Strojno učenje – domaća zadaća 3

UNIZG FER, ak. god. 2013./2014.

Zadano: 11.11.2013. Rok: 15.11.2013.

Napomena: Zadatke možete rješavati samostalno ili u grupi. Ako zadatke rješavate u grupi, pobrinite se da svi članovi grupe pridonose rješenju i da ga naposlijetku svi razumiju. Po potrebi konzultirajte sve dostupne izvore informacija. Rješenja zadataka ponesite na iduće auditorne vježbe. Zabilježite sve nejasnoće i nedoumice, kako bismo ih prodiskutirali.

1. [***Svrha: Razumjeti model Bayesovog klasifikatora i njegove komponente. Razumjeti kako bi generativni model mogao generirati podatke, kada bismo to htjeli.***]

   (a) Definirajte model Bayesovog klasifkatora i navedite sve veličine koje se pojavljuju u definiciji modela. Objasnite ulogu nazivnika i objasnite kada ga možemo zanemariti.

   (b) Je li taj model parametarski ili neparametarski? Obrazložite odgovor.

   (c) Objasnite zašto Bayesov model nazivamo generativnim i opišite postupak generiranja primjera, odnosno opišite postupak kojim biste mogli generirati primjere koji se pokoravaju zajedničkoj gustoći $P(\mathbf{x}, \mathcal{C}_j)$.

2. [***Svrha: Isprobati izračun maksimalne aposteriorne hipoteze i najvjerojatnije hipoteze uz minimizaciju rizika. Razumjeti kako se neuravnoteženost klasa može kompenzirati pri izračunu izglednosti.***] Razmotrimo problem klasifikaciji neželjene el. pošte u klase *spam* ($\mathcal{C}_1$), *important* ($\mathcal{C}_2$) i *normal* ($\mathcal{C}_3$). Neka su apriorne vjerojatnosti tih klasa $P(\mathcal{C}_1) = 0.2$, $P(\mathcal{C}_2) = 0.05$ i $P(\mathcal{C}_3) = 0.75$. Za neku poruku el. pošte $\mathbf{x}$ izglednosti iznose $p(\mathbf{x}|\mathcal{C}_1) = 0.8$ i $p(\mathbf{x}|\mathcal{C}_2) = p(\mathbf{x}|\mathcal{C}_3) = 0.5$.

   (a) Izračunajte aposteriorne vjerojatnost za svaku od klasa te maksimalnu aposteriornu hipotezu za primjer $\mathbf{x}$.

   (b) Budući da je klasifikacija važne poruke kao spama neprihvatljiva, umjesto matrice gubitka 0-1 treba definirati drugačiju matricu gubitka L. Definirajte tu matricu vodeći se sljedećim načelima: klasifikacija važne poruke kao obične jednako je loša kao klasifikacija spama kao važne poruke, još je gora klasifikacija obične poruke kao spama, ali najgora je klasifikacija važne poruke kao spama. Definirajte matricu gubitka L te klasificirajte primjer $\mathbf{x}$ minimizacijom rizika uz tako definiran gubitak.

   (c) Budući da su važne poruke slabo zastupljene u skupu za učenje, pa je teško naučiti model koji dobro generalizira, odučili smo poduzorkovanjem umjetno uravnotežiti skup za učenje. U novom skupu apriorne vjerojatnosti klasa su $P'(\mathcal{C}_1) = 0.3 = P'(\mathcal{C}_2) = 0.3$ i $P'(\mathcal{C}_3) = 0.4$. Koja je sada klasifikacija primjera $\mathbf{x}$ (uz iste izglednosti), i to u slučaju gubitka 0-1 te u slučaju asimetričnog gubitka definiranog matricom L?

3. [***Svrha: Razumjeti faktorizaciju zajedničke vjerojatnosti uz pretpostavku uvjetne nezavisnosti te povezanost toga s induktivnom pristranošću i, posljedično, brojem značajki modela.***]

    (a) Definirajte naivan Bayesov klasifikator.

    (b) Uz pretpostavku o uvjetnoj nezavisnosti značajki za zadanu klasu, izvedite dekompoziciju zajedničke n-dimenzijske vjerojatnosti na umnožak n faktora.

    (c) Zašto nam treba pretpostavka o uvjetnoj nezavisnosti značajki te kojoj vrsti induktivne pristranosti ona odgovara?

    (d) Naivan Bayesov klasifikator koristimo za klasifikaciju rukom pisanih znamenki u jednu od deset klasa. Znamenke su prikazane kao vektor binarnih značajki (crno/bijeli slikovni elementi) u matrici s razlučivošću 32×32. Odredite ukupan broj parametara naivnog Bayesovog klasifikatora.

4. [***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 ML-procjene svih parametara modela te klasificirajte primjere (Istra, ne, kamp, bus) i (Dalmacija, da, hotel, bus).

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

5. [***Svrha: Razumjeti definiciju uzajamne informacije i način njezina izračuna.***]

    (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) Izračunajte mjeru uzajamne informacije $I(X, Y)$ za varijable X i Y s razdiobom definiranom u zadatku (1a) u zadaći 2. Biste li, temeljem vrijednosti uzajamne informacije, rekli da su varijable X i Y nezavisne?

6. [***Svrha: Isprobati na konkretnom primjeru izgradnju polunaivnog Bayesovog klasifikatora i razumjeti kako modeliranje zavisnosti utječe na broj parametara.***]

    Učimo polunaivan Bayesov klasifikator za klasifikaciju primjera iz $\{0, 1\}^5$ u dvije klase. Na temelju skupa primjera $\mathcal{D}$ dobili smo sljedeće procjene za mjeru uzajamne informacije:

$$I(x_3, \mathcal{C}) > I(x_1, \mathcal{C}) > I(x_2, \mathcal{C}) > I(x_4, \mathcal{C}) > I(x_5, \mathcal{C})$$
$$I(x_1, x_3|\mathcal{C}) > I(x_2, x_4|\mathcal{C}) > I(x_1, x_2|\mathcal{C}) > I(x_3, x_4|\mathcal{C}) > I(x_1, x_4|\mathcal{C}) >$$
$$I(x_3, x_5|\mathcal{C}) > I(x_1, x_5|\mathcal{C}) > I(x_2, x_3|\mathcal{C}) > I(x_2, x_5|\mathcal{C}) > I(x_4, x_5|\mathcal{C})$$

(a) Naučite polunaivan Bayesov klasifikator algoritmom TAN. Skicirajte model kao Bayesovu mrežu.

(b) Naučite polunaivan Bayesov klasifikator algoritmom 2-DB. Skicirajte model kao Bayesovu mrežu.

(c) Izračunajte broj parametara oba modela i usporedite ga s brojem parametara naivnog modela.

7. [***Svrha: Razviti intuiciju za model kontinuiranog bayesovog klasifiaktora.***]

Izrađujemo Bayesov model za klasifikaciju primjera iz $\mathcal{X} = \mathbb{R}$ u tri klase. Učenjem na skupu primjera dobili smo sljedeće parametre modela: $P(\mathcal{C}_1) = 0.3$, $P(\mathcal{C}_2) = 0.2$, $\mu_1 = -5$, $\mu_2 = 0$, $\mu_3 = 5$, $\sigma_1^2 = 5$, $\sigma_2^2 = 1$, $\sigma_3^2 = 10$. Skicirajte funkcije gustoće vjerojatnosti $p(x|\mathcal{C}_j)$, $p(x, \mathcal{C}_j)$, $p(x)$ i $p(\mathcal{C}_j|x)$.

8. [***Svrha: Razumjeti izvod modela kontinuiranog Bayesovog klasifikatora i osvježiti potrebno znanje matematike.***] Izvedite model višedimenzijskog Bayesovog klasifikatora s kontinuiranim ulazima s dijeljenom i dijagonalnom kovarijacijskom matricom. U skripti je taj model izveden iz izraza za $p(\mathbf{x}|\mathcal{C}_j)$ (izraz 3.30); vi izvedite model postepeno, krenuvši od najopćenitijeg modela, preko modela s dijeljenom matricom (izraz 3.29). Ukratko opišite svaki korak izvoda.

9. [***Svrha: Razviti intuiciju za složenost modela kontinuiranog Bayesovog klasifikatora i shvatiti kako se problem u konačnici svodi na odabir optimalnog modela.***] Želimo izgraditi klasifikator za klasifikaciju brucoša u jednu od dvije klase: $\mathcal{C}_1 =$ *"Završava FER u roku"* i $\mathcal{C}_2 =$ *"Produljuje studij"*. Svaki je primjer opisan sa šest ulaznih varijabli: prosjek ocjena 1.–4. razreda (četiri varijable), bodovi državne mature iz matematike te bodovi državne mature iz fizike. Raspolažemo trima modelima: modelom $\mathcal{H}_1$ s dijeljenom kovarijacijskom matricom, modelom $\mathcal{H}_2$ s dijagonalnom (i dijeljenom) kovarijacijskom matricom i modelom $\mathcal{H}_3$ s izotropnom kovarijacijskom matricom.

(a) Koliko svaki od ova tri modela ima parametara?

(b) Za koji od ova tri modela očekujete da će najbolje generalizirati u ovom konkretnom slučaju (uzmite u obzir prirodu problema i očekivane odnose između značajki)? Zašto?

(c) Nacrtajte skicu funkcije empirijske pogreške i pogreške generalizacije i naznačite na njoj točke koje označavaju navedenim trima modelima.

(d) Kako biste u praksi odredili koji ćete model upotrijebiti?

10. [***Svrha: Uočiti vezu između probabilističke interpretacije i interpetacije u smislu funkcije pogreške, odnosno razumjeti dualnost između funkcije gubitka i log-izglednosti.***] Jedna od triju komponenata svakog algoritma strojnog učenja jest funkcija gubitka $L(y, h(\mathbf{x}|\boldsymbol{\theta}))$, temeljem koje je definirana pogreška modela $E(\boldsymbol{\theta}|\mathcal{D})$

na skupu za učenje, a koju algoritam nastoji minimizirati odabirom prikladne hipoteze $h(\cdot|\boldsymbol{\theta})$. Kod probabilističkih modela problem učenja međutim tipično je formuliran ne kao minimizacija funkcije pogreške, već kao maksimizacija log-izglednosti $\ln \mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$ modela na skupu za učenje.

(a) Dovedite u vezu ova dva problema, odnosno izrazite funkciju gubitka Bayesovog klasifikatora preko funkcije izglednosti.

(b) Skicirajte log-izglednost na skupu za učenje i log-izglednost na ispitnome skupu u ovisnosti o složenosti Bayesovog modela.