

## 21. Vrednovanje modela

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

Jan Šnajder, natuknice s predavanja, v1.3

### 1 Osnovne mjere vrednovanja

- **Matrica zabune** (*confusion matrix*) – usporeba stvarnih oznaka i predikcija modela

		Stvarno	
		1	0
Model	1	TP	FP
	0	FN	TN

TP – true positives, FP – false positives, FN – false negatives, TN – true negatives

- **Točnost** (*accuracy*) je udio točno klasificiranih primjera u skupu svih primjera:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} = 1 - E(h|\mathcal{D})$$

- Ako je udio klasa izrazito neuravnotežen, točnost nije indikativna mjera

- **Preciznost** (*precision*):

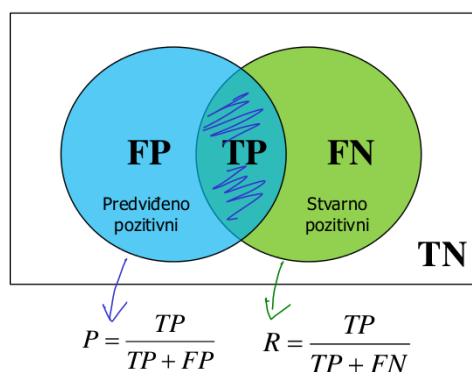
$$P = \frac{TP}{TP + FP}$$

⇒ udio pozitivno klasificiranih primjera u skupu pozitivno klasificiranih primjera

- **Odziv** (*recall, true positive rate, sensitivity*):

$$R = TPR = \frac{TP}{TP + FN}$$

⇒ udio pozitivno klasificiranih primjera u skupu svih pozitivnih primjera



- **Fall-out** (false positive rate)

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

⇒ udio primjera pogrešno proglašanih pozitivnima

- **Specifičnost** (*specificity*):

$$S = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

⇒ udio negativno klasificiranih primjera u skupu svih negativnih primjera

- **Mjera F1** – harmonijska sredina preciznosti i odziva:

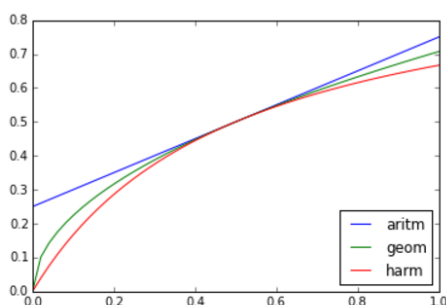
$$F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2PR}{P + R}$$

- **Mjera F-beta** – poopćenje mjere F1 koje različito naglašava  $P$  i  $R$ :

$$F_\beta = \frac{(1 + \beta^2)PR}{\beta^2 P + R}$$

⇒  $F_{0.5}$  dvostruko naglašava preciznost,  $F_2$  dvostruko naglašava odziv

- Harmonijska sredina je “najstroža” od triju sredina; npr. za  $R = 0.5$  i  $P \in [0, 1]$ :



- Primjer:  $N = 1000$ , od čega 100 poz. Ispravno klasificiranih 90 poz. i 650 neg.

		Stvarno	
		1	0
Model	1	90	250
	0	10	650

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} = 0.74$$

$$P = \frac{TP}{TP + FP} = \frac{90}{90 + 250} = 0.265$$

$$R = \frac{TP}{FP + FN} = \frac{90}{90 + 10} = 0.9$$

$$F_1 = \frac{2PR}{P + R} = \frac{2 \cdot 0.265 \cdot 0.9}{0.265 + 0.9} = 0.409$$

## 2 Višeklasna klasifikacija

- Iz matrice  $K \times K$  ( $K > 2$ ) izvodimo matricu  $2 \times 2$  za svaku klasu  $j$ , s elementima:
  - $TP_j$  –  $j$ -ti element dijagonale
  - $FP_j$  – zbroj nedijagonalnih elemenata  $j$ -tog retka
  - $FN_j$  – zbroj nedijagonalnih elemenata  $j$ -tog stupca
  - $TN_j = N - TP_j - FP_j - FN_j$  – zbroj po elementima izvan retka  $j$  i stupca  $j$

- **Makro-prosjek** ( $M$ ): izračun mjere za svaku klasu pa uprosječivanje kroz klase

$$Acc^M = \frac{1}{K} \sum_{j=1}^K Acc_j, \quad P^M = \frac{1}{K} \sum_{j=1}^K P_j, \quad R^M = \frac{1}{K} \sum_{j=1}^K R_j, \quad F_1^M = \frac{1}{K} \sum_{j=1}^K F_{1,j}$$

$\Rightarrow$  jednak utjecaj svih klasa  $\Rightarrow$  loš rezultat na manjim klasama narušava mjeru

- **Mikro-prosjek** ( $\mu$ ): zbrajanje matrica pojedinačnih klasa pa izračun mjere

$$TP = \sum_{j=1}^K TP_j, \quad FP = \sum_{j=1}^K FP_j, \quad FN = \sum_{j=1}^K FN_j, \quad TN = \sum_{j=1}^K TN_j$$

$\Rightarrow$  vrijedi  $FP = FN \Rightarrow$  vrijedi  $P^\mu = R^\mu = F_1^\mu$

- Vrijedi  $Acc^M = Acc^\mu$
- Alternativa: neuprosječena točnost –  $Acc = \frac{1}{N} \sum_{j=1}^K TP_j = P^\mu = R^\mu = F_1^\mu$
- Primjer ( $N = 13, K = 3$ ):

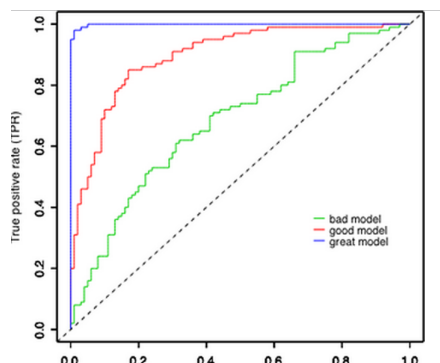
$$\begin{array}{c}
 y = 1 \quad y = 2 \quad y = 3 \\
 y = 1 \begin{pmatrix} 1 & 1 & 0 \\ 2 & 2 & 3 \\ 0 & 0 & 4 \end{pmatrix} \Rightarrow \overbrace{\begin{pmatrix} y = 1 & y = 2 & y = 3 \\ (1 \ 1) & (2 \ 5) & (4 \ 0) \\ (2 \ 9) & (1 \ 5) & (3 \ 6) \end{pmatrix}}^{\text{Makro}} \Rightarrow \overbrace{\begin{pmatrix} \text{zbroj} \\ (7 \ 6) \\ (6 \ 20) \end{pmatrix}}^{\text{Mikro}}
 \end{array}$$

$$\begin{array}{ll}
 Acc^M = \frac{1}{3} \left( \frac{10}{13} + \frac{7}{13} + \frac{10}{13} \right) = 0.69 & Acc^\mu = \frac{27}{39} = 0.69 \\
 P^M = \frac{1}{3} \left( \frac{1}{2} + \frac{2}{7} + \frac{4}{4} \right) = 0.60 & P^\mu = \frac{7}{13} = 0.54 \\
 R^M = \frac{1}{3} \left( \frac{1}{3} + \frac{2}{3} + \frac{4}{7} \right) = 0.52 & R^\mu = \frac{7}{13} = 0.54 \\
 F_1^M = \frac{1}{3} (0.40 + 0.40 + 0.73) = 0.51 & F_1^\mu = \frac{2P^M R^M}{P^M + R^M} = 0.54
 \end{array}$$

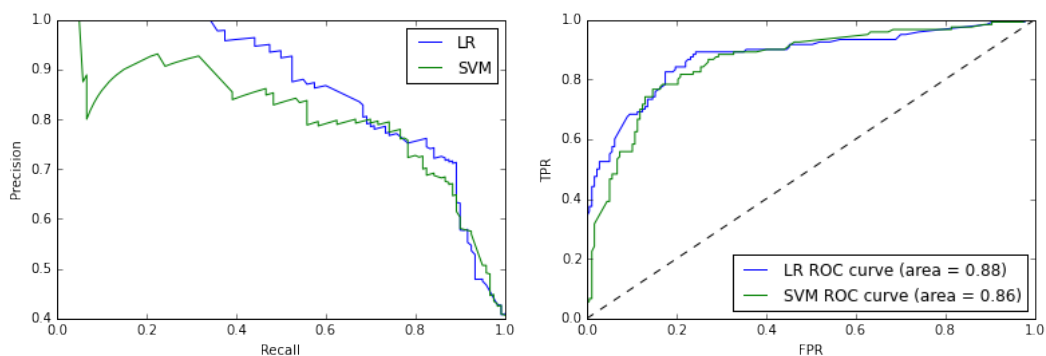
- Tipično (ali ne nužno)  $M < \mu$  jer klasifikatori rade lošije na manjim klasama

### 3 Vrednovanje klasifikatora s pragom

- Ugađanjem klasifikacijskoj praga može se ugađati  $P$  i  $R$  modela
- **Krivulja preciznost-odziv (P-R)** – preciznost kao funkcija odziva (monotono opada)
- Agregatna mjera: **prosječna preciznost (AP)** (*average precision*)
- **Krivulja ROC** – odziv kao funkcija od FPR (fall-out)

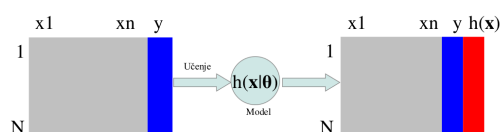


- Nasumična predikcija  $\Rightarrow$   $TPR = FPR$ , neovisno o udjelu pozitivnih primjera
- Agregatna mjera: **površina ispod ROC krivulje (AUC)** (*area under curve*)
- Najbolji model:  $(1, 1)$  za krivulju P-R,  $(0, 1)$  za krivulju ROC



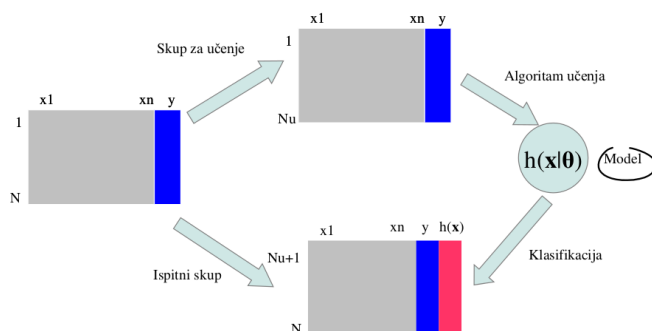
### 4 Procjena pogreške modela

- Ispitni skup je **slučajan uzorak**  $\Rightarrow$  svaka mjera točnosti je funkcija **slučajne varijable**
- Procjena pogreške (točnosti) treba biti **dobra** (nepristrana) i **poštena** (realistična)
- Procjena na skupu za učenje  $\Rightarrow$  ne mjerimo pogrešku generalizacije  $\Rightarrow$  nepošteno



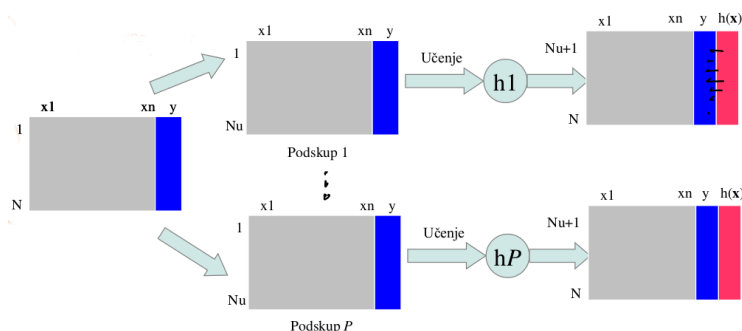
- **Metoda izdvajanja** (*holdout method*)

- Podjela na skup za učenje i skup za ispitivanje (npr., 70%–30%)
- Prednost: mjerimo pogrešku generalizacije
- Nedostatci: gubitak primjera za učenje, procjena na samo jednom uzorku



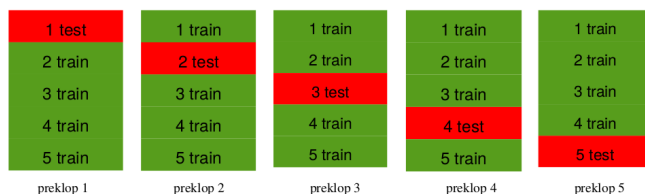
- **Ponovljeno izdvajanje** (*repeated holdout*)

- Višestruko uzorkovanje skupova za učenje/ispitivanje pa izračun prosjeka mjere
- Prednost: procjena pogreške generalizacije na više uzorka
- Nedostatak: ne kontroliramo koji su primjeri i koliko puta upotrijebljeni



- **k-struka unakrsna provjera (CV)** (*k-folded cross-validation*)

- Podjela na  $k$  **preklopa** (*folds*) (tipično  $k = 5$  ili  $k = 10$ )
- Učenje na  $(k - 1)$  preklopa, ispitivanje na jednom preklopu, ponovljeno  $k$  puta
- Prednost: svaki je primjer iskorišten i za učenje i za ispitivanje
- Nedostatak: modeli nisu međusobno nezavisni  $\Rightarrow$  visoka varijanca procjene



- **Metoda izdvoji jednoga (LOOCV)** (*leave-one-out cross-validation*)
  - $k$ -struka unakrsna provjera uz  $k = N$
  - Prednost: gotovo svi primjeri se koriste za učenje u svakoj iteraciji
  - Nedostatci: računalno skupo, visoka varijanca procjene pogreške
- Procjena pogreške uz **odabir modela**:
  - Podjela na skup za **učenje** ( $\mathcal{D}_{\text{train}}$ ), **provjeru** ( $\mathcal{D}_{\text{validate}}$ ) i **ispitivanje** ( $\mathcal{D}_{\text{test}}$ )
  - Odabir modela: učenje na  $\mathcal{D}_{\text{train}}$  i ispitivanje na  $\mathcal{D}_{\text{validate}}$
  - Ispitivanje odabranog modela: učenje na  $\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{validate}}$  i ispitivanje na  $\mathcal{D}_{\text{test}}$
- $k$ -struka CV uz odabir modela  $\Rightarrow$  **ugniježđena unakrsna provjera** (*nested CV*)

### Ugniježđena unakrsna provjera $k \times l$

- 1: podijeli  $\mathcal{D}$  na vanjske preklape  $\mathcal{D}_i, i = 1, \dots, k$
- 2: **za**  $i = 1, \dots, k$  **radi**: *vanjska petlja*
- 3:      $\mathcal{D}_{\text{train}} \leftarrow \mathcal{D} \setminus \mathcal{D}_i, \mathcal{D}_{\text{test}} \leftarrow \mathcal{D}_i$
- 4:     **za svaku** odabranu vrijednost hiperparametra  $\alpha$  **radi**:
- 5:         podijeli  $\mathcal{D}_{\text{train}}$  na unutarnje preklape  $\mathcal{D}_j, j = 1, \dots, l$
- 6:         **za**  $j = 1, \dots, l$  **radi**: *unutarnja petlja*
- 7:              $\mathcal{D}_{\text{train}'} \leftarrow \mathcal{D}_{\text{train}} \setminus \mathcal{D}_j, \mathcal{D}_{\text{validate}} \leftarrow \mathcal{D}_j$
- 8:             nauči model na  $\mathcal{D}_{\text{train}'}$  i ispitaaj na  $\mathcal{D}_{\text{validate}}$
- 9:             izračunaj prosjek mjere na  $l$  unutarnjih preklopa
- 10:         odaberi hiperparametar  $\alpha$  koji maksimizira prosjek mjere
- 11:         nauči odabrani model na  $\mathcal{D}_{\text{train}}$  i ispitaaj na  $\mathcal{D}_{\text{test}}$
- 12:         izračunaj prosjek mjere na  $k$  vanjskih preklopa

- Odabir hiperparametara (redak 4) može biti vođen heurističkim pretraživanjem
- Kao optimalan model odabrati onaj koji je najčešće odabran u  $k$  vanjskih preklopa
- Paziti da se pri učenju modela koristi isključivo informacija iz skupa za učenje