Strojno učenje – domaća zadaća 2

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

Zadano: 20.10.2013. Rok: 25.10.2013.

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: Prisjetiti se zajedničke, marginalne i uvjetne vjerojatnosti. Prisjetiti se očekivanja, varijacije, kovarijacije i korelacije varijabli.***]

   (a) 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 marginalne vjerojatnosti $P(X)$ i $P(Y)$ te uvjetne vjerojatnosti $P(X|Y)$ i $P(Y|X)$. Uvjerite se da Bayesov teorem daje isti rezultat.

   (b) Za prethodni primjer izračunajte očekivanje $\mathbb{E}[X]$, varijancu $\mathrm{Var}(X)$, kovarijancu $\mathrm{Cov}(X,Y)$, koeficijent korelacije $\rho_{X,Y}$ i kovarijacijsku matricu Σ.

   (c) Dokažite:

      i. $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

      ii. $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$

      iii. $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

2. [***Svrha: Razumjeti nezavisnost slučajnih varijabli i shvatiti da linearna nekoreliranost ne znači nezavisnost.***]

   (a) Definirajte nezavisnost slučajnih varijabli (preko zajedničke vjerojatnosti i preko uvjetne vjerojatnosti).

   (b) Sudeći po iznosu koeficijenta korelacije $\rho_{X,Y}$, jesu li varijable iz zadatka 1 linearno zavisne? Jesu li nezavisne?

   (c) Za koje od sljedećih varijabli očekujete da su zavisne, a za koje da je ta zavisnost linearna: (i) dob i veličina cipela, (ii) dob i sati spavanja, (iii) razina buka i udaljenost od izvora buke, (iv) razina unosa masnoća i rizik od bolesti srca?

   (d) Dokažite da su nezavisne varijable linearno nekorelirane.

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

   (a) Definirajte uvjetnu nezavisnost sl. 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 = danas je ponedjeljak, S = danas je subota, L = danas je listopad.
ii. S = sunčano je; V = vruće je; K = ljudi se kupaju.
iii. L = dokument sadrži riječ "lopta"; N = dokument sadrži riječ "nogomet"; S = dokument je o sportu.
iv. K = pada kiša; C = pukla je cijev; M = ulica je mokra.

(c) Razmotrite ponovo posljednji primjer. Neka je $P(K) = 0.2$, $P(C) = 0.01$ te neka $K \perp C$. Definirajte zajedničku vjerojatnost $P(K, C, M)$ te pokažite da $P(K|C, M) \leqslant P(K|M)$. Ovaj se fenomen zove *explaining away* (činjenica da je cijev pukla objašnjava mokru ulicu i smanjuje vjerojatnost da padanje kiše objašnjava mokru ulicu.)

(d) 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 kako podatci određuju izglednost parametara posredstvom funkcije izglednosti.***]

(a) Definirajte funkciju izglednosti $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$. Na kojoj se pretpostavci o skupu $\mathcal{D}$ temelji ta definicija?

(b) Raspolažemo skupom (neoznačenih) primjera $\mathcal{D} = \{x^{(i)}\}_i = \{-2, -1, 1, 3, 5, 7\}$. Pretpostavljamo da se primjeri pokoravaju Gaussovoj distribuciji, $x^{(i)} \sim \mathcal{N}(\mu, \sigma^2)$. Napišite funkciju izglednosti $\mathcal{L}(\mu, \sigma^2|\mathcal{D})$. Koliko iznosi izglednost parametara $\mu = 0$ i $\sigma^2 = 1$, a koliko vjerojatnost uzorka $\mathcal{D}$ uz te parametre?

(c) Novčić bacamo N puta, pri čemu smo m puta dobili glavu, a $N - m$ puta pismo. Ishodi bacanja novčića sačinavaju naš uzorak $\mathcal{D}$. Napišite izraz za funkciju izglednosti parametriziranu s N i m, tj. $\mathcal{L}(\mu|N, m)$.

(d) Skicirajte funkciju izglednosti za slučaj $N = 10$ i $m = 1$. Koja je vrijednost parametra μ najizglednija?

5. [***Svrha:*** ***Osvježiti znanje matematike potrebno za izvođenje ML-procjenitelja dviju osnovnih univarijatnih razdioba.***]

(a) Definirajte ML-procjenitelj $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$.

(b) Izvedite ML-procjenitelj $\hat{\mu}_{\mathrm{ML}}$ za parametar μ Bernoullijeve razdiobe $p(x|\mu)$.

(c) Izvedite ML-procjenitelje $\hat{\mu}_{\mathrm{ML}}$ i $\hat{\sigma}^2$ za parametre μ odnosno σ^2 univarijatne Gaussove razdiobe $p(x|\mu, \sigma^2)$.

6. [***Svrha:*** ***Isprobati izračun pristranost procjenitelja i shvatiti da ML-procjenitelj može biti pristran, tj. da najveća izglednost ne jamči nepristranost.***]

(a) Dokažite da je $\hat{\mu}_{\mathrm{ML}}$ nepristran, a $\hat{\sigma}^2_{\mathrm{ML}}$ pristran. Koliko iznosi pristranost $b(\hat{\sigma}^2)$?

(b) Što mislite: je li ta pristranost u praksi problematična? Obrazložite.

7. [***Svrha:*** ***Izvježbati izračun procjene parametara multivarijatne Gaussove razdiobe (v. zadatak 3.5 u skripti). Uočiti da koreliranost varijabli dovodi do problema.***] Raspolažemo uzorkom $\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{6}$ za koji znamo da potječe iz multivarijatne normalne razdiobe:

$$
\begin{aligned}
\mathbf{x}^{(1)} &= (9.59, -0.75, 2.88) & \mathbf{x}^{(4)} &= (2.24, \quad 0.02, 0.67)\\
\mathbf{x}^{(2)} &= (2.30, \quad 0.37, 0.69) & \mathbf{x}^{(5)} &= (6.59, -0.20, 1.98)\\
\mathbf{x}^{(3)} &= (8.87, -0.84, 2.66) & \mathbf{x}^{(6)} &= (3.69, \quad 0.35, 1.11)
\end{aligned}
$$

(a) Izračunajte ML-procjenu vektora srednje vrijednosti i ML-procjenu kovarijacijske matrice.

(b) Napišite formulu za multivarijatnu Gaussovu gustoću $p(\mathbf{x}|\boldsymbol{\mu}, \Sigma)$ (raspišite formulu do kraja). Je li ta funkcija dobro definirana? Zašto?

(c) Matrica kovarijacije Σ mora biti pozitivno definitna da bi imala pozitivnu determinantu i inverz. Koreliranost varijabli jedan je od mogućih razloga zašto matrica nije pozitivno definitna. Izračunate Pearsonov koeficijent korelacije između varijabli, izbacite varijablu koja je previše korelirana s nekom drugom varijablom te pokušajte ponovno definirati funkciju gustoće.

8. [***Svrha: Razumjeti MAP-procjenitelj i način njegovog izračuna za Bernoullijevu varijablu. Uočiti kako svojstvo konjugatnosti olakšava izračun aposteriorne distribucije.***]

(a) Definirajte MAP-procjenitelj $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ i objasnite zašto je on bolji od $\hat{\boldsymbol{\theta}}_{\text{ML}}$.

(b) Apriornu distribuciju parametra μ Bernullijeve varijable modeliramo beta-distribucijom $p(\mu|\alpha, \beta)$. Beta-distribucija konjugatna je Bernoullijevoj distribuciji $p(x|\mu)$, pa će umnožak izglednosti $\mathcal{L}(\mu|N, m)$ (zadatak 4(c)) i beta-distribucije $p(\mu|\alpha, \beta)$ opet biti neka beta-distribucija $p(\mu|\alpha', \beta')$, što pojednostavljuje izračun. Izračunajte taj umnožak i odredite parametre α' i β' aposteriorne beta-distribucije.

(c) Recimo da vjerujemo da je novčić pravedan, ali da u to nismo baš jako uvjereni. To možemo modelirati beta-distribucijom $p(\mu|\alpha = 2, \beta = 2)$. Zatim smo u $N = 10$ bacanja novčića samo $m = 1$ puta dobili glavu. Skicirajte apriornu gustoću $p(\mu|\alpha = 2, \beta = 2)$, funkciju izglednosti $\mathcal{L}(\mu|N = 10, m = 1)$ te njihov umnožak. Iskoristite činjenicu da je maksimum (mod) beta-distribucije jednak $\frac{\alpha-1}{\alpha+\beta-2}$.

(d) Izračunajte $\hat{\mu}_{\text{MAP}}$ i $\hat{\mu}_{\text{ML}}$ te komentirajte razliku. Kako bi porast broja primjera N utjecao na ovu razliku?

9. [***Svrha: Razumjeti bayesovski procjenitelj za Bernoullijevu varijablu. Shvatiti da je Laplaceov procjenitelj poseban slučaj bayesovskog procjenitelja.***]

(a) Definirajte bayesovski procjenitelj $\hat{\boldsymbol{\theta}}_{\text{Bayes}}$ i objasnite zašto je on bolji od $\hat{\boldsymbol{\theta}}_{\text{MAP}}$.

(b) Izračunajte $\hat{\boldsymbol{\theta}}_{\text{Bayes}}$ za slučaj iz prethodnog zadatka. Iskoristite činjenicu da je očekivanje beta-distribucije jednako $\frac{\alpha}{\alpha+\beta}$. Je li procjena drugačija od MAP-procjene? Zašto?

(c) Ako nemamo nikakvih spoznaja o apriornoj razdiobi parametra, možemo je modelirati uniformnom beta-razdiobom s parametrima $\alpha = \beta = 1$. Takvu apriornu razdiobu parametara zovemo neinformativnom ili slabo informativnom (engl. *uninformative, weakly informative prior*). Bayesov procjenitelj za Bernoullijevu varijablu (također i za multinomijalnu varijablu) s uniformnom apriornom razdiobom zovemo Laplaceovim procjeniteljem (ili *add-one rule*). Izvedite Laplaceov procjenitelj Bernoullijeve varijable za skup od N primjera od kojih je m pozitivno.

(d) Izračunajte Laplaceov procjenitelj za slučaj $N = 10$ bacanja novčića i $m = 0$, $m = 1$, $m = 9$ odnosno $m = 10$ glava. Usporedite procjenu s ML-procjenom.