

Strojno učenje - 1. Međuispit

1.

- a) Objasniti razliku između nadziranog i nenadziranog učenja.
- b) Kakvi su neparametarski postupci? Jeli regresija neparametarski postupak? Zašto?

2.

Želimo naučiti Booleovu funkciju u prostoru primjera $X=\{0,1\}^3$. Razmatramo tri modela: H_1 je skup svih logičkih funkcija, H_2 je skup logičkih funkcija prikazivih u konjunktivnoj normalnoj formi s točno dva konjunkt (npr. $h(x_1, x_2, x_3)=((x_1 \vee \neg x_2) \wedge (x_2 \vee x_3))$), a H_3 je linearan model u R^3 . Skup primjera za učenje je:

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

- a) Odredite $|H_1|$ te objasnite odnose (prema relaciji \subset) između skupova H_1, H_2, H_3 .
- b) Neka je hipoteza $h \in H_3$ definirana s $h(x_1, x_2, x_3)=1(x_3 < 0,5)$. Koji od primjera iz D zadovoljava hipoteza h , koji su primjeri konzistentni s h te kolika je empirijska pogreška ($E(h|D)$)?
- c) Formalno definirati relaciju \geq_g . Za hipotezu $h_1(x_1, x_2, x_3)=(x_1 \rightarrow (x_2 \wedge x_3))$ i hipotezu $h_2(x_1, x_2, x_3)=1(x_3 \leq 0,5)$ dokažite da vrijedi $h_1 \geq_g h_2$.
- d) Formalno definirajte prostor inačica. Koje hipoteze iz H_3 čine prostor inačica za gore navedeni skup primjera D i koja je od njih najopćenitija?
- e) Objasniti ideju aktivnog učenja. Koji bi upit algoritam trebao generirati za prostor inačica u gornjem slučaju?
- f) Objasniti što je induktivna pristranost i koje vrste pripadnosti postoje. Kakvu vrstu pristranosti predstavlja odabir modela H_1 , a kakvu odabir modela H_2 ?

3.

- a) Formalno definirajte razdjeljivanje primjera i VC-dimenziju modela H .
- b) Izvedite VC- dimenziju za model $h(x_1, x_2 | \Theta_0, \Theta_1, \Theta_2)=\text{sgn}(\Theta_1 x_1 + \Theta_2 x_2 + \Theta_0)$. Jeli model s tako malom VC-dimenzijom primjenjiv u praksi i zašto?
- c) Pokažite da je VC-dimenzija sljedećeg modela barem 4:
 $h(x_1, x_2 | \Theta_{x_1}, \Theta_{y_1}, \Theta_{x_2}, \Theta_{y_2})=1((\Theta_{x_1} \leq x_1 \leq \Theta_{x_2}) \wedge (\Theta_{y_1} \leq x_2 \leq \Theta_{y_2}))$.
- d) U nekom prostoru primjera X razmatramo 2 modela, H_1 i H_2 , takva da $H_1 \cap H_2 \neq \emptyset$. Neka $VC(H_1)=2$ i $VC(H_2)=5$. Odredite vrijednost (ili skup mogućih vrijednosti) za $VC(H_1 \cup H_2)$ odnosno $VC(H_1 \cap H_2)$ te obrazložite odgovor.

4.

- a) Izvedite jednadžbe za parametre w_0 i w_1 linearno regresijskog modela $h(x|w_0, w_1)=w_1 x + w_0$ uz funkciju gubitka:
 $L(y^{(i)}, h(x^{(i)}|w_0, w_1)) = (y^{(i)} - h(x^{(i)}|w_0, w_1))^2$.

- b) Neka je skup primjera generiran (nama u stvarnosti nepoznatom funkcijom) $f(x) = 6x^2 - 4x - 2 - \epsilon$, gdje je ϵ slučajan, uniformno distribuiran šum. Za regresiju koristimo polinom 5. stupnja. Objasniti kako parametri w_i optimalne hipoteze ovise o (i) broju primjera i (ii) količini šuma.

5.

- a) Objasniti pojmove generalizacije, prenaučivosti i podnaučivosti.
 b) Skicirajte odnos empirijske pogreške i pogreške generalizacije u ovisnosti o složenosti modela te ukratko objasniti odabir modela unakrsnom provjerom.
 c) Odabrali smo model H i naučili hipotezu koja minimizira empirijsku pogrešku. Unakrsnom provjerom ustanovili smo da je pogreška generalizacije približno izjednačena s empirijskom pogreškom. Što možemo zaključiti o modelu H (u smislu generalizacije) i što možemo očekivati ako povećamo složenost modela odnosno broj primjera?

6. Tu je bio još neki uvodni tekst....

i	tipkovnica	OS	rezolucija	$y^{(i)}$
1	ima	Android	visoka	1
2	nema	Android	niska	0
3	nema	Android	visoka	1
4	nema	WP7	visoka	0

Uporabom CE algoritma naći prostor inačica. Skicirati dobiveni prostor inačica i uređaj općenitosti hipoteza. Primjere obraditi zadanim redoslijedom, u svakom koraku ispisati sadržaj S i G i ukratko objasniti što se događa.