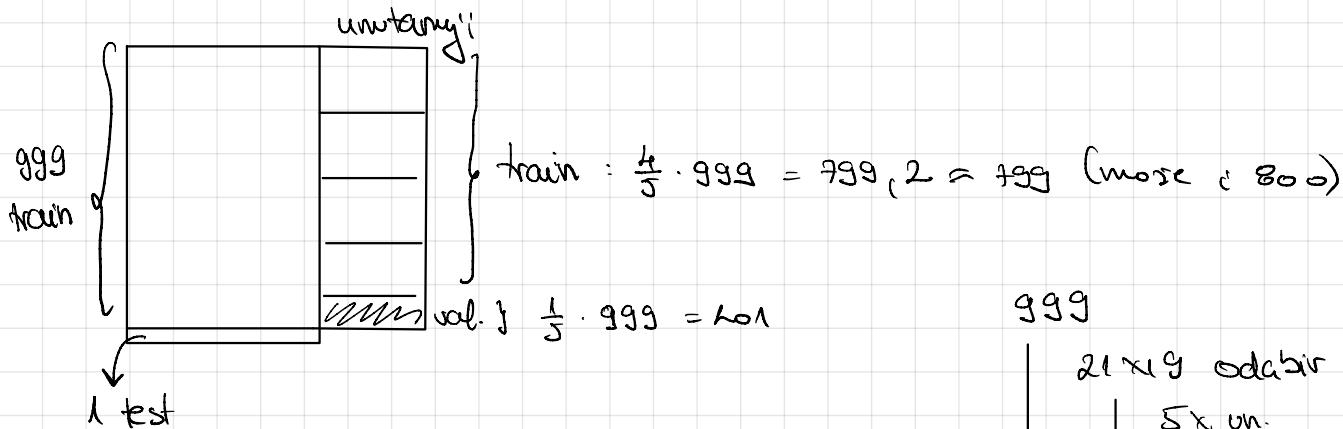
$$\rightarrow \text{val. un. } \frac{4}{5} \cdot \frac{4}{5} \cdot 1000 = 160$$

test vanjski:

$$\frac{1}{5} \cdot 1000 = 200$$

c) LOOCV : analiza powtarzajce koliko rame prinyer

→ 1 prinyer za test, ostali za train



$$999(21 \cdot 19 \cdot 5 + 1) = \dots$$

d) SVM s lin. j. ili bez jezgre → racuna skalarni proiz.

→ svaka dimenzija ima jednaku značaj

$$\text{z-score računanje } x' = \frac{x - \mu}{\sigma}$$

ugradnja: nikako na test setu \emptyset

→ tek u untrainedj petlj'

→ iste je i o karakteru i pravoj

e) spada u odabir modela ⇒ kao i svi hiperparam.

1. (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):

$$\begin{array}{c} y=1 \quad y=2 \quad y=3 \\ \begin{array}{l} y=1 \\ y=2 \\ y=3 \end{array} \left(\begin{array}{ccc} 15 & 3 & 1 \\ 6 & 5 & 4 \\ 4 & 2 & 23 \end{array} \right) \end{array}$$

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 0.01 B 0.05 C 0.09 D 0.13

$$F_1^{\mu} = \text{Acc} = \frac{15+5+23}{63} = 0,68$$

$$\begin{array}{cc} \text{TP} & \text{FP} \\ \text{FN} & \text{TN} \end{array}$$

MAKRO

		$y=1$		$y=2$		$y=3$	
		1	0	1	0	1	0
1	1	15	4	5	10	23	6
	0	10	34	5	43	5	29
		trebalo je biti 1 \rightarrow FN		$\hookrightarrow 63 - 15 - 4 - 10$			

trebalo je
biti 1 \rightarrow FN

$$P_1 = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{15}{15+4} = 0,79 \quad P_2 = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{5}{5+10} = 0,33 \quad P_3 = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{23}{23+6} = 0,79$$

$$R_1 = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{15}{15+10} = 0,6 \quad R_2 = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{5}{5+5} = 0,5 \quad R_3 = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{23}{23+5} = 0,82$$

$$F_{1,1}^M = \frac{2P_1R_1}{P_1+R_1} = 0,68 \quad F_{1,2}^M = \frac{2P_1R_1}{P_1+R_1} = 0,4 \quad F_{1,3}^M = \frac{2P_1R_1}{P_1+R_1} = 0,8$$

$$F_1^M = \frac{1}{3} \sum_{i=1}^3 F_{1,i} = \frac{1}{3} (0,68 + 0,4 + 0,8) = 0,63$$

$$F_1^{\mu} - F_1^M = 0,68 - 0,63 = 0,05$$

2. (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?**

- A Preciznost modela lošija je od odziva, ali model ne radi znatno lošije na primjerima iz klase s manjim brojem primjera, pa su zbog toga mjere mikro- F_1 i makro- F_1 izjednačene
- B Odziv modela lošiji je od preciznosti, osim na primjerima iz klase s velikim brojem primjera, gdje se preciznost i odziv znatno ne razlikuju
- C 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
- D 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

3. (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 -mjelu umjesto F_1 -mjere. Izračunajte F_2 -mjelu logističke regresije i referentnog modela. **Koliko će očekivano logistička regresija biti bolja od referentnog modela po F_2 -mjeri?**

- A 0.095 B 0.051 C 0.101 D 0.176

$F_2 \rightarrow$ dvostrukno naglašavanje odgov

$$F_2 = \frac{(1+2) PR}{2^2 P + R}$$

		true		
		1	0	
pred	1	3	4	
	0	1	2	$\Sigma = 10$
				$\alpha = 0.1$

$$P = \frac{TP}{TP+FP} = \frac{3}{3+2} = \frac{3}{5}$$

$$R = \frac{TP}{TP+FN} = \frac{3}{3+1} = \frac{3}{4}$$

$$F_2 = 0,652$$

NASUMIČNI

$$\text{H } y \text{ su } 1 : \quad TP : 2 \quad FN : 2$$

$$\text{G } y \text{ su } 0 : \quad TN : 3 \quad FP : 3$$

$$P = \frac{TP}{TP+FP} = \frac{2}{2+3} = 0,4$$

$$R = \frac{TP}{TP+FN} = \frac{2}{2+2} = 0,5$$

$$F_2 = 0,476$$

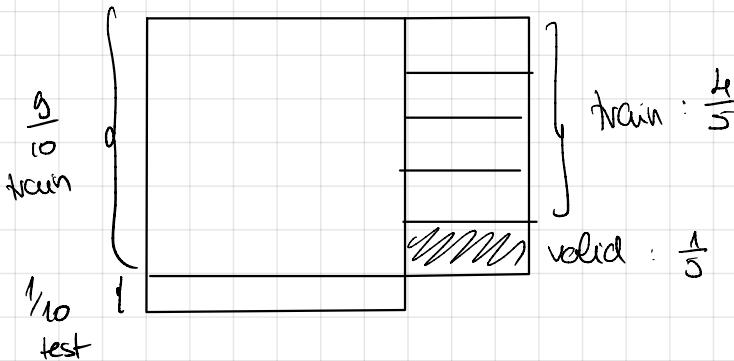
$$\Delta = 0,652 - 0,476 = 0,176$$

4. (T) Procjena pogreške modela metodom unakrsne provjere omogućava nam da procijenimo prediktivnu moć modela, mjerenu kao točnost modela na neviđenom skupu primjera. Daljnja razrada te ideje je ugniježđena višestruka unakrsna provjera, koja se u praksi vrlo često koristi. **Koja je motivacija za korištenje ugniježđene višestruke unakrsne provjere, umjesto obične unakrsne provjere?**

- A Razdvaja skup za učenje od skupa za ispitivanje te time osigurava da doista mjerimo prediktivnu moć modela, odnosno ispitnu pogrešku, a ne pogrešku učenja
- B Omogućava nam da odredimo točnost modela s klasifikacijskim pragom, na način da u obzir uzimamo preciznost i odziv za različite vrijednosti klasifikacijskog praga
- C Provodi optimizaciju hiperparametra modela na uniji skupa za provjeru i skupa za testiranje, čime postiže bolju točnost modela jer više primjera ostaje za treniranje
- D Omogućava nam da procijenimo prediktivnu moć modela optimalne složenosti te maksimalno iskoristimo raspoložive podatke za učenje i ispitivanje

5. (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 35721 B 44640 C 49600 D 69201



$$C : 15 + 1 + 15 = 31$$

$$\gamma : 15 + 1 + 15 = 31$$

$$\text{valid} : \frac{1}{5}$$

$$\begin{array}{c|c} \text{linearna} & \text{RBF} \\ \downarrow & \downarrow \\ \text{svaki } C & C \& \gamma \end{array}$$

$g \times \text{vanjska}$

$$\text{lin} : 31 \rightarrow \text{RBF} : 31 \times 31 = 961 \Rightarrow 992$$

$4 \times \text{vn.}$

1 train

$$g(992 \cdot 4 - 1) = 35721$$

6. (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?

- A $\lambda_1 > \lambda_2 > \lambda_3, F_1^1 < F_1^2, F_1^3 < F_1^2$
- B $\lambda_1 < \lambda_3, F_1^1 < F_1^2 < F_1^3$
- C $\lambda_1 < \lambda_2 = \lambda_3, F_1^1 > F_1^2, F_1^3 > F_1^2$
- D $\lambda_1 = \lambda_3 < \lambda_2, F_1^2 < F_1^1, F_1^3 < F_1^2$

λ : odr. se u unutarnjoj petljiji

\hookrightarrow razlik. samo u 1. $\rightarrow \lambda_1 \neq \lambda_2, \lambda_3 \rightarrow C)$

$$\boxed{\lambda_2 = \lambda_3}$$

uvijek ispitivamo ne 1. preklopu

\rightarrow jednom i treniramo ne mjeru $F_1^1 > F_1^2$

$F_1^3 > F_1^2 \rightarrow$ treniramo na većem stepenu primitiva.

\rightarrow očekujemo bolju

činjenost