Strojno učenje – međuispit

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

21. studenog 2013.

Napomena: Ispit traje 180 minuta i nosi 35 bodova. Svaki zadatak rješavajte na zasebnoj stranici.

1. (*3 boda*) Osnovni koncepti nadziranog učenja.

   (a) Nabrojite (bez opisivanja) tri osnovne komponente algoritma nadziranog strojnog učenja. Koja od tih komponenti uvodi koju vrstu induktivne pristranosti?

   (b) Obrazložite u kakvim su odnosima (i) složenost modela, (ii) broj parametara modela i (iii) broj primjera za učenje.

2. (*5 bodova*) Prostor inačica i VC-dimenzija.

   (a) Model $h(\mathbf{x}|\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2+(x_2-\theta_2)^2 \leqslant \theta_0\}$, gdje $\boldsymbol{\theta} \in \mathbb{R}^3$, koristimo za klasifikaciju u prostoru $\mathcal{X} = \mathbb{Z}^2$. Raspolažemo skupom primjera: $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \big\{((0,0),0),((2,3),0),((1,2),1),((2,2),1)\big\}$. Odredite klasifikaciju primjera $\mathbf{x} = (1,4)$ te odredite skup prostora inačica $VS_{\mathcal{H},\mathcal{D}}$.

   (b) Formalno definirajte VC-dimenziju modela $\mathcal{H}$ i objasnite zašto je ta karakterizacija složenosti modela u praksi problematična.

   (c) Odredite VC-dimenziju modela pod (a).

   (d) Redefinirajte model pod (a) tako da (i) prostor inačica sadrži samo jednu hipotezu (model $\mathcal{H}_1$) i (ii) povećate VC-dimenziju modela (model $\mathcal{H}_2$).

3. (*4 boda)* Odabir modela.

   (a) Definirajte prenaučenost modela i u jednoj rečenici objasnite zašto ne želimo prenaučene modele.

   (b) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Odabrali smo dva modela, $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$, oba trenirana istim algoritmom L. Pogreške na skupu za učenje su takve da $E(h_2|\mathcal{D}) < E(h_1|\mathcal{D})$. Koji biste model odabrali i zašto?

   (c) Raspolažemo dvama algoritmima strojnog učenja, L_1 i L_2. Algoritam L_1 analitički optimizira parametre modela, dok algoritam L_2 parametre optimizira heuristički i sklon je zaglavljivanju u lokalnom optimimu tako da u više od 50% slučajeva nalazi suboptimalne parametre. Skicirajte funkciju empirijske pogreške i pogrešku generalizacije u ovisnosti o složenosti modela za algoritme L_1 i L_2.

4. (*5 bodova*) Procjenitelji.

   (a) Definirajte funkciju log-izglednosti $\ln\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$ i izračunajte $\ln\mathcal{L}(\mu = 0, \sigma^2 = 1|\mathcal{D})$ za skup primjera $\mathcal{D} = \{0,1,2\}$ uz pretpostavku Gaussove razdiobe $p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\}$.

   (b) Definirajte procjenitelje $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ i $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}$ te ukratko objasnite prednost drugog nad prvim.

   (c) Izvedite, korak po korak, ML-procjenitelj za parametar μ univarijatne Gaussove razdiobe.

   (d) U deset bacanja novčića ($N = 10$) dva puta smo dobili glavu ($m = 2$). Apriornu razdiobu parametra μ Bernoullijeve razdiobe modeliramo beta-distribucijom $p(\mu|\alpha,\beta) = \mu^{\alpha-1}(1-\mu)^{\beta-1}/B(\alpha,\beta)$ uz $\alpha = \beta = 2$. Skicirajte funkciju izglednosti, gustoće $p(\mu)$ i $p(\mu|\mathcal{D})$ te odredite $\hat{\mu}_{\mathrm{MAP}}$ i $\hat{\mu}_{\mathrm{Bayes}}$. Iskoristite činjenicu da je mod beta-distribucije jednak $\frac{\alpha-1}{\alpha+\beta-2}$ a očekivanje $\frac{\alpha}{\alpha+\beta}$.

   (e) Iz bayesovskog procjenitelja izvedite Laplaceov procjenitelj za Bernoullijevu varijablu.

5. (*5 boda*) Bayesov klasifikator za diskretne ulaze.

   (a) Definirajte naivan Bayesov klasifikator i napišite općenit izraz za broj parametara tog modela.

   (b) Poruke el. pošte klasificiramo u dvije klase: *spam* ($\mathcal{C}_1$) i *no-spam* ($\mathcal{C}_2$). Pogrešna klasifikacija *no-spam* poruke kao *spam* je deset puna nepoželjnija od ostalih pogrešaka. Definirajte matricu gubitka L i izračunajte klasifikaciju poruke $\mathbf{x}$ ako $P(\mathbf{x}|\mathcal{C}_1) = 0.7$, $P(\mathbf{x}|\mathcal{C}_2) = 0.1$ i $P(\mathcal{C}_1) = 0.8$.

   (c) 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:

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

   Naučite polunaivan Bayesov klasifikator algoritmom 3-DB, skicirajte model kao Bayesovu mrežu te napišite izraz za zajedničku vjerojatnost.

   (d) Skicirajte funkciju log-izglednosti na skupu za učenje i skupu za provjeru u ovisnosti o broju k za model k-DB.

6. (*4 boda*) Bayesov klasifikator za kontinuirane ulaze.

   (a) Izgrađujemo Bayesov model za klasifikaciju primjera iz $\mathcal{X} = \mathbb{R}$ u tri klase. Učenjem na skupu primjera dobili smo sljedeće parametre modela: $P(\mathcal{C}_1) = 0.7$, $P(\mathcal{C}_2) = 0.2$, $\mu_1 = -2$, $\mu_2 = 0$, $\mu_3 = 1$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $\sigma_3^2 = 5$. Skicirajte funkcije gustoće vjerojatnosti $p(x|\mathcal{C}_j)$, $p(x)$ i $p(\mathcal{C}_j|x)$.

   (b) Kod multivarijatnog Bayesovog klasifikatora, izglednosti klasa definirane su gustoćom:

   $$p(\mathbf{x}|\mathcal{C}_j) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_j|^{1/2}} \exp\Big\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\Big\}.$$

   Izvedite općeniti izraz za model $h(\mathbf{x}|\boldsymbol{\theta})$ s dijeljenom kovarijacijskom matricom.

   (c) Multivarijatni Bayesov klasifikator koristimo za predikciju mjesta prvog zapošljavanja studenata FER-a. Ulazne varijable su ocjene na dvadeset odabranih predmeta te duljina studija. Ciljne klase su: *industrija-hrv*, *industrija-ino*, *phd-hrv*, *phd-ino*. Koliki je ukupan broj parametara ovog modela uz pretpostavku (i) dijeljene kovarijacijske matrice te (ii) dijeljene i dijagonalne kovarijacijske matrice.

7. (*4 boda*) Linearan model regresije.

   (a) Empirijska pogreška linearnog modela regresije jest $E(\mathbf{w}|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^{N}\big(y^{(i)} - \boldsymbol{\phi}(\mathbf{x}^{(i)})^{\mathrm{T}}\mathbf{w}\big)^2$. Pokažite da je minimizacija ovog izraza istovjetna maksimizaciji log-izglednosti $\ln\mathcal{L}(\mathbf{w}|\mathcal{D})$ (odnosno minimizaciji negativne log-izglednosti) uz pretpostavku normalno distribuiranog šuma $\mathcal{N}(h(\mathbf{x}|\mathbf{w}),\sigma^2)$.

   (b) Rješenje za vrijednosti težina koje minimizira regularizirani pogrešku najmanjih kvadrata jest $\mathbf{w} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \lambda\mathbf{I})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{y}$. Treniramo regresijski model za funkciju $\mathbb{R}^2 \to \mathbb{R}$. Raspolažemo primjerima $\mathcal{D} = \big\{(\mathbf{x}^{(i)}, y^{(i)})\big\} = \big\{((0,1),4),((1,2),1),((2,3),2),((4,1),5)\big\}$. Napišite izraz za $\mathbf{w}$ i odredite dimenziju matrice koju treba invertirati ako koristimo funkciju preslikavanja $\boldsymbol{\phi}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$.

8. (*5 bodova*) Regularizacija.

   (a) Napišite općenit izraz za L2-regulariziranu funkciju pogreške.

   (b) Regresijskim modelom aproksimiramo funkciju koja se ponaša slično jednoj periodi sinusne funkcije (no to nam nije unaprijed poznato). Raspolažemo s malo podataka koji sadrže razmjerno mnogo šuma. Neka je $\mathcal{H}_{d,\lambda}$ model polinoma d-tog stupnja s regularizacijskim faktorom λ. Iskušavamo četiri modela: $\mathcal{H}_{2,0}$, $\mathcal{H}_{2,100}$, $\mathcal{H}_{5,0}$, $\mathcal{H}_{5,100}$. Na dva odvojena crteža skicirajte (i) funkciju $h(x)$ za sva četiri modela te (ii) empirijsku pogrešku i pogrešku generalizacije za sva četiri modela.

   (c) Na jednom crtežu skicirajte empirijsku pogrešku te *regulariziranu* funkciju pogreške za model $\mathcal{H}_{d,\lambda}$ za $\lambda = 0$, $\lambda = 100$ i $\lambda = 1000$.

   (d) Na jednom crtežu skicirajte empirijsku pogrešku te *regulariziranu* funkciju pogreške za model $\mathcal{H}_{d,100}$ za slučajeve kada ima mnogo i kada ima malo šuma u podatcima.

   (e) Odgovorite u jednoj rečenici: koja je probabilistička interpretacija L2-regularizacije?