1. ZADATAK

Problem maksimalne margine

Linearni model bez aktivacijske f-je

$$h(x) = \vec{w}^T \phi(\vec{x}) + w_0$$

$$y \in \{-1, +1\}$$ // $y = sgn(h(x))$

pretpostavka: primjeri iz D su linearno odvojivi (ili nakon preslikavanja ϕ u PZ)

znači da $\exists\ \vec{w}$ i w_0 takvi da $h(x^i) \geq 0$ za $y = +1$

te $h(x^i) < 0$ za $y = -1$

tj. vrijedi $\forall (x^{(i)}, y^{(i)}) \in D . \, y^{(i)} h(x^{(i)}) \geq 0$

-> postoji više rj koji zadovoljavaju $y^{(i)}(h(x^i)) \geq 0$, ali nas zanima ono max margine

udaljenost pr. od hiperravnine

$$\frac{y^{(i)} h(x^{(i)})}{\|w\|} = \frac{y^{(i)}(w^T \phi(x^{(i)}) + w_0)}{\|w\|}$$

po def. margina je jednaka udaljenosti do najbližeg pr.

$$\frac{1}{\|\vec{w}\|} \min_i \{y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0)\}$$

-> zanimaju nas $\vec{w}$ i w_0 za koje će margina biti maksimalna

$$\underset{w, w_0}{\operatorname{argmax}} \left\{\frac{1}{\|\vec{w}\|} \min_i \{y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0)\}\right\}$$

Ako je skup primjera D linearno odvojiv, onda postoji i takva margina

=> za $x^{(i)}$ koji je najbliži margini vrijedi

$$y^{(i)}(\vec{w}^T \phi(x^{(i)}) + w_0) = 1$$

onda za sve $(x^{(i)}, y^{(i)}) \in D$ mora vrijediti

$$y^{(i)}(\vec{w}^T \phi(x^{(i)}) + w_0) \geq 1 \quad , n = 1, \dots, N$$

-> opt. problem se svodi na $\underset{w, w_0}{\operatorname{argmax}} \frac{1}{\|w\|}$ uz ograničenja

što je ekvivalentno $\underset{w, w_0}{\operatorname{argmin}} \frac{1}{2}\|w\|^2$

Lagrangeova f-ja

PRIMARNI PROBLEM

$$L(\vec{w}, w_0, \alpha) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{N} \alpha_i \left\{ y^{(i)} \left(\vec{w}^T \phi(x^{(i)}) + w_0 \right) - 1 \right\}$$

b) $\frac{dL}{dw} = 0 \qquad w = \sum_{i=1}^{N} \alpha_i \left(y^{(i)} \phi(x^{(i)}) \right)$

$\frac{dL}{dw_0} = 0 \qquad 0 = \sum_{i=1}^{N} \alpha_i y^{(i)}$

uvrstim u primarni primarni problem

$$L(\alpha) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{N} \alpha_i y^{(i)} \vec{w}^T \phi(x^{(i)}) - \sum_{i=1}^{N} \alpha_i y^{(i)} w_0 + \sum_{i=1}^{N} \alpha_i$$

$$= \frac{1}{2} \sum_{i=1}^{N} \alpha_i y^{(i)} \phi(x^{(i)})^T \cdot \sum_{j=1}^{N} \alpha_j y^{(j)} \phi(x^{(j)})$$

$$- \sum_{i=1}^{N} \alpha_i y^{(i)} \phi(x^{(i)})^T \sum_{j=1}^{N} \alpha_j y^{(j)} \phi(x^{(j)}) + \sum_{i=1}^{N} \alpha_i$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \phi(x^{(i)})^T \cdot \phi(x^{(j)})$$

dualni problem je maksimizirati $\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \phi(x^{(i)})^T \cdot \phi$

uz ograničenja na α: $\alpha_i \geq 0, \; i = 1, \dots, N$

$$\sum_{i=1}^{N} \alpha_i y^{(i)} = 0$$

KKT uvjeti: $\alpha_i \geq 0$

$y^{(i)} h(x^{(i)}) - 1 \geq 0$

$\alpha_i \left(y^{(i)} h(x^{(i)}) - 1 \right) = 0$

c) kod meke granice dozvoljavamo da primjeri budu na pogrešnoj strani, ali to kažnjavamo što su dublje na pogrešnoj strani

uvode se restrevne varijable ξ, $\xi_i \geq 0$, $i = 1, \dots, N$ $\forall$ pr. $x^{(i)} \in D$

→ za primjere na ispravnoj strani $\xi_i = 0$, za ostale $\xi_i = |y^{(i)} - h(x^{(i)})|$

→ pa tako će primjeri unutar margine, ali na ispravnoj strani biti kažnjeni $\Rightarrow$ $0 \leq \xi_i < 1$

→ reformulacija optimizacijskih ograničenja

$$y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0) \geq 1 - \xi_i, \quad i = 1, \dots, N$$

→ cilj: maksimizirati marginu i kazniti pr. na krivoj str. granice

$$\frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^{N} \xi_i$$

, $C > 0$ - kompromis vel. margine i ukupne kazne

$$\underset{w, w_0}{\operatorname{argmin}} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i \right\}$$

...ni meke granice

$$L(w, w_0, \xi, \alpha, \beta) = \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i (y^{(i)} h(x^{(i)}) - 1 + \xi_i) - \sum_{i=1}^{N} \beta_i \xi_i$$

$$\frac{\partial L}{\partial w} \Rightarrow \vec{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \phi(x^{(i)})$$

$$\frac{\partial L}{\partial w_0} \Rightarrow \sum_{i=1}^{N} \alpha_i y^{(i)} = 0$$

$$\frac{\partial L}{\partial \xi_i} \Rightarrow \alpha_i = C - \beta_i$$

...ualna

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \phi(x^{(i)})^T \phi(x^{(j)})$$

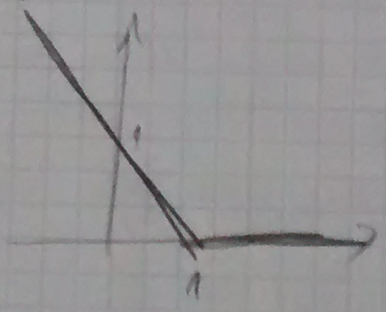
uz ograničenja $\quad 0 \leq \alpha_i \leq C$

$$\sum_{i=1}^{N} \alpha_i y^{(i)} = 0$$

d) f-ja zglobnice

$$L(h(x), y) = max(0, 1 - y h(x))$$

→ f-ja zglobnice je za razliku od logističke f-je jednaka 0 za ispravno klasificirane primjere izvan margine

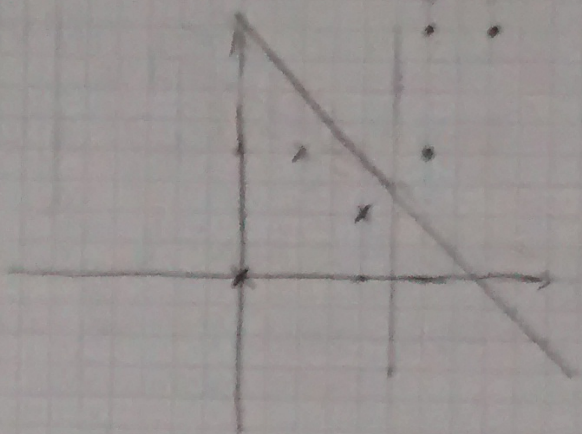


(nedostatak nije konveksna, pa nije pogodna za optimizaciju)

② $D = \{(x^{(i)}, y^{(i)})\}_{i=1}$

$\{((0,0), 1), ((2,4), -1), ((4,2), -1), ((6,4), 1), ((6,?), 1), ((?,?), 1)$

a)



x −1
• 1

$h(x) = w^T \phi(x) + w_0$

$= x_1 + x_2 - 8 \quad / :2 \quad \boxed{\frac{1}{2}x_1 + \frac{1}{2}x_2 - 4}$

b) $d = \frac{2}{\|w\|}$ → ukloniti točka

Širina margine je $2\sqrt{2}$ — do potpornih vektora

c) potporni vektori $x^{(2)}, x^{(3)}$

$\alpha = ?$

$w = \sum_{i=1}^{N} \alpha_i y^{(i)} \phi(x^{(i)})$

$0 = \sum_{i=1}^{N} \alpha_i y^{(i)}$

$$\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \alpha_1 (-1) \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \alpha_2 \, 1 \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

$0 = \alpha_1(-1) + \alpha_2 \, 1 \Rightarrow \alpha_1 = \alpha_2$

$$\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \alpha_1 \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$\alpha_1 = \alpha_2 = \frac{1}{4}$

b) $w_0 = \frac{1}{|S|} \sum_{i \in S} \left(y^{(i)} - \sum_{j \in S} \alpha_j y^{(j)} \phi(x^{(j)})^T \phi(x^{(i)}) \right)$

$w_0 = \frac{1}{2} \left(-1 - \left\{ \frac{1}{4}(-1)[4\ 2]\begin{bmatrix}4\\2\end{bmatrix} + \frac{1}{4}(1)[6\ 4]\begin{bmatrix}6\\4\end{bmatrix} \right\} + 1 - \left\{ \frac{1}{4}(-1)[4\ 2]\begin{bmatrix}4\\2\end{bmatrix} + \frac{1}{4}(1)[6\ 4]\begin{bmatrix}6\\4\end{bmatrix} \right\} \right)$

$= -\frac{1}{4}\left(-\frac{1}{4} \cdot 20 + \frac{1}{2} \cdot 52 - \frac{1}{4} 20 + \frac{1}{4} \cdot 52 \right) =$

$= -\frac{1}{4}\left(-10 + 26 - 10 + 26 \right) = -4$

c) $x^{(1)} = (5, 6)$

dualna reprezentacija

$\alpha_i \geq 0$

$y^{(i)} h(x^{(i)}) - 1 \geq 0$ — ispravno?

$\alpha_i (y^{(i)} h(x^{(i)}) - 1) = 0$ — pravilo

$y^{(1)} (5 + 6 - 6) - 1 \geq 0$

$5 y^{(1)} \geq 1$

$y^{(1)} \geq \frac{1}{5}$

$\frac{1}{2}(y^{(1)} 5 - 1) = 0$

$\frac{5}{2} y^{(1)} = \frac{1}{2}$

$y^{(1)} = \frac{1}{5}$

$y^{(1)} = \mathrm{sgn}\left(\frac{1}{5}\right) = 1$

d) 1.1

$L(h(x), y) = \max\{0,$

③

a) • model ⇒ prostor hipoteza koji alg. pretražuje. Hipoteza je pojedino stablo odluke ⇒ ID3 pretražuje potpun skup hipoteza pa je model skup svih mogućih stabala odluke

• funkcija gubitka ⇒ 0-1

• optimizacijski postupak ⇒ pretraživanje prostora hipoteza metodom "uspona na vrh" kojem je heuristika info INFORMACIJSKA DOBIT

(pristranost pretraživanja ⇒ preferiraju se kraća stabla, ona koja atribute s većom info dobiti dodaje bliže korijenu

b) "Dobro ljetovanje na Jadranu"

$y^{(i)} \in \{da, ne, možda\}$

• odrediti atribut s najvećom info dobiti

[entropija] $E(D) = -\sum_{k=1}^{K} p_k \log_2 p_k = -\left(\frac{6}{10}\log_2\frac{6}{10} + \frac{4}{10}\log_2\frac{4}{10} + 0\log_2 0\right)$

$= 0.971$

[info dobit] $IG(D,A) = E(D) - \sum_{v \in \text{vrijednosti } A} \frac{|D_v|}{|D|} E(D_v)$

I Mjesto → Istra ×4 (da×4) E(Istra) = 0
Kvarner ×3 (ne×3) E(Kvarner) = 0
Dalmacija ×3 (da×2 ne×1) E(Dalm.) = 0.92

$IG(D, Mjesto) = 0.971 - \frac{4}{10}\cdot 0 - \frac{3}{10}\cdot 0 - \frac{3}{10}\cdot 0.92 = \boxed{[illegible]}$

II Otok → da ×3 (da×3) E(da) = 0
ne ×7 (da×3 ne×4) E(ne) = 0.985

$IG(D, Otok) = 0.971 - \frac{3}{10}\cdot 0 - \frac{7}{10}\cdot 0.985 = \boxed{0.2815}$

III Smještaj ⇒ privatni ×5 (da×4 ne×1) E(privatni) = [illegible]
hotel ×4 (da×2 ne×2) E(hotel) = 1
[illegible] ×1 (ne×1) E([illegible]) = 0

$IG(D, Smještaj) = 0.971 - \frac{5}{10}\cdot 0.72 - \frac{4}{10}\cdot 1 = \boxed{0.211}$

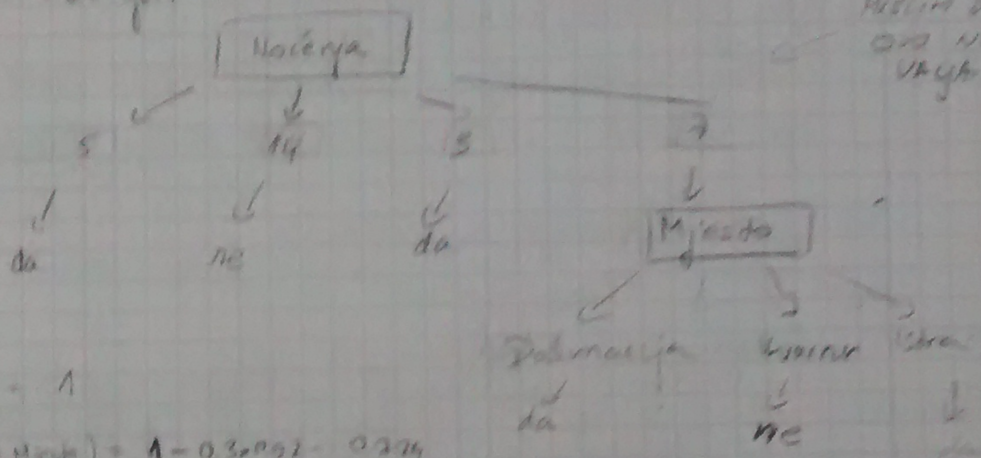
II Noćenja 5 × 3 (da × 3) E(5) = 0
14 × 3 (da × 0 ne × 3) E(14) = 0
3 × 2 (da × 2) E(3) = 0
7 × 2 (da × 1 ne × 1) E(7) = 1

IG(D, Noćenja) = 0.971 − $\frac{2}{10}$ · 1 = 0.771

I Prijevoz auto × 6 (da × 4 ne × 2) E(auto) = 0.92
avion × 3 (da × 2 ne × 1) E(avion) = 0.92
autobus × 1 (ne × 1) E(autobus) = 0

IG(D, Prijevoz) = 0.971 − $\frac{6}{10}$ · 0.92 − $\frac{3}{10}$ · 0.92 = 0.143

Noćenja → korijen



Mislim da ovo ne valja

E(7) = 1

IG(7, Mjesto) = 1 − 0.3 · 0.92 = 0.724

IG(7, Oblik) = 1 − 0.7 · 0.985 = 0.3105

IG(7, Smještaj) = 1 − 0.5 · 0.72 − 0.5 · 0.24

IG(7, Prijevoz) = 1 − 0.6 · 0.92 − 0.3 · 0.92 = 0.172

E(Istra)

c) Kada bi k primjera nedostajala vrijednost značajke Smještaj → dodala bi vrijednost koja se najčešće pojavljuje u D

(5) a) k-NN

$D = \{((4,4,3),1), ((4,3,4),1), ((4,0,2),1), ((5,3,3),0),$
$((5,4,1),0), ((3,2,0),0)\}$

$x^{(1)} = (4,2,1) \to ?$ $x^{(2)} = (0,3,3) \to ?$

Euklidska udaljenost, težinski faktor obrnuto proporcionalan kvadratu udaljenosti

$$h(x_q) = \underset{v \in \{0,1\}}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(v, y^{(i)})$$

Euklidska $d(x^a, y^b) = \sqrt{\sum_{i=1}^{n} (x_i^a - x_i^b)^2}$

Težinski faktor $w_i = \dfrac{1}{d(x_i, x_q)^2}$

$$h(x_q) = \frac{\sum_{i=1}^{k} w_i y^{(i)}}{\sum_{i=1}^{k} w_i}$$

5 $d((4,4,0),(4,2,1)) = \sqrt{5}$
1 $d((4,3,1),(4,2,1)) = 1$
$d((4,0,2),(4,2,1)) = 3$
2 $d((5,3,3),(4,2,1)) = \sqrt{2}$
3 $d((5,4,1),(4,2,1)) = \sqrt{5}$
$d((3,2,0),(4,2,1)) = \sqrt{10}$

$d((4,4,0),(0,3,3)) = \sqrt{26}$
$d((4,3,1),(0,3,3)) = 2\sqrt{5}$
$d((4,0,2),(0,3,3)) = \sqrt{46}$
2 $d((5,3,3),(0,3,3)) = 3\sqrt{3}$
6 $d((5,4,1),(0,3,3)) = \sqrt{30}$
$d((3,2,0),(0,3,3)) = \sqrt{19}$

1,20
5,94

(4,2,1)

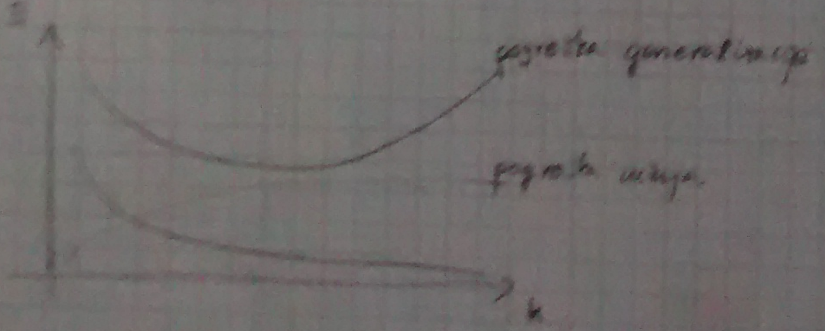
Razred 1: $\frac{1}{5} + 1 = \frac{6}{5}$ $\frac{6}{5} > 1 \Rightarrow$ (4,2,1) pripada 1

Razred 0: $\frac{1}{2} + \frac{1}{2} = 1$

(0,3,3)

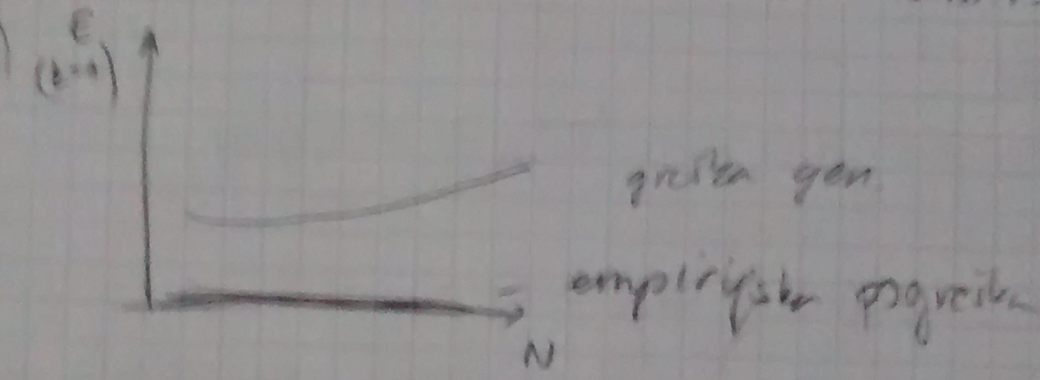
Razred 1: $\frac{1}{20} + \frac{1}{26} = 0{,}0885$
Razred 0: $\frac{1}{30} + \frac{1}{27} = 0{,}067$ $\to$ (0,3,3) pripada 1

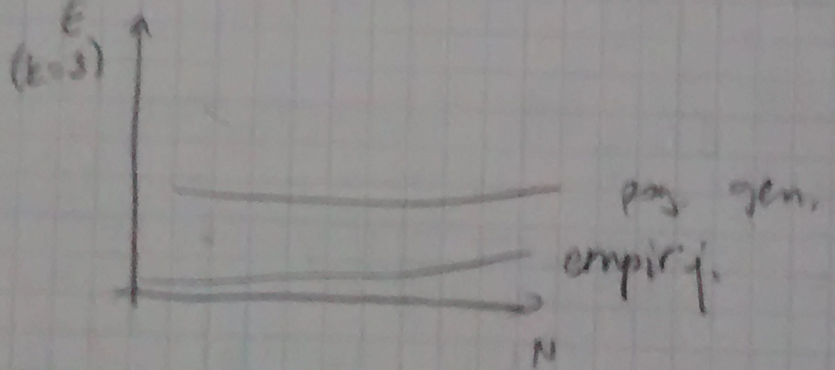
b)



IV MATRICA 5 × 3 (do ×3) → $E(S)=0$

a)





b)

(5) a) $\{(y^{(i)}, h(x^{(i)})\}_i =$

$\{(1,1), (0,2), (2,2), (1,2), (1,1), (0,0), (1,1), (2,1), (0,1), (2,0), (2,1)\}$

suma nedijagonalnih el. u retku

$N = 11 \leftarrow T0 + F0 + T1 + F1 + T2 + F2$

$T0 = 1$, $F0 = 1$, $T1 = 3$, $F1 = 3$, $T2 = 1$, $F2 = 2$

$N0 = 2$, $N1 = 1$, $N2 = 3$

$\downarrow$ suma nedij. el. u stupcu

		stvarno 0	stvarno 1	stvarno 2
pred.	0	1	0	1
	1	1	3	2
	2	1	1	1

preciznost = udio točno klasificiranih primjera u skupu poz. klasif. primjera

$$P0 = \frac{T0}{T0+F0} = \frac{1}{2} = 0.5$$

$$P1 = \frac{T1}{T1+F1} = \frac{3}{3+3} = \frac{1}{2}$$

$$P2 = \frac{T2}{T2+F2} = \frac{1}{1+2} = \frac{1}{3}$$

odziv = udio točno klasificiranih primjera u skupu svih pozitivnih primjera

$$R0 = \frac{T0}{T0+N0} = \frac{1}{1+2} = \frac{1}{3}$$

$$R1 = \frac{T1}{T1+N1} = \frac{3}{3+1} = \frac{3}{4}$$

$$R2 = \frac{T2}{T2+N2} = \frac{1}{1+3} = \frac{1}{4}$$

F_1 mjera - harmonijska sredina preciznosti i odziva

$$F_\beta = \frac{(1+\beta^2)PR}{\beta^2 P + R} \qquad F_j = \frac{2P_jR_j}{P_j+R_j}$$

$TP = \sum TP_j$

$FP = \sum FP_j$

$FN = \sum FN_j$

$$F^{macro} = \frac{1}{k}\sum_{j=1}^{k} F_j \qquad F^{micro} = \frac{2PR}{P+R} = \frac{2TP}{2TP+FP+FN}$$

$$F_0 = \frac{2 \cdot P_0 R_0}{P_0 + R_0} = \frac{2 \cdot \frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}+\frac{1}{3}} = 0.4$$

$$F_1 = \frac{2 \cdot P_1 R_1}{P_1 + R_1} = \frac{2 \cdot \frac{1}{2} \cdot \frac{3}{4}}{\frac{1}{2}+\frac{3}{4}} = 0.6$$

$$F_2 = \frac{2 \cdot \frac{1}{3} \cdot \frac{1}{4}}{\frac{1}{3}+\frac{1}{4}} = \frac{2}{7}$$

$$F_{1,macro} = \frac{1}{3}\left(0.4 + 0.6 + \frac{2}{7}\right) = \frac{3}{7} = 0.423$$

$TP = T0 + T1 + T2 = 5$

$FP = F0 + F1 + F2 = 6$

$FN = 6$

$$F_1^{micro} = \frac{2 \cdot 5}{2 \cdot 5 + 6 + 6} = \frac{10}{22} = \frac{5}{11} = 0.45$$

b) $N = 1000$

ugniježđena unakrsna provjera 5×5

$25 \cdot$ kombinacija hiperparametara

25/25

$(5 \cdot K \cdot 5 +$

~~5 puta~~ (jedna iteracija 5)

koliko puta provesti učenje modela?

koliko primjera u svakoj iteraciji?

1) za učenje? $\frac{4}{5} \cdot \frac{4}{5} \cdot 1000 = 640$

2) za provjeru? $\frac{4}{5} \cdot \frac{1}{5} \cdot 1000 = 160$

3) za ispitivanje? $\frac{1}{5} \cdot 1000 = 200$