2. Osnovni koncepti

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

Jan Šnajder, vježbe, v3.1

1 Zadatci za učenje

1. [***Svrha: Na stvarim problemima razlikovati klasifikaciju od regresije.***] Objasnite razliku između klasifikacije i regresije. Koji je od ta dva pristupa prikladan za: (a) filtriranje neželjene e-pošte (*spam*), (b) predviđanje kretanja dionica, (c) rangiranje rezultata tražilice? Kako biste u ovim slučajevima definirali ciljne oznake y?

2. [***Svrha: Razumjeti što je hipoteza, što je model i koja je veza između njih.***]

   (a) Dopunite praznine:

   Hipoteza je funkcija koja preslikava ________ u ________, definirana do na ________. Model je ________ hipoteza, koje su indeksirane ________. Tako parametrizirani skup hipoteza također možemo prikazati kao prostor ________, a dimenzija tog prostora jednaka je ________. Učenje modela odgovara pretraživanju ________ u potrazi za ________ hipotezom. To je ona hipoteza koja ________ klasificira označene primjere, što procjenjujemo pomoću ________ mjerene na ________. Drugim riječima, učenje modela svodi se na ________ parametara modela s ________ kao kriterijskom funkcijom.

   (b) Rješavamo problem binarne klasifikacije u prostoru primjera $\mathcal{X} = \{0,1\}^2$. Definirajte linearan model koji će primjere odvajati pravcem.

   (c) Koja je dimenzija prostora parametra? Koliko različitih hipoteza postoji u $\mathcal{H}$?

   (d) Neka je skup označenih primjera sljedeći:

   $$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \big\{((0,0),0),((1,1),0),((1,0),1),((0,1),1)\big\}.$$

   Odredite konkretnu hipotezu $h \in \mathcal{H}$ koja ima najmanju empirijsku pogrešku.

3. [***Svrha: Shvatiti što je to induktivna pristranost i kako ona određuje klasifikaciju neviđenih primjera.***] Pročitajte poglavlje 2.3 u skripti (tu temu nismo obradili na predavanju).

   (a) Definirajte induktivnu pristranost (neformalno i formalno). Koje su dvije vrste pristranosti koje sačinjavaju induktivnu pristranost?

   (b) Raspolažemo skupom označenih primjera u ulaznome prostoru $\mathcal{X} = \{0,1\}^3$:

   $$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \big\{((0,0,0),0),((1,0,0),1),((1,0,1),1),((0,1,0),1),((0,1,1),1)\big\}.$$

   Koja je klasifikacija neviđenih primjera?

   (c) Definirajte linearan model $\mathcal{H}$ za $\mathcal{X} = \{0,1\}^3$. Koja je to vrsta pristranosti?

   (d) Možete li odrediti klasifikaciju neviđenih primjera uz odabrani model $\mathcal{H}$? Je li pristranost koja proizlazi iz odabira modela dovoljna za jednoznačnu klasifikaciju svih primjera iz $\mathcal{X}$?

   (e) Definirajte (neformalno) neku dodatnu pristranost takvu da klasifikacija svakog primjera slijedi jednoznačno na temelju skupa primjera $\mathcal{D}$. Koje je vrste ta dodatna pristranost?

4. [***Svrha: Znati nabrojati osnovne komponente algoritma strojnog učenja i povezati ih s induktivnom pristranošću.***]

   (a) Nabrojite tri osnovne komponente algoritma strojnog učenja.

(b) Identificirajte uz koje se komponente veže koja vrsta induktivne pristranosti.

5. [***Svrha: Razumjeti vezu između funkcije gubitka i empirijske pogreške te mogućnost njihove prilagodbe konkretnom problemu.***]

   (a) Pogreška hipoteze je očekivanje funkcije gubitka L. Nad kojom distribucijom je definirano to očekivanje? Koji je problem s takvom definicijom u praksi?

   (b) Definirajte *empirijsku* pogrešku preko funkcije gubitka L. Koja je pretpostavka implicitno ugrađena u tu definiciju?

   (c) Kod asimetričnih gubitaka funkciju L možemo definirati preko matrice gubitka (v. skriptu: poglavlje 2.7 i primjer 2.6). Definirajte takvu matricu za problem klasifikacije neželjene e-pošte te izračunajte funkciju pogreške za slučaj pet pogrešno negativnih i dvije pogrešno pozitivne klasifikacije od ukupno deset ($N = 10$) primjera.

6. [***Svrha: Razviti ispravnu intuiciju za odabir modela temeljem unakrsne provjere.***]

   (a) Skicirajte krivulje pogreške učenje i ispitne pogreške u ovisnosti o složenosti modela. Naznačite područje prenaučenosti i podnaučenosti.

   (b) Objasnite zašto pogreška učenja s povećanjem složenosti modela teži k nuli.

   (c) Raspolažemo modelom $\mathcal{H}_\alpha$ koji ima hiperparametar α kojim se može ugađati složenost modela. Za odabrani α naučili smo hipotezu koja minimizira empirijsku pogrešku. Unakrsnom provjerom utvrdili smo da je ispitna pogreška znatno veća od pogreške učenja. Je li naš odabir hiperparametra α suboptimalan?

   (d) Raspolažemo modelom $\mathcal{H}_\alpha$ s hiperparametrom α (veći α daje složeniji model). Raspolažemo dvama optimizacijskim algoritmima: L_1 i L_2. Algoritam L_2 lošiji je od algoritma L_1, u smislu da L_2 pronalazi parametre $\boldsymbol{\theta}_2$ koji su lošiji od parametara $\boldsymbol{\theta}_1$ koje pronalazi L_1, tj. $E(\boldsymbol{\theta}_2|\mathcal{D}) > E(\boldsymbol{\theta}_1|\mathcal{D})$. Neka α_1^* označava optimalnu vrijednost hiperparametra za $\mathcal{H}_\alpha$ učenog algoritmom L_1, a α_2^* optimalnu vrijednost za $\mathcal{H}_\alpha$ učenog algoritmom L_2. Načinite skicu analognu onoj iz zadatka (a) i naznačite vrijednosti pogrešaka za modele $\mathcal{H}_{\alpha_1^*}$ i $\mathcal{H}_{\alpha_2^*}$.

   (e) Može li model učen lošijim algoritmom L_2 imati manju ispitnu pogrešku od modela koji je učen boljim algoritmom L_1, ali nije optimalan? Skicirajte takvu situaciju na prethodnoj skici.

2 Zadatci s ispita

1. (P) U ulaznom prostoru $\mathcal{X} = \{0,1\}^3$ definiramo sljedeći klasifikacijski model:

$$h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \geqslant 0\big\}$$

   Koja je dimenzija prostora parametara te koliko različitih hipoteza postoji u ovom modelu?

   [A] Dimenzija prostora parametara je 4, a hipoteza ima beskonačno mnogo

   [B] Dimenzija prostora parametara je 4, a hipoteza ima manje od 256

   [C] Dimenzija prostora parametara i broj hipoteza su beskonačni

   [D] Dimenzija prostora parametara je 256, a hipoteza ima 14

2. (P) Za ulazni prostor $\mathcal{X} = \{0,1\}^3$ definiramo klasifikacijski model $\mathcal{H}$ kao skup parametriziranih funkcija definiranih na sljedeći način:

$$h(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\big\{(\theta_{1,1} \leqslant x_1 \leqslant \theta_{1,2}) \wedge (\theta_{2,1} \leqslant x_2 \leqslant \theta_{2,2}) \wedge (\theta_{3,1} \leqslant x_3 \leqslant \theta_{3,2})\big\}$$

   Parametri su trodimenzijski vektori realnih brojeva, tj. prostor parametara definiran je kao $\boldsymbol{\theta} \in \mathbb{R}^6$. **Koliko iznosi $|\mathcal{H}|$?**

   [A] 42   [B] ∞ [C] 56   [D] 28

3. (P) Skup označenih primjera u dvodimenzijskome ulaznom prostoru je:

$$\mathcal{D} = \{((0,0),0),((0,1),0),((1,1),1)\}$$

Koliko hipoteza ostvaruje empirijsku pogrešku jednaku nuli?

[A] 16 [B] Pitanje nema smisla jer nije definiran model [C] Beskonačno mnogo [D] 14

4. (P) Za linearan klasifikator u $\mathcal{X} = \{0,1\}^3$ zadan je sljedeći skup primjera za učenje:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \big\{((0,0,0),0),((1,0,0),1),((1,0,1),1),((0,1,0),1),((0,1,1),1),((1,1,0),0)\big\}$$

Razmatramo dva modela:

$$\begin{aligned}\mathcal{H}_a :& h_a(\mathbf{x}|\boldsymbol{\theta}) = \mathbf{1}\{\theta_0 + x_1\theta_1 + x_2\theta_2 + x_3\theta_3 \geqslant 0\}\\ \mathcal{H}_b :& h_b(\mathbf{x}|\boldsymbol{\theta}) = h_a(\mathbf{x};\boldsymbol{\theta}_1)\cdot h_a(\mathbf{x};\boldsymbol{\theta}_2)\end{aligned}$$

Uočite da svaka hipoteza iz modela $\mathcal{H}_b$ kombinira dvije hipoteze iz modela $\mathcal{H}_a$ (operacijom množenja). Neka:

$$\begin{aligned}h_a^* &= \text{argmin}_{h\in\mathcal{H}_a} E(h|\mathcal{D})\\ h_b^* &= \text{argmin}_{h\in\mathcal{H}_b} E(h|\mathcal{D})\end{aligned}$$

Koja je od navedenih tvrdnji točna?

[A] $E(h_a^*|\mathcal{D}) = E(h_b^*|\mathcal{D}) > 0$

[B] $E(h_a^*|\mathcal{D}) > E(h_b^*|\mathcal{D}) = 0$

[C] $0 < (E(h_a^*|\mathcal{D}) < E(h_b^*|\mathcal{D}) < 1$

[D] $E(h_a^*|\mathcal{D}) = E(h_b^*|\mathcal{D}) = 0$

5. (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \{0,1\}^2$. Razmatramo sljedeće modele:

$$\mathcal{H}_1 : h_1(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\} \qquad \mathcal{H}_3 : h_3(\mathbf{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = h_1(\mathbf{x};\boldsymbol{\theta}_1) \wedge h_2(\mathbf{x};\boldsymbol{\theta}_2)$$

$$\mathcal{H}_2 : h_2(\mathbf{x};\boldsymbol{\theta}) = \mathbf{1}\{(x_1-\theta_1)^2 + (x_2-\theta_2)^2 \leqslant \theta_0^2\} \qquad \mathcal{H}_4 = \mathcal{H}_1 \cup \mathcal{H}_2$$

Parametri svih modela realni su brojevi, $\boldsymbol{\theta} \in \mathbb{R}^3$. **Koji odnosi vrijede između ovih modela?**

[A] $\mathcal{H}_1 = \mathcal{H}_2 \subset \mathcal{H}_3 = \mathcal{H}_4$

[B] $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_4 \subset \mathcal{H}_3$

[C] $\mathcal{H}_1 \subset \mathcal{H}_2 \subset \mathcal{H}_3 \subset \mathcal{H}_4$

[D] $\mathcal{H}_1 \subset \mathcal{H}_2 = \mathcal{H}_3 \subset \mathcal{H}_4$

6. (P) Za linearan model u $\mathcal{X} = \{0,1\}^3$ zadan je sljedeći skup primjera za učenje:

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \big\{((0,0,0),0),((1,0,0),1),((1,0,1),1),((0,1,0),1),((0,1,1),1)\big\}$$

Optimizacijski postupak klasifikatora funkcionira tako da minimizira empirijsku pogrešku, definiranu kao očekivanje funkcije gubitka 0-1, i postupak u tome uvijek uspijeva. Želimo znati koju bi klasu ovaj klasifikator dodijelio primjeru $\mathbf{x} = (1,1,1)$. **Možemo li, na temelju iznesenih informacija, odrediti klasifikaciju dotičnog primjera i što nam to govori o induktivnoj pristranosti ovog algoritma**?

[A] Ne možemo, jer nije definirana induktivna pristranost preferencijom, pa činjenica da je model linearan nije dovoljan skup pretpostavki da bismo jednoznačno odredili klasifikaciju svih novih primjera

[B] Možemo, klasifikacija je $y = 1$, i ovaj klasifikator ima definiranu induktivnu pristranost pomoću koje može jednoznačno odrediti klasifikaciju svakog primjera

[C] Možemo, klasifikacija je $y = 1$, premda dane informacije nisu dovoljne za definiciju induktivne pristranosti, pa za ovaj skup primjera više hipoteza savršeno točno klasificira primjere

[D] Možemo, $y = 1$, jer klasifikator ima induktivnu pristranost jezikom (linearan model) i preferencijom (primjeri za koje je $h(x) \geqslant 0$ klasificiraju se pozitivno)

7. (P) Optimizacija parametara modela temelji se na funkciji gubitka $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$, gdje je $L(y, h(\mathbf{x}))$ gubitak na primjeru $(\mathbf{x}, y)$. U većini primjena koristimo simetričan gubitak 0-1. Međutim, u nekim primjenama ima više smisla definirati asimetričan gubitak. Jedan takav primjer je zadatak detekcije karcinoma iz medicinskih slika. Taj zadatak možemo formalizirati kao problem binarne klasifikacije s oznakama $\mathcal{Y} = \{0, 1\}$, gdje $y = 1$ označava postojanje karcinoma, a $y = 0$ nepostojanje karcinoma. **Koje od sljedećih svojstava bi trebala zadovoljiti asimetrična funkcija gubitka za takav zadatak?**

   A $L(0,1) = 1$ i $L(1,0) = L(1,1) = L(0,0) = 0$

   B $L(0,1) > L(1,0)$ i $L(1,1) = L(0,0) > 0$

   C $L(1,0) > L(0,1)$ i $L(1,1) = L(0,0) = 0$

   D $L(0,1) = L(1,0) > 0$ i $L(1,1) = L(0,0) = 0$

8. (P) Zadan je sljedeći skup sa $N = 6$ označenih primjera iz $\mathbb{R}^3$:

$$\begin{aligned}\mathcal{D} &= \{(\mathbf{x}^{(i)}, y^{(i)})\}\\ &= \big\{((0,0,0),0), ((1,1,0),0), ((1,0,0),1), ((1,0,1),1), ((0,1,0),1), ((0,1,1),1)\big\}\end{aligned}$$

   Razmatramo linearan model i računamo empirijsku pogrešku $E(h|\mathcal{D})$ hipoteza iz tog modela definiranu kao očekivanje asimetričnog gubitka. Gubitak je definiran tako da lažno negativne primjere kažnjava sa 1, a lažno pozitivne primjere sa 0.5. **Koliko iznosi najmanja, a koliko najveća moguća vrijednost tako definirane empirijske pogreške $E(h|\mathcal{D})$?**

   A $0 \leqslant E(h|\mathcal{D}) \leqslant 1/4$

   B $1/4 \leqslant E(h|\mathcal{D}) \leqslant 2/3$

   C $\frac{1}{48} \leqslant E(h|\mathcal{D}) \leqslant 2/3$

   D $1/12 \leqslant E(h|\mathcal{D}) \leqslant 3/4$

9. (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \big\{((0,0),0), ((0,2),0), ((0,-1),0), ((-1,0),1), ((0,1),1), ((1,0),1)\big\}$. Razmatramo sljedeće modele, parametrizirane sa $\boldsymbol{\theta} \in \mathbb{R}^{n+1}$:

$$\begin{aligned}&\mathcal{H}_1 : h_1(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}\\ &\mathcal{H}_2 : h_2(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{(x_1 - \theta_1)^2 + (x_2 - \theta_2)^2 \geqslant \theta_0^2\}\end{aligned}$$

   Pored ova dva modela, razmatramo i njihove kombinacije, modele $\mathcal{H}_3$ i $\mathcal{H}_4$. Neka je $\mathcal{H}_3 = \mathcal{H}_1 \cup \mathcal{H}_2$ te neka je $\mathcal{H}_4$ skup funkcija definiranih kao $h_4(\mathbf{x}; \boldsymbol{\theta}) = h_1(\mathbf{x}) \cdot h_2(\mathbf{x})$. Neka je E_k minimalna empirijska pogreška koja se modelom $\mathcal{H}_k$ može ostvariti na skupu $\mathcal{D}$, tj. $E_k = \mathrm{argmin}_{h \in \mathcal{H}_k} E(h|\mathcal{D})$. **Koji odnosi vrijede između minimalnih empirijskih pogrešaka ovih modela?**

   A $E_1 > E_2 = E_3 > E_4$

   B $E_1 = E_2 > E_3 = E_4$

   C $E_1 > E_2 > E_3 = E_4$

   D $E_1 = E_2 = E_3 > E_4$

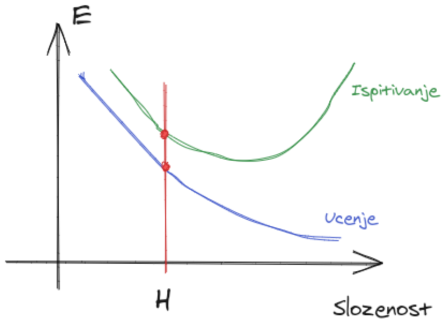
10. (P) Razmatramo klasifikacijski problem u ulaznome prostoru $\mathcal{X} = \mathbb{Z}^2$. Skup označenih primjera je $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\} = \big\{((0,0),1), ((-1,-1),0), ((1,1),0)\big\}$. Razmatramo sljedeće modele $\mathcal{H}$ i funkcije preslikavanja $\boldsymbol{\phi} : \mathbb{Z}^2 \to \mathbb{Z}^2$, kojom primjere iz $\mathcal{D}$ preslikavamo u matricu dizajna $\boldsymbol{\Phi}$:

$$\begin{aligned}&\mathcal{H}_1 : h_1(\mathbf{x}; \theta_0, \theta_1) = \mathbf{1}\{\theta_1 x_1 + \theta_0 \geqslant 0\} && \boldsymbol{\phi}_1(\mathbf{x}) = (1, x_2, x_1)\\ &\mathcal{H}_2 : h_2(\mathbf{x}; \theta_0, \theta_2) = \mathbf{1}\{\theta_2 x_2 + \theta_0 \geqslant 0\} && \boldsymbol{\phi}_2(\mathbf{x}) = (1, x_1, x_1 x_2)\\ &\mathcal{H}_3 : h_3(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{1}\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x} \geqslant 0\}\\ &\mathcal{H}_4 : h_4(\mathbf{x}; \theta_0) = \mathbf{1}\{x_1^2 + x_2^2 \geqslant \theta_0\}\end{aligned}$$

U svim modelima parametri su realni brojevi, $\theta_j \in \mathbb{R}$. Razmotrite sve kombinacije modela $\mathcal{H}$ i funkcije preslikavanja ϕ. **Za koju kombinaciju modela $\mathcal{H}$ i funkcije preslikavanja ϕ postoji samo jedna hipoteza $h \in \mathcal{H}$ za koju $E(h|\mathcal{D}) = 0$?**

A $h_2 + \phi_2$ B $h_4 + \phi_1$ C $h_3 + \phi_2$ D $h_1 + \phi_1$

11. (P) Na slici ispod prikazan je graf funkcije pogreške učenje i pogreške ispitivanja za neku familiju modela i neki označeni skup primjera:



Crvenom linijom označena je složenost nekog modela $\mathcal{H}$. Crvene točke odgovaraju ispitnoj pogrešci i pogrešci učenja za hipotezu $h \in \mathcal{H}$ iz tog modela, dobivenoj nekim optimizacijskim algoritmom. **Što možemo reći o modelu $\mathcal{H}$ i o hipotezi h?**

A Model $\mathcal{H}$ nije optimalne složenosti, a čak ni hipoteza h ne mora biti optimalna na skupu za učenje, ako je optimizacijski algoritam loš

B Model H je podnaučen, ali je barem hipoteza h hipoteza s najmanjom ispitnom pogreškom unutar takvog suboptimalnog modela

C Model $\mathcal{H}$ je nedovoljne složenosti, ali je barem hipoteza h optimalna u smislu najmanje moguće pogreške na skupu za učenje

D Model $\mathcal{H}$ je prenaučen, a hipoteza h će loše generalizirati na neviđene primjere

12. (P) Raspolažemo modelom $\mathcal{H}_\alpha$, koji ima hiperparametar α kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra: α_1 i α_2. Treniramo modele $\mathcal{H}_{\alpha_1}$ i $\mathcal{H}_{\alpha_2}$ te dobivamo hipoteze h_{α_1} i h_{α_2}. Zatim računamo empirijske pogreške tih hipoteza na skupu za učenje $\mathcal{D}_u$ i na skupu za ispitivanje $\mathcal{D}_i$. Utvrđujemo da vrijedi:

$$E(h_{\alpha_1}|\mathcal{D}_i) - E(h_{\alpha_1}|\mathcal{D}_u) < E(h_{\alpha_2}|\mathcal{D}_i) - E(h_{\alpha_2}|\mathcal{D}_u)$$

Što iz toga možemo zaključiti?

A Model $\mathcal{H}_{\alpha_2}$ je prenaučen

B Optimalan model je onaj s vrijednošću hiperparametra iz intervala $[\alpha_1, \alpha_2]$

C Model $\mathcal{H}_{\alpha_1}$ je podnaučen

D Model $\mathcal{H}_{\alpha_1}$ je manje složenosti od modela $\mathcal{H}_{\alpha_2}$