

Strojno učenje – domaća zadaća 1

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

Zadano: 13.10.2013. Rok: 18.10.2013.

Napomena: Zadatke možete rješavati samostalno ili u grupi. Ako zadatke rješavate u grupi, pobrinite se da svi članovi grupe pridonose rješenju i da ga naposljetku 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: 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, (b) predviđanje kretanja dionica, (c) rangiranje rezultata tražilice? Kako biste u ovim slučajevima definirali ciljne oznake y ?
2. [*Svrha: Razumjeti koncept VC-dimenzije na konkretnim primjerima.*] Razmatramo sljedeće modele u ulaznom prostoru $\mathcal{X} = \mathbb{R}^2$:
 - (a) kružnice sa središtem u ishodištu koordinatnog sustava;
 - (b) skup parova koncentričnih kružnica sa središtem u ishodištu (primjeri unutar prstena pozitivni, a svi ostali su negativni);
 - (c) kružnice sa središtem u proizvoljnoj točki.

Definirajte svaki od ovih modela (napišite definiciju funkcije hipoteze $h(\mathbf{x})$) te odredite VC-dimenziju svakog modela.

3. [*Svrha: Razviti intuiciju o tome što je sve obuhvaćeno definicijom modela.*] Definirajte model $\mathcal{H}_1 =$ “skup pravokutnika sa stranicama paralelnima s osima”. Koliko parametara ima taj model i koja je njegova VC-dimenzija? Definirajte model \mathcal{H}_2 kod kojeg pozitivni primjeri mogu biti ili unutar pravokutnika ili izvan njega (različito za svaku hipotezu). Koja je VC-dimenzija takvog modela?
4. [*Svrha: Produbiti shvaćanje modela kao skupa hipoteza.*] Pokažite da, ako za dva modela \mathcal{F} i \mathcal{G} vrijedi $\mathcal{F} \subseteq \mathcal{G}$, onda vrijedi $VC(\mathcal{F}) \leq VC(\mathcal{G})$.
5. [*Svrha: Razumjeti kako induktivna pristranost određuje klasifikaciju nevidenih primjera.*] Definirajte linearan model \mathcal{H} za ulazni prostor $\mathcal{X} = \{0, 1\}^3$.
 - (a) Koja je dimenzija prostora hipoteza \mathcal{H} ? Koliko različitih hipoteza postoji u tom prostoru, tj. koliko iznosi $|\mathcal{H}|$?
 - (b) Koja je veličina i koja je dimenzija prostora parametara modela?
 - (c) Raspoložom skupom primjera:

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

Odredite klasifikaciju nevidenih primjera te odredite skup prostora inačica VS .

- (d) Definirajte induktivnu pristranost takvu da $|VS| = 1$, tj. da klasifikacija svakog primjera slijedi jednoznačno na temelju skupa primjera \mathcal{D} .
6. [*Svrha: Shvatiti uzročne veze između naoko nevezanih veličina.*] Obrazložite u kakvim su odnosima sljedeći pojmovi: (a) složenost modela, (b) broj parametara modela, (c) dimenzija ulaznog prostora n i (d) broj primjera N . Analizirajte odnose između svih parova pojmova, posebno za parametarske, a posebno za neparametarske modele.
7. [*Svrha: Podsjetiti se matematike, iskušati regresiju na konkretnom primjeru te shvatiti vezu između ulaznog prostora i prostora parametara.*]
- (a) Definirajte model linearne regresije za $n = 1$, $\mathcal{X} = \mathbb{R}$, definirajte empirijsku pogrešku kao kvadratno odstupanje te izvedite parametre koji minimiziraju tu pogrešku.
- (b) Želimo predviđati prihod osobe na temelju njezine dobi. Raspoložemo malim skupom primjera:
- $$\mathcal{D} = \{(20, 6K), (22, 5K), (25, 5K), (30, 8K), (40, 7K), (50, 9K), (65, 3K)\}.$$
- Naučite model linearne regresije i odredite prihod osobe stare 35 godina.
- (c) Skicirajte (a) ulazni prostor i (b) prostor hipoteza (prostor parametara) za naučenu (optimalnu) hipotezu i dvije (suboptimalne) hipoteze. Odredite $E(h|\mathcal{D})$ za sve tri hipoteze. Kako izgleda funkcija $E(h(\mathbf{x}|w_1, w_0)|\mathcal{D})$ kao funkcija od (w_1, w_0) ?
- (d) Bi li složeniji model (npr. polinom drugog stupnja) bio prikladniji za ovaj problem? Obrazložite.
8. [*Svrha: Naučiti uočavati zajedničku strukturu u različitim algoritmima.*]
- (a) Nabrojite tri osnovne komponente algoritma strojnog učenja.
- (b) Identificirajte uz koje se komponente veže koja vrsta induktivne pristranosti.
- (c) Identificirajte te tri komponente kod linearne regresije.
9. [*Svrha: Razumjeti vezu između funkcije gubitka i empirijske pogreške te mogućnost njihove prilagodbe konkretnom problemu.*]
- (a) Definirajte empirijsku pogrešku preko funkcije gubitka L .
- (b) Kod asimetričnih gubitaka funkciju L možemo definirati preko matrice gubitka. Definirajte takvu matricu za problem klasifikacije neželjene e-pošte te izračunajte funkciju pogreške za slučaj pet pogrešno negativnih i dvije pogrešno pozitivne klasifikacije.
10. [*Svrha: Razviti ispravnu intuiciju za odabir modela temeljem unakrsne provjere.*]
- (a) Skicirajte krivulje empirijske pogreške (pogreške učenja) i pogreške generalizacije u ovisnosti o složenosti modela. Naznačite područje prenaučivosti i podnaučenosti.

- (b) Raspoložemo modelom \mathcal{H}_α koji ima hiperparametar α kojim se može ugađati složenost modela. Za odabrani α naučili smo hipotezu koja minimizira empirijsku pogrešku. Unakrsnom provjerom ustanovili smo da je pogreška generalizacije znatno veća od empirijske pogreške. Je li naš odabir hiperparametra α suboptimalan? Obrazložite odgovor.
- (c) Raspoložemo modelom \mathcal{H}_α s hiperparametrom α (veći α daje složeniji model). Raspoložemo dvama optimizacijskim algoritmima: L_1 i L_2 . Algoritam L_2 lošiji je od algoritma L_1 , u smislu da L_2 pronalazi parametre θ_2 koji su lošiji od parametara θ_1 koje pronalazi L_1 , tj. $E(\theta_2|\mathcal{D}) > E(\theta_1|\mathcal{D})$. Neka α_1^* označava optimalnu vrijednost hiperparametra za \mathcal{H}_α učenog algoritmom L_1 , a α_2^* optimalnu vrijednost za \mathcal{H}_α učenog algoritmom L_2 . Načinite skicu analognu onoj iz zadatka (a) i naznačite vrijednosti pogrešaka koje odgovaraju modelima $\mathcal{H}_{\alpha_1^*}$ i $\mathcal{H}_{\alpha_2^*}$.
- (d) Može li model učen lošijim algoritmom L_2 imati manju pogrešku generalizacije od modela koji je učen boljim algoritmom L_1 , ali nije optimalan? Skicirajte takvu situaciju na prethodnoj skici.