

## 2. Osnovni koncepti

Strojno učenje 1, UNIZG FER, ak. god. 2021./2022.

Jan Šnajder, vježbe, v2.3

### 1 Zadaci za učenje

- [*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  $y$ ?
- [*Svrha: Razumjeti što je hipoteza, što je model i koja je veza između njih. Razumjeti što je prostor inačica.*]

(a) Dopunite praznine:

Hipoteza je funkcija koja preslikava \_\_\_\_\_ u \_\_\_\_\_, definirana do na \_\_\_\_\_. Model je \_\_\_\_\_ hipoteza, indeksiranih \_\_\_\_\_. Model također nazivamo prostorom \_\_\_\_\_, 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:

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

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

(e) Definirajte prostor inačica  $VS_{\mathcal{H}, \mathcal{D}}$ . Odredite  $VS_{\mathcal{H}, \mathcal{D}}$  za ovaj konkretan problem.

- [*Svrha: Shvatiti što je to induktivna pristranost i kako ona određuje klasifikaciju nevidenih 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) Raspoložemo skupom označenih primjera u ulaznome prostoru  $\mathcal{X} = \{0, 1\}^3$ :

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

Koja je klasifikacija nevidenih primjera?

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

(d) Možete li odrediti klasifikaciju nevidenih primjera uz odabrani model  $\mathcal{H}$ ? Je li pristranost koja proizlazi iz odabira modela dovoljna za jednoznačnu klasifikaciju primjera iz  $\mathcal{D}$ ? Odredite skup prostora inačica  $VS_{\mathcal{H}, \mathcal{D}}$ .

(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 i povezati ih s induktivnom pristranošću.]
- Nabrojite tri osnovne komponente algoritma strojnog učenja.
  - 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.]
- Pogreška hipoteze je očekivanje funkcije gubitka  $L$ . Nad kojom distribucijom je definirano to očekivanje? Koji je problem s takvom definicijom u praksi?
  - Definirajte empirijsku pogrešku preko funkcije gubitka  $L$ . Koja je pretpostavka implicitno ugrađena u tu definiciju?
  - 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.]
- Skicirajte krivulje pogreške učenja i ispitne pogreške u ovisnosti o složenosti modela. Naznačite područje prenaučivosti i podnaučivosti.
  - Objasnite zašto pogreška učenja s povećanjem složenosti modela teži k nuli.
  - Raspolažemo modelom  $\mathcal{H}_\alpha$  koji ima hiperparametar  $\alpha$  kojim se može ugađati složenost modela. Za odabrani  $\alpha$  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  $\alpha$  suboptimalan?
  - Raspolažemo modelom  $\mathcal{H}_\alpha$  s hiperparametrom  $\alpha$  (veći  $\alpha$  daje složeniji model). Raspolažemo dvama optimizacijskim algoritima:  $L_1$  i  $L_2$ . Algoritam  $L_2$  lošiji je od algoritma  $L_1$ , u smislu da  $L_2$  pronalazi parametre  $\theta_2$  koji su lošiji od parametara  $\theta_1$  koje pronalazi  $L_1$ , tj.  $E(\theta_2|\mathcal{D}) > E(\theta_1|\mathcal{D})$ . Neka  $\alpha_1^*$  označava optimalnu vrijednost hiperparametra za  $\mathcal{H}_\alpha$  učenog algoritmom  $L_1$ , a  $\alpha_2^*$  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  $\mathcal{H}_{\alpha_2^*}$ .
  - 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. (T) Model  $\mathcal{H}$  je skup svih parametriziranih funkcija  $h(\mathbf{x}; \theta)$  indeksiran parametrima  $\theta$ . To jest:

$$\mathcal{H} = \{h(\mathbf{x}; \theta)\}_\theta$$

Što to zapravo znači?

- A Da različite funkcije  $h$  imaju različite parametre  $\theta$ , i da su sve one sadržane u modelu, to jest za sve njih vrijedi  $h \in \mathcal{H}$
- B Da za različite parametre  $\theta$  dobivamo različite funkcije  $h$ , i da su sve one sadržane u modelu, to jest za sve njih vrijedi  $h \in \mathcal{H}$
- C Da model sadrži beskonačno mnogo funkcija  $h$  čija konkretna definicija ovisi o vrijednostima parametara  $\theta$
- D Da su funkcije  $h$  definirane sa slobodnim parametrima  $\theta$  i da broj različitih funkcija odgovara broju parametara

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

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

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

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

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

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  $\infty$     C 56    D 28

4. (P) Skup označenih primjera u dvodimenzijaskome ulaznom prostoru je:

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

Koja je veličina prostora inačica,  $|\mathcal{V}_{\mathcal{H}, \mathcal{D}}|$ ?

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

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

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

Razmatramo dva modela:

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

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

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

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

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

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

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, što se vidi iz toga da je za ovaj skup primjera prostor inačica veći od jedan
- D Možemo,  $y = 1$ , jer klasifikator ima induktivnu pristranost jezikom (linearan model) i preferencijom (primjeri za koje je  $h(x) \geq 0$  klasificiraju se pozitivno)

7. (P) Optimizacija parametara modela temelji se na funkciji gubitka  $L : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}_0^+$ , 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. (T) Pogreška modela definirana je kao očekivanje funkcije gubitka na primjerima iz  $\mathcal{X} \times \mathcal{Y}$ . Međutim, u praksi tu pogrešku aproksimiramo empirijskom pogreškom, koju računamo kao srednju vrijednost funkcije gubitka na skupu označenih primjera  $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$ . **Zašto pogrešku modela aproksimiramo empirijskom pogreškom i na kojoj se pretpostavci temelji ta aproksimacija?**

- A Očekivanje gubitka ne možemo izračunati jer primjera iz  $\mathcal{X} \times \mathcal{Y}$  ima potencijalno beskonačno, stoga pogrešku računamo na temelju skupa  $\mathcal{D}$  za koji pretpostavljamo da je konačan
- B Različiti primjera iz  $\mathcal{X} \times \mathcal{Y}$  potencijalno ima beskonačno mnogo, pa pogrešku računamo na uzorku  $\mathcal{D}$  za koji pretpostavljamo da je reprezentativan
- C Funkciju gubitka jednostavnije je definirati nego funkciju pogreške, a aproksimacija je točna uz pretpostavku i.i.d.
- D Ne možemo izračunati očekivanje gubitka jer nam nije poznata distribucija primjera iz  $\mathcal{X} \times \mathcal{Y}$ , no pretpostavljamo da je  $\mathcal{D}$  reprezentativan uzorak iz te distribucije

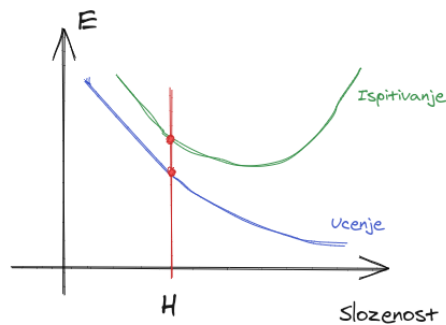
9. (P) Zadan je sljedeći skup sa  $N = 6$  označenih primjera iz  $\mathbb{R}^3$ :

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

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  $0 \leq E(h|\mathcal{D}) \leq 1/4$
- B  $1/4 \leq E(h|\mathcal{D}) \leq 2/3$
- C  $\frac{1}{48} \leq E(h|\mathcal{D}) \leq 2/3$
- D  $1/12 \leq E(h|\mathcal{D}) \leq 3/4$

10. (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 empirijskoj 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 prenaučan, a hipoteza  $h$  će loše generalizirati na neviđene primjere
11. (T) Modeli strojnog učenja tipično imaju i parametre i hiperparametre. **Koja je razlika između parametara i hiperparametara?**
- A Algoritam strojnog učenja minimizira parametre te istovremeno maksimizira hiperparametre
  - B Hiperparametri mogu biti diskretni ili kontinuirani, dok su parametri uvijek kontinuirani
  - C Parametre optimira algoritam strojnog učenja, dok optimizacija hiperparametara nije u nadležnosti tog algoritma
  - D Parametri određuju iznos empirijske pogreške na skupu za učenje, a hiperparametri iznos te pogreške na skupu za provjeru
12. (P) Raspoložemo modelom  $\mathcal{H}_\alpha$  koji ima hiperparametar  $\alpha$  kojim se može ugađati složenost modela. Isprobavamo dvije vrijednosti hiperparametra:  $\alpha_1$  i  $\alpha_2$ . Treniramo modele  $\mathcal{H}_{\alpha_1}$  i  $\mathcal{H}_{\alpha_2}$  te dobivamo hipoteze  $h_{\alpha_1}$  i  $h_{\alpha_2}$ . 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 prenaučan
- B Optimalan model je onaj s hiperparametrom 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  $\mathcal{H}_{\alpha_2}$