21. Vrednovanje modela

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

Jan Šnajder, vježbe, v1.1

1 Zadatci za učenje

1. [***Svrha: Izvježbati izračun mjera uspješnosti modela na konkretnom primjeru.***]

   Raspolažemo skupom od 11 ispitnih primjera koje želimo klasificirati u tri klase. Oznaka $y^{(i)}$ i izlaz modela $h(\mathbf{x}^{(i)})$ za svaki od 11 primjera su sljedeći:

$$\left\{\left(y^{(i)}, h(\mathbf{x}^{(i)})\right)\right\}_{i=1}^{11} = \\ \left\{(1,1),(0,2),(2,2),(1,2),(1,1),(0,0),(1,1),(2,1),(0,1),(2,0),(2,1)\right\}.$$

   (a) Izračunajte točnost klasifikatora.

   (b) Izračunajte preciznost, odziv i mjeru F_1, i to *mikro* i *makro* varijante.

2. [***Svrha: Izvježbati izračun mjere F_1 na temelju parcijalno zadane matrice zabune.***] Od $N = 1000$ primjera, klasifikator je za prvu, drugu i treću klasu ispravno pozitivno klasificirao njih 620, 146 odnosno 134. Od preostalih 100 neispravno klasificiranih primjera, 50 ih je klasificirano u drugu klasu umjesto u prvu, 20 u drugu umjesto u treću, a 30 u treću umjesto u drugu klasu. Izračunajte makro-F_1.

3. [***Svrha: Znati kako na temelju probabilističkog izlaza klasifikatora skicirati krivulju ROC. Znati mjerom AUC možemo usporediti klasifikator s nasumičnim klasifikatorom. Znati kako pomoću krivulje ROC uspoređivati klasifikatore međusobno.***] Na ispitnome skupu od $N = 10$ primjera evaluiramo tri binarna klasifikatora: logističku regresiju (h_{LR}), naivan Bayesov klasifikator (h_{NB}) i stroj potpornih vektora s probabilističkim izlazom dobivenim metodom Plattove kalibracije (h_{SVM}). Stvarne oznake primjera $y^{(i)}$ i vjerojatnosne predikcije triju klasifikatora $h(\mathbf{x}^{(i)}) = p(y = 1|\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

i	1	2	3	4	5	6	7	8	9	10
$y^{(i)}$	1	1	0	0	1	1	1	0	0	1
$h_{\mathrm{LR}}(\mathbf{x})$	0.8	0.6	0.8	0.6	0.8	0.8	0.8	0.2	0.2	0.2
$h_{\mathrm{NB}}(\mathbf{x})$	0.3	0.8	0.3	0.5	0.8	0.3	0.8	0.5	0.3	0.5
$h_{\mathrm{SVM}}(\mathbf{x})$	0.6	0.1	0.7	0.6	0.1	0.7	0.7	0.6	0.1	0.7

   Na temelju ovog uzorka želimo procijeniti krivulju ROC te mjeru AUC (površinu ispod krivulje ROC). Prisjetite se da krivulja ROC opisuje TPR (odziv) kao funkciju od FPR (stopa lažnog alarma).

   (a) Skicirajte krivulje ROC za ova tri klasifikatora, linearno interpolirajući između točaka koje odgovaraju opaženim vjerojatnosnim izlazima klasifikatora.

   (b) Skicirajte krivulje ROC za ova tri klasifikatora.

   (c) Izračunajte mjere AUC za sva tri klasifikatora.

   (d) Kako izgleda krivulja ROC za nasumični klasifikator. Zašto?

   (e) Koji je od navedenih klasifikatora lošiji od nasumičnog klasifikatora, a koji biste klasifikator odabrali kao najbolji?

4. [***Svrha: Razumjeti na koji se način provodi ugniježđena unakrsna provjera, kako se razdjeljuju primjeri kroz iteracije petlji te kako ugraditi dodatne predobradbe značajki, a pritom ne kompromitirati podjelu na skup za učenje i skup za ispitivanje.***] Raspolažemo sa 1000 označenih primjera. Za vrednovanje SVM-a s hiperparametrima C i γ koristimo ugniježđenu unakrsnu provjeru sa po 5 ponavljanja u obje petlje. Hiperparametre optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-5}, 2^{-4}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^{3}\}$.

   (a) Koliko ćemo ukupno puta trenirati model?

   (b) Koliko ćemo primjera u svakoj od iteracija koristiti za treniranje, koliko za provjeru, a koliko za ispitivanje?

   (c) Kako glase odgovori na prethodna dva pitanja, ako bismo u vanjskoj petlji umjesto peterostruke unakrsne provjere koristili unakrsnu provjeru *izdvoji jednoga* (engl. *leave one out, LOOCV*)?

   (d) Klasifikator SVM posebno je osjetljiv na razlike u rasponima između značajki (zašto?), pa se preporuča standardizirati značajke. Što to točno znači i kako biste standardizaciju značajki ugradili u ugniježđenu unakrsnu provjeru?

   (e) Gdje biste u ugniježđenu unakrsnu provjeru ugradili odabir značajki modela i optimizaciju praga po mjeri AUC?

2 Zadatci s ispita

1. (N) Na ispitnome skupu evaluiramo klasifikator sa $K = 3$ klase. Dobili smo sljedeću matricu zabune (stupci su stvarne oznake, a retci oznake koje daje klasifikator):

$$\begin{array}{c} \\ y=1 \\ y=2 \\ y=3 \end{array} \begin{array}{c} \begin{array}{ccc} y=1 & y=2 & y=3 \end{array} \\ \begin{pmatrix} 15 & 3 & 1 \\ 6 & 5 & 4 \\ 4 & 2 & 23 \end{pmatrix} \end{array}$$

   Izračunajte mikro-F1 (F_1^{μ}) i makro-F1 (F_1^{M}) mjere na ovoj matrici zabune. **Koliko iznosi razlika između vrijednosti mikro-F1 i makro-F1 mjere, $F_1^{\mu} - F_1^{M}$?**

   A 0.01   B 0.05   C 0.09   D 0.13

2. (P) Evaluirali smo model višeklasne logističke regresije i dobili da je vrijednost mjere makro-$F_{0.5}$ znatno manja od vrijednosti mjere makro-F_1, a da se vrijednosti mjere mikro-F_1 i mjere makro-F_1 znatno ne razlikuju. **Što možemo zaključiti o ovom modelu?**

   A Preciznost modela lošija je od odziva, ali model ne radi znatno lošije na primjerima iz klasa s manjim brojem primjera, pa su zbog toga mjere mikro-F_1 i makro-F_1 izjednačene

   B Odziv modela lošiji je od preciznosti, osim na primjerima iz klasa s velikim brojem primjera, gdje se preciznost i odziv znatno ne razlikuju

   C Model ima manji odziv na primjerima iz manjih klasa, ali veću preciznost na primjerima iz većih brojem primjera, pa su mjere mikro-F_1 i makro-F_1 izjednačene

   D Preciznost i odziv modela za pojedine klase znatno se ne razlikuju, neovisno o broju primjera u dotičnoj klasi, pa se onda znatno ne razlikuju niti mjere mikro-F_1 i makro-F_1

3. (N) Na ispitnome skupu od $N = 10$ primjera evaluiramo binarnu logističku regresiju. Stvarne oznake primjera $y^{(i)}$ i predikcije klasifikatora $h(\mathbf{x}^{(i)})$ na tom skupu su sljedeće:

$$\big\{(y^{(i)}, h(\mathbf{x}^{(i)})\big\}_{i=1}^{10} = \big\{(1,1),(0,0),(0,1),(1,0),(1,1),(1,1),(0,1),(0,0),(0,1),(0,1)\big\}$$

   Kao referentni model koristimo klasifikator koji nasumično pogađa oznaku y, birajući s jednakom vjerojatnošću između oznaka $y = 0$ i $y = 1$. Za evaluaciju klasifikatora koristimo F_2-mjeru umjesto F_1-mjere. Izračunajte F_2-mjeru logističke regresije i referentnog modela. **Koliko će očekivano logistička regresija biti bolja od referentnog modela po F_2-mjeri?**

   A 0.095   B 0.051   C 0.101   D 0.176

4. (T) Procjena pogreške modela metodom unakrsne provjere omogućava nam da procijenimo prediktivnu moć modela, mjerenu kao točnost modela na neviđenom skupu primjera. Daljnja razrada te ideje je ugniježđena višestruka unakrsna provjera, koja se u praksi vrlo često koristi. **Koja je motivacija za korištenje ugniježđene višestruke unakrsne provjere, umjesto obične unakrsne provjere?**

   A Razdvaja skup za učenje od skupa za ispitivanje te time osigurava da doista mjerimo prediktivnu moć modela, odnosno ispitnu pogrešku, a ne pogrešku učenja

   B Omogućava nam da odredimo točnost modela s klasifikacijskim pragom, na način da u obzir uzimamo preciznost i odziv za različite vrijednosti klasifikacijskog praga

   C Provodi optimizaciju hiperparametra modela na uniji skupa za provjeru i skupa za testiranje, čime postiže bolju točnost modela jer više primjera ostaje za treniranje

   D Omogućava nam da procijenimo prediktivnu moć modela optimalne složenosti te maksimalno iskoristimo raspoložive podatke za učenje i ispitivanje

5. (P) Raspolažemo sa 1000 označenih primjera. Na tom skupu treniramo i evaluiramo algoritam SVM. Pritom razmatramo tri hiperparametra: jezgra (linearna ili RBF), regularizacijski faktor C i preciznost RBF jezgre γ. Posljednja dva hiperparametra optimiramo rešetkastim pretraživanjem u rasponima $C \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$ i $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^{15}\}$. Naravno, ako ne koristimo RBF-jezgru, onda hiperparametar γ ne optimiramo. Za treniranje i evaluaciju modela koristimo ugniježđenu unakrsnu provjeru s 10 ponavljanja u vanjskoj petlji i 5 ponavljanja u unutarnjoj petlji. **Koliko će puta svaki primjer biti iskorišten za treniranje modela?**

   A 35721 B 44640 C 49600 D 69201

6. (P) Evaluiramo model L_2-regularizirane logističke regresije. Za evaluaciju koristimo ugniježđenu unakrsnu provjeru u kojoj optimiramo regularizacijski faktor λ. Neka je λ_1 prosjek optimalnih vrijednosti regularizacijskog faktora, i neka je F_1^1 prosječna F_1-mjera na ispitnom skupu vanjske petlje. Međutim, naknadno smo ustanovili da nam se potkrala pogreška i da smo u unutarnjoj petlji model uvijek ispitivali na prvom preklopu. Kada to ispravimo, dobivamo λ_2 kao prosjek optimalnih vrijednosti regularizacijskog faktora i F_1^2 kao prosjek F_1-mjere na ispitnom skupu vanjske petlje. Nažalost, kasnije smo ustanovili da nam se potkrala još jedna pogreška: umjesto da u vanjskoj petlji optimalan model treniramo na cijelom skupu za treniranje, mi smo ga trenirali samo na skupu za treniranje zadnje iteracije unutarnje petlje. Kada i tu pogrešku ispravimo, dobivamo λ_3 odnosno F_1^3. **Što možemo očekivati o odnosima između procjena za optimalni λ i za F_1-mjeru na ispitnom skupu**?

   A $\lambda_1 > \lambda_2 > \lambda_3$, $F_1^1 < F_1^2$, $F_1^3 < F_1^2$

   B $\lambda_1 < \lambda_3$, $F_1^1 < F_1^2 < F_1^3$

   C $\lambda_1 < \lambda_2 = \lambda_3$, $F_1^1 > F_1^2$, $F_1^3 > F_1^2$

   D $\lambda_1 = \lambda_3 < \lambda_2$, $F_1^2 < F_1^1$, $F_1^3 < F_1^2$