16. Bayesov klasifikator II

Strojno učenje 1, UNIZG FER, ak. god. 2023./2024.

Jan Šnajder, vježbe, v3.2

1 Zadatci za učenje

1. [***Svrha: Razumjeti vezu između Bayesovog klasifikatora i logističke regresije, odnosno probabilističku interpretaciju logističke regresije. Razumjeti razliku u broju parametara između diskriminativnog i generativnog modela te utjecaj broja klasa i broja primjera na taj odnos.***]

   (a) Izvedite model logističke regresije krenuvši od generativne definicije za $P(y = 1|\mathbf{x})$. Izvod napravite korak po korak te se uvjerite da možete obrazložiti svaki korak u izvodu. Napišite sve pretpostavke koje ste ugradili u izvod.

   (b) Model logističke regresije koristimo za binarnu klasifikaciju primjera s $n = 100$ značajki. Odredite broj parametara modela logističke regresije te njemu odgovarajućeg generativnog modela.

   (c) Izračunajte broj parametara za isti slučaj, ali sa $K = 5$ klasa.

   (d) Pretpostavite da klasificiramo u $K = 10$ klasa. Izračunajte koliko velika mora biti dimenzija prostora značajki n, a da bi se logistička regresija isplatila jer ima manje parametara od odgovarajućega generativnog modela.

2. [***Svrha: Isprobati na konkretnom primjeru procjenu parametara naivnog Bayesovog klasifikatora.***] Naivan Bayesov klasifikator 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 $(\text{Istra}, \text{ne}, \text{kamp}, \text{bus})$ i $(\text{Dalmacija}, \text{da}, \text{hotel}, \text{bus})$.

   (b) Izračunajte Laplaceove (zaglađene) procjene za sve parametre modela te klasificajte nanovo iste primjere.

3. [***Svrha: Razviti intuiciju o uvjetnoj nezavisnosti i odnosu između nezavisnosti i uvjetne nezavisnosti.***]

   (a) Definirajte uvjetnu nezavisnost slučajnih varijabli. Pokažite da je definicija pomoću zajedničke vjerojatnosti istovjetna definiciji pomoću uvjetne vjerojatnosti.

   (b) Za sljedeće primjere razmotrite sve parove varijabli i odredite za koje parove možemo pretpostaviti nezavisnost odnosno uvjetnu nezavisnost:

      i. $P \equiv$ danas je ponedjeljak, $S =$ danas je subota, $L \equiv$ danas je listopad.

      ii. $S \equiv$ sunčano je; $V =$ vruće je; $K =$ ljudi se kupaju.

iii. $L \equiv$ dokument sadrži riječ "lopta"; $N \equiv$ dokument sadrži riječ "nogomet"; $S \equiv$ dokument je o sportu.

iv. $K \equiv$ pada kiša; $C =$ pukla je cijev; $M \equiv$ ulica je mokra.

(c) Temeljem prethodnih primjera, odgovorite implicira li nezavisnost dviju varijabli njihovu uvjetnu nezavisnost, $A \perp B \Rightarrow A \perp B|C$? Vrijedi li obrnut slučaj, $A \perp B \Rightarrow A \perp B|C$?

4. [***Svrha: Razumjeti definiciju uzajamne informacije i način njezina izračuna. Razumjeti razliku između zavisnosti i lienarne zavisnosti.***]

(a) Krenuvši od definicija za entropiju i relativnu entropiju, izvedite mjeru uzajamne informacije $I(X, Y)$ kao Kullback-Leiblerovu divergenciju između zajedničke razdiobe, $P(X, Y)$, i zajedničke razdiobe uz pretpostavku nezavisnosti, $P(X)P(Y)$.

(b) Neka je zajednička vjerojatnost $P(X, Y)$ varijabli X i Y sljedeća: $P(1, 1) = 0.2$, $P(1, 2) = 0.05$, $P(1, 3) = 0.3$, $P(2, 1) = 0.05$, $P(2, 2) = 0.3$, $P(2, 3) = 0.1$. Izračunajte mjeru uzajamne informacije $I(X, Y)$ za varijable X i Y. Biste li, temeljem vrijednosti uzajamne informacije, rekli da su varijable X i Y nezavisne? Jesu li varijable linearno zavisne?

(c*) Uzajamna informacija nije odozgo ograničena, ali je ograničena odozdo. Primjenom Jensenove nejednakosti, dokažite da vrijedi $I(X, Y) \geqslant 0$.

5. [***Svrha: Shvatiti kako uvjetna nezavisnost varijabli određuje optimalnu strukturu polunaivnog Bayesovog modela te kako to onda određuje broj parametara.***] Želimo naučiti model za klasifikaciju pacijenata s obzirom na rizik oboljenja od kardiovaskularnih bolesti. Ciljne klase su $\mathcal{C}_1 = \mathit{VisokRizik}$, $\mathcal{C}_2 = \mathit{UmjerenRizik}$, $\mathcal{C}_3 = \mathit{NizakRizik}$. Koristimo sedam diskretiziranih ulaznih varijabli: spol, dob, težina, visina, indeks tjelesne mase (BMI), indikacija je li osoba pušač (binarna varijabla) i indikacija bavi li se osoba sportom (binarna varijabla).

(a) Bi li naivan Bayesov model u ovom slučaju bio dobar odabir? Zašto? Predložite polunaivni model.

(b) Izračunajte broj parametara predloženog polunaivnog modela i usporedite ga s brojem parametara naivnog modela.

(c) Razmatramo familiju modela polunaivnog Bayesovog klasifikatora $\mathcal{H}_\alpha$ kod kojeg se združivanje varijabli provodi za sve parove varijabli (x_i, x_j) za koje $I(x_i, x_j) \geqslant \alpha$. Skicirajte pogreške učenja i ispitivanja modela $\mathcal{H}_\alpha$ kao funkcije praga α (dvije krivulje na istoj skici).

2 Zadatci s ispita

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

2. (N) Treniramo naivan Bayesov klasifikator 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

3. (P) Naivan Bayesov klasifikator pretpostavlja uvjetnu nezavisnost značajki unutar 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 $P(\text{stožer}|y = 1) = P(\text{stožer}|\text{pandemija}, y = 1)$

B $P(\text{general}|y = 0) = P(\text{general}|\text{stožer}, y = 0)$

C $P(\text{koronavirus}|y = 0) = P(\text{koronavirus}|\text{general}, y = 0)$

D $P(\text{pandemija}|y = 1) = P(\text{stožer}|y = 1)$

4. (N) Treniramo Bayesov klasifikator za odluku o dobroj destinaciji za Erasmus+ studijski boravak. Skup primjera za učenje, izgrađen na temelju iskustava prijatelja i prijatelja prijatelja, je sljedeći:

i	Država	Stipendija	Semestar	Studij	GovoriJezik	$y^{(i)}$
1	Njemačka	da	ljetni	dipl	da	1
2	Poljska	ne	zimski	preddipl	ne	1
3	Italija	da	ljetni	dipl	da	1
4	Njemačka	ne	zimski	preddipl	ne	0
5	Austrija	da	ljetni	dipl	da	1
6	Poljska	ne	zimski	dipl	ne	1
7	Austrija	da	zimski	dipl	ne	1
8	Njemačka	ne	zimski	dipl	ne	0

Očekujemo zavisnost između varijabli *Država* i *Stipendija*, pa koristimo polunaivan Bayesov klasifikator u kojemu su te dvije varijable združene. Procjene izglednosti klasa radimo Laplaceovim MAP-procjeniteljem. Zanima nas klasifikacija za $\mathbf{x} = (\text{Italija}, \text{ne}, \text{zimski}, \text{dipl}, \text{ne})$. **Koliko iznosi aposteriorna vjerojatnost $P(y = 1|\mathbf{x})$?**

A 0.322 B 0.488 C 0.588 D 0.741

5. (P) 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

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

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

Prije treniranja klasifikatora, koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Odlučili smo združiti onaj par varijabli koje imaju uzajamnu informaciju veću od 0.01. Ako to vrijedi za dva para varijabli, onda ćemo sve tri varijable združiti u jedan faktor. Izračunajte uzajamne informacije između svih parova varijabli te odredite koje varijable ćemo združiti u zajedničke faktore prema gornjem pravilu. **Kako glasi faktorizacija zajedničke vjerojatnosti tog polunaivnog Bayesovog klasifikatora?**

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 $P(y)P(x_1|y)P(x_2|y)P(x_3|y)$

7. (N) Treniramo polunaivan Bayesov klasifikator sa tri binarne značajke, x_1, x_2 i x_3. Skup primjera za učenje $\mathcal{D}$ sastoji se od sljedećih deset primjera:

x_1	x_2	x_3	y	x_1	x_2	x_3	y
1	1	0	1	1	1	0	0
0	1	0	1	1	0	1	1
1	1	0	0	1	0	0	0
0	1	1	1	0	1	1	1
0	1	0	1	0	0	1	0

Prije treniranja koristimo uzajamnu informaciju kako bismo procijenili koje su varijable najviše statistički zavisne, jer se te varijable isplati združiti u zajednički faktor. Izračun provodimo tako da za svaki par varijabli x_i i x_j procjenjujemo parametre zajedničke distribucije $P(x_i, x_j)$, a zatim iz zajedničke distribucije računamo marginalne vjerojatnosti i uzajamnu informaciju $I(x_i, x_j)$. Budući da je skup $\mathcal{D}$ malen, za procjenu parametara distribucije $P(x_i, x_j)$ koristimo Laplaceov procjenitelj. **Koliko iznosi na taj način izračunata uzajamna informacija između varijabli x_1 i x_2?**

A 0.0078 B 0.0112 C 0.0334 D 0.0423