Strojno učenje – domaća zadaća 6

UNIZG FER, ak. god. 2016./2017.

Zadano: 15. 11. 2016. Rok: 18. 11. 2016.

Napomena: Zadatke možete rješavati samostalno ili u grupi. Ako zadatke rješavate u grupi, pobrinite se da svi članovi grupe pridonose rješenju i da ga naposlijetku 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: Znati definirati osnovne jezgrene funkcije. Znati definirati jezgreni stroj i razumjeti razliku između jezgrenog stroja i rijetkog jezgrenog stroja.***]

   (a) Definirajte jezgrenu funkciju, RBF-jezgru i Gaussovu jezgru.

   (b) Definirajte Mahalanobisovu udaljenost i RBF-jezgru koja koristi tu udaljenost. Navedite primjer u kojemu biste koristili tu jezgru umjesto Gaussove jezgre.

   (c) Definirajte jezgreni stroj i rijetki jezgreni (vektorski) stroj. Koji od njih je parametarski a koji neparametarski algoritam i što to znači?

2. [***Svrha: Isprobati preslikavanje primjera u prostor značajki primjenom Gaussovih baznih funkcija. Razumjeti kako preslikavanje utječe na broj parametara i hiperparamera modela.***] Raspolažemo skupom primjera:

$$\mathcal{D} = \big\{(\mathbf{x}^{(i)}, y^{(i)})\big\}_{i=1}^{5} = \big\{\big((-1,-1),0\big), \big((0,0),0\big), \big((3,-3),1\big), \big((-2,1),1\big), \big((-4,2),1\big)\big\}\,.$$

   (a) U ulaznome prostoru skicirajte diskriminacijsku granicu $h(\mathbf{x}) = 0$ koju biste dobili logističkom regresijom uz $\boldsymbol{\phi}(\mathbf{x}) = (1, \mathbf{x})$, tj. bez preslikavanja (izračun nije potreban).

   (b) Na isti skup primjera primijenite jezgreni stroj s baznim funkcijama:

$$\phi_j(\mathbf{x}) = \exp\Big(-\frac{\|\mathbf{x}-\boldsymbol{\mu}_j\|^2}{2\sigma_j^2}\Big)\,.$$

   Konkretno, koristite dvije bazne funkcije s parametrima $\boldsymbol{\mu}_1 = (0,0)$, $\boldsymbol{\mu}_2 = (-3,3)$, $\sigma_1 = \sigma_2 = 1$. Skicirajte primjere u prostoru značajki (dimenzije ϕ_1 i ϕ_2) i granicu koju biste dobili linearnom regresijom (izračun nije potreban).

   (c) Koliko ovaj jezgreni stroj ima parametara a koliko hiperparametara? Kako biste u praksi odredili vrijednosti hiperparametara modela? Utječu li u ovom slučaju hiperparametri na složenost modela? Obrazložite odgovor.

3. [***Svrha: Razumjeti jezgreni trik kod SVM-a.***]

   (a) Koristimo polinomijalnu jezgrenu funkciju $\kappa(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^{\mathrm{T}}\mathbf{z} + 1)^2$. Pokažite da je za $n = 2$ jezgra κ Mercerova jezgra. Zašto je to bitno?

(b) Odredite vektor $\boldsymbol{\phi}(\mathbf{x})$ u koji će efektivno biti preslikan primjer $\mathbf{x} = (2,3)$ primjenom jezgre $\kappa(\mathbf{x},\mathbf{z}) = (\mathbf{x}^{\mathrm{T}}\mathbf{z} + 1)^2$.

(c) Veza između primarnih i dualnih parametara SVM-a jest $\mathbf{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \boldsymbol{\phi}(\mathbf{x}^{(i)})$. Na skupu za učenje trenirali smo SVM s polinomijalnom jezgrenom funkcijom, $\kappa(\mathbf{x},\mathbf{z}) = (\mathbf{x}^{\mathrm{T}}\mathbf{z} + 1)^3$. Potporni vektori su $\mathbf{x}^{(1)} = (-2,3,5)$, $\mathbf{x}^{(2)} = (6,4,3)$ i $\mathbf{x}^{(3)} = (8,8,2)$. Prvi primjer je negativan, a druga dva su pozitivna. Lagrangeovi koeficijenti su $\alpha_1 = 0.131$, $\alpha_2 = 0.048$ i $\alpha_3 = 0.013$. Pomak je $w_0 = -0.51$. Iskoristite jezgreni trik te odredite klasifikaciju primjera $\mathbf{x}^{(4)} = (1,2,3)$.

4. [***Svrha: Razumjeti karakteristike Gaussove jezgre.***]

(a) Primjenom operacija za izgradnju složenijih Mercerovih jezgri iz jednostavnijih Mercerovih jezgri, dokažite da je Gaussova jezgra Mercerova jezgra. (Pomoć: raspišite izraz $\|\mathbf{x} - \mathbf{x}'\|^2$.)

(b) Kako parametar $\gamma = 1$ Gaussove jezgre utječe na složenost modela? Koji je odnos između parametra C i γ?

(c) Koristimo Gaussovu jezgru uz $\gamma = 1$. Možemo li u ovom slučaju odrediti u koji vektor $\boldsymbol{\phi}(\mathbf{x})$ u prostoru značajki će biti preslikan primjer $\mathbf{x}$? Možemo li odrediti težine $\mathbf{w}$. Zašto?

(d) Jamči li primjena Gaussove jezgre (1) empirijsku pogrešku jednaku nuli na skupu za učenje? (2) savršenu linearnu odvojivost primjera za učenje u prostoru značajki? (3) minimalnu pogrešku na ispitnom skupu? (4) preslikavanje primjera u beskonačan prostor značajki?

5. [***Svrha: Isprobati klasifikator k-nn na konkretnom primjeru. Razumjeti kako parametri k i broj primjera N utječu na složenost modela.***]

(a) Klasifikator 4-NN s euklidskom udaljenošću učen je na sljedećim primjerima iz $\mathbb{R}^3 \times \{0,1\}$:

$$\mathcal{D} = \big\{\big((\mathbf{x}^{(i)}, y^{(i)}\big)\big\}_{i=1}^{6} =$$
$$\big\{\big((4,4,0),1\big),\ \big((4,3,1),1\big),\ \big((6,0,2),1\big),\ \big((5,2,2),0\big),\ \big((5,1,1),0\big),\ \big((7,2,0),0\big)\big\}.$$

Odredite klasifikaciju primjera $\mathbf{x}^{(1)} = (4,2,1)$ i $\mathbf{x}^{(2)} = (0,3,3)$.

(b) Ponovite klasifikaciju s težinskim modelom 4-NN, primjenom inverzne kvadratne jezgre.

(c) Skicirajte (za općenit slučaj) pogrešku učenja i pogrešku generalizacije kao funkciju od k.

(d) Skicirajte (za općenit slučaj) pogrešku učenja i pogrešku generalizacije kao funkciju broja primjera N za $k = 1$ i $k = 3$ (nacrtajte dva zasebna grafikona).

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 metode.

⑥ DOMAĆA ZADAĆA

① a) DEF. JEZGRENU Fju, RBF - jezgru i GAUSSOVU JEZGRU

- JEZGRENA Fja: $\mathcal{K}: \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$
- RBF - jezgra: $\mathcal{K}(x, x') = \phi(\|\vec{x} - \vec{x}'\|)$
- GAUSSOVA JEZGRA: $\mathcal{K}(x, x') = \exp(-\gamma \|\vec{x} - \vec{x}'\|^2)$ gdje je $\gamma = 1/2\sigma^2$ PRECIZNOST

b) DEF. MAHALANOBISOVU UDALJENOST I RBF JEZGRU KOJA KORISTI TU UDALJ. NAVEDI PR. KADA BI TU JEZGRU KORISTIO UMJESTO GAUSSOVE.

$$\boxed{D_{MAX} = \sqrt{(x - x')^T \Sigma^{-1} (x - x')}}$$

$$\boxed{\mathcal{K}(x, x') = \exp\left(-\frac{1}{2} (x - x')^T \Sigma^{-1} (x - x')\right)}$$

Σ - KOVARIJANCIJSKA MATRICA ZNAČAJKI

RBF jezgru s Mahalanobisovom udaljenošću bolje je koristiti od Gaussove jezgre u slučajevima kada:

- → Kada su ZNAČAJKE RAZLIČITIH REDOVA VELIČINE ili RAZL. MJ. JEDINICE
- → Kada imamo VISOKU KORELACIJU IZMEĐU NEKIH ZNAČAJKI.

c) DEF. JEZGRENI STROJ I RIJETKI JEZGRENI STROJ. KOJI OD NJIH JE PARAM. A KOJI NEPARAM. I ŠTO TO ZNAČI?

→ PARAMETARSKI ALG. su algoritmi kod kojih BROJ PARAM. NE OVISI O BROJU PRIMJERA N!

→ NEPARAMETARSKI ALG. su alg. kod kojih BR. PARAM. OVISI O BR. PR.

→ JEZGRENI STROJ - Poopćeni linearni model s preslikavanjem ϕ koje za bazne fje ϕ_j koristi jezgrene funkcije:

$$\boxed{\phi(x) = (\mathcal{K}(x, \mu_1), \mathcal{K}(x, \mu_2), \ldots, \mathcal{K}(x, \mu_m))}$$

gdje su μ_j - centroidi u prostoru primjera.

→ RIJETKI JEZGRENI STROJEVI - umjesto centroida koriste primjere za učenje

$$\boxed{\phi(x) = \mathcal{K}(x, x_1), \mathcal{K}(x, x_2), \ldots, \mathcal{K}(x, x_N))}$$

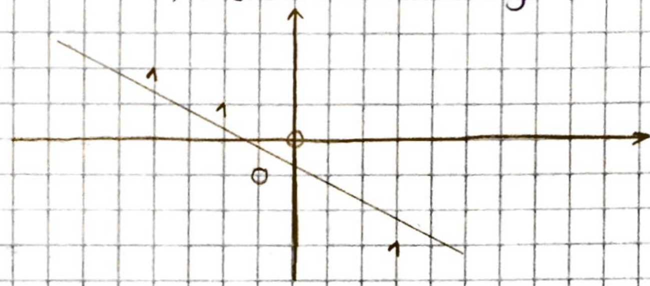
Provjeravaju sličnost ispitnog pr. s pr. za učenje

→ Iz prethodnih jednadžbi možemo primjetiti kako Rjetki jezgreni strojevi ovise o broju primjera N! Dok jezgreni strojevi ne ovise o N.

↳ Iz toga zaključujemo kako JEZGRENI STROJEVI padaju pod PARAMETARSKE ALG., dok RIJETKI JEZGRENI STROJEVI pod NEPARAMETARSKE ALG.

② $D=\{(x^{(i)},y^{(i)})\}_{i=1}^{5} = \{((-1,-1),0),((0,0),0),((3,-3),1),((-2,1),1),((-4,2),1)\}$

a) U UL. PROSTORU SKIC. DISKRIM. GRANICU $h(x)=0$ KOJU BI DOBILI LOG. REG. UZ $\Phi(x) = (1,x)$ (BEZ PRESLIKAVANJA)



b) NA ISTI SKUP PRIMJENI $\phi_j = \exp\left(-\frac{\|x-\mu_j\|^2}{2\sigma_j^2}\right)$ $\mu_1=(0,0)$ $\mu_2=(-3,3)$
$\sigma_1=\sigma_2=1$

$\phi_1 = \exp\left(-\frac{1}{2}\|x-\begin{bmatrix}0\\0\end{bmatrix}\|^2\right)$ $\qquad$ $\phi_2 = \exp\left(-\frac{1}{2}\|x-\begin{bmatrix}-3\\3\end{bmatrix}\|^2\right)$

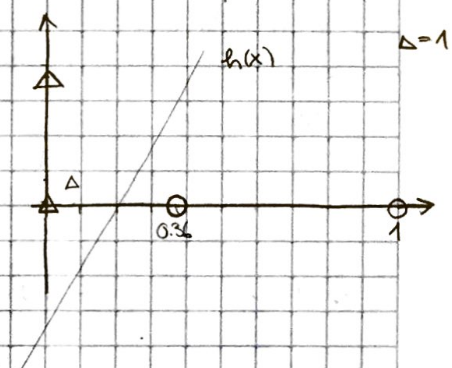
$\phi_1(x_1) = e^{-1}$ $\qquad$ $\phi_2(x_1) = e^{-10}$ $\qquad$ $X_1 = (e^{-1}, e^{-10})$ (0)

$\phi_1(x_2) = e^{0} = 1$ $\qquad$ $\phi_2(x_2) = e^{-9}$ $\qquad$ $X_2 = (e^{0}, e^{-9})$ (0)

$\phi_1(x_3) = e^{-9}$ $\qquad$ $\phi_2(x_3) = e^{-36}$ $\qquad$ $X_3 = (e^{-9}, e^{-36})$ (1)

$\phi_1(x_4) = e^{-5/2}$ $\qquad$ $\phi_2(x_4) = e^{-5/2}$ $\qquad$ $X_4 = (e^{-5/2}, e^{-5/2})$ (1)

$\phi_1(x_5) = e^{-10}$ $\qquad$ $\phi_2(x_5) = e^{-1}$ $\qquad$ $X_5 = (e^{-10}, e^{-1})$ (1)



c) W=? hiperparam? U PRAKSI ODREĐIVANJE VR. HIPERPARAM? UTJECAJ U U OVOM SLUČ HIPERPARAM NA SLOŽ. MODELA?

$W=3 \Rightarrow (w_0,w_1,w_2)$

$\text{Hip}=6 \rightarrow \left.\begin{matrix}\mu_1=(\mu_{11},\mu_{12}) \rightarrow 2\\ \mu_2=(\mu_{21},\mu_{22}) \rightarrow 2\\ \sigma_1,\sigma_2 \rightarrow 2\end{matrix}\right\} 6$

- Unakrsnom provjerom
- μ ne, ali σ da. $\sigma = 1/\gamma$, $\sigma\uparrow$ slož$\downarrow$

③ a) $K(\vec{x},\vec{z}) = (\vec{x}^T\vec{z}+1)^2$ → Polinomijalna jezgrena fja
POKAŽI DA JE ZA $n=2$ K MERCEROVA JEZGRA. Zašto je to bitno?

$K(\vec{x},\vec{z}) = (x^Tz+1)^2 = (x_1z_1+x_2z_2+1)^2 = (x_1z_1+x_2z_2+1)(x_1z_1+x_2z_2+1) =$
$= (x_1z_1)^2 + x_1x_2z_1z_2 + x_1z_1 + x_1x_2z_1z_2 + (x_2z_2)^2 + x_2z_2 + x_1z_1 + x_2z_2 + 1 =$
$= (x_1z_1)^2 + (x_2z_2)^2 + 1 + 2x_1z_1x_2z_2 + 2x_1z_1 + 2x_2z_2 =$
$= (x_1z_1)^2 + (x_2z_2)^2 + 1 + \sqrt{2}x_1z_1\sqrt{2}x_2z_2 + \sqrt{2}x_1\sqrt{2}z_1 + \sqrt{2}x_2\sqrt{2}z_2 =$
$= (x_1^2, x_2^2, 1, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2)^T (z_1^2, z_2^2, 1, \sqrt{2}z_1z_2, \sqrt{2}z_1, \sqrt{2}z_2)$
$= \Phi(x)^T\Phi(z)$ //

↳ $K(x,z)$ ODGOVARA SKALARNOM UMNOŠKU U PROSTORU ZNAČAJKI ŠTO POKAZUJE DA JE $K(x,z)$ MERCEROVA JEZGRA: $K(x,z) = \Phi(x)^T\Phi(z)$. TO JE BITNO IZ RAZLOGA ŠTO JEZGRENI TRIK KOD SVM-A NEĆE FUNKCIONIRATI AKO $K(x,z)$ NIJE MERCEROVA JEZGRA.

b) $\Phi(x) = ?$ u koji će efektivno biti preslikan $x = (2,3)$ primjenom jezgre $K(x,z) = (x^T z + 1)^2$

$$\Phi(x) = (x_1^2, x_2^2, 1, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2) = (4, 9, 1, 6\sqrt{2}, 2\sqrt{2}, 3\sqrt{2})$$

c) $\vec{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \Phi(x^{(i)})$; $K = (x^T z + 1)^3$

POTP. VEKT. → $x^{(1)} = (-2,3,5)$ $x^{(2)} = (6,4,3)$ $x^{(3)} = (8,8,2)$ $y = (-1,1,1)$

$\alpha_1 = 0.131$ $\alpha_2 = 0.048$, $\alpha_3 = 0.013$ $w_0 = -0.51$

KLASIFIKAC. PR. $x^{(4)} = (1,2,3) = ?$

$$h(x) = w^T \Phi(x) + w_0 = \sum \alpha_i y^{(i)} \underbrace{\Phi^T(x^{(i)}) \Phi(x)}_{K(x, x^{(i)})} + w_0 = \sum_{i=1}^{N} \alpha_i y^{(i)} K(x, x^{(i)}) + w_0$$

$$h(x^{(4)}) = \alpha_1 y^{(1)} K(x^{(4)}, x^{(1)}) + \alpha_2 y^{(2)} K(x^{(4)}, x^{(2)}) + \alpha_3 y^{(3)} K(x^{(4)}, x^{(3)}) + w_0$$
$$= -(0.131)(8000) + (0.048)(13824) + (0.013)(29791) - 0.51$$
$$= -1048 + 663.552 + 387.283 - 0.51$$
$$h(x^{(4)}) = 2.325$$

$$y(x^{(4)}) = \mathrm{sgn}(h(x^{(4)})) = +1$$

4 a) DOKAZ DA JE GAUSSOVA JEZGRA MERCEROVA JEZGRA

GAUSSOVA JEZGRA ⇒ $K(x,x') = \exp(-\gamma \|x - x'\|^2) = \exp(-\gamma(x^2 - 2xx' + x'^2)) =$

$$= \underbrace{\exp(-\gamma)}_{\text{SKALAR}} \underbrace{\exp(x^2)\, \exp(-2xx')\, \exp(x'^2)}_{f(x)\, K_1(x,x')\, f(x')}$$

→ f → bilo koja f-ja

↳ Potrebno je provjeriti da li je K_1 mercerova jezgra

$K_1(x,x') = -2xx' = -\sqrt{2}x\sqrt{2}x' = -\Phi^T(x)\Phi(x')$ → MERCEROVA JEZGRA //

b) UTJECAJ $\gamma = 1$ GAUSSOVE JEZGRE NA SLOŽENOST MODELA? ODNOS C i γ = ?

⇒ $\gamma \uparrow$ → $\sigma^2 \downarrow$ (varijanca Gaussove f) → "UŽA" Gaussova f-ja → Primjeri efektivno RAZLIČITIJI →
→ SKALARNI PRODUKT PRIMJERA MANJI → $K(x,x') \to 0$ (osim za $x = x'$) →
→ PRENAUČENOST! → SLOŽENOST MODELA ↑

↳ $\gamma \uparrow$ složenost ↑

→ C↑ kazna ↑ složenost ↑

⇒ POVEZANOST, ODNOS, C i γ:

↳ Ako odaberemo visoku preciznost, $\gamma \uparrow$, model će postati složeniji pa trebamo pojačati regularizaciju odabirom manjeg C, C↓ (α 1/λ)

c) $\gamma = 1$ MOŽEMO LI ODREDITI U KOJI $\Phi(x)$ ĆE SE PRESLIKATI X? MOŽEMO LI ODREDITI TEŽINE w? ZAŠTO?

Ne! $\Phi(x)$ definiran je samo implicitno, s obzirom da Gaussova jezgra definira $K(x, x')$

↳ S obzirom da ne možemo odrediti $\Phi(x)$ ne možemo odrediti ni $\vec{w}$ jer su def. izrazom:

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

d)

i) Primjena Gaussove jezgre NE GARANTIRA empirijsku pogr. $=\emptyset$ na skupu za učenje.

ii) -||- NE GARANTIRA savršenu lin. odvojivost pr. za učenje u prostoru značajki

iii) -||- NE GARANTIRA minimalnu pogrešku na ispitnom skupu

iv) -||- GARANTIRA preslikavanje primjera u ∞ prostor značajki

5 a) $D = \{((4,4,0),1), ((4,3,1),1), ((6,0,2),1), ((5,2,2),0), ((5,1,1),0), ((7,2,0),0)\}$
$\underline{K=4}$

$\underbrace{X^{(1)} = (4,2,1) \quad X^{(2)} = (0,3,3)}$ } odredi klasifikaciju

Računamo euklidsku udaljenost novih pr. od pr. iz D, gledamo 4 najbliža susjeda i na temelju njihove klasifikacije klasificiramo nove primjere → na temelju VEĆINSKE OZNAKE K NAJBLIŽIH SUSJEDA

↳ $\boxed{h(x) = \underset{j \in \{0,1\}}{\mathrm{argmax}} \sum 1\{j = y^{(i)}\}}$ $\boxed{d = \|X - X^{(i)}\|^2}$

$X^{(1)}$: $d_{1,1} = 5$ $d_{1,2} = 1$ $d_{1,3} = 9$ $d_{1,4} = 2$ $d_{1,5} = 2$ $d_{16} = 10$

↳ najbliži susjedi X_1, X_2, X_4, X_5 s oznakama $y = [1 \ 1 \ 0 \ 0]$

Ne možemo eksplicitno odrediti klasifikaciju, potreban nam je neki dodatan kriterij/uvjet

$X^{(2)}$: $d_{21} = 26$ $d_{22} = 20$ $d_{23} = 46$ $d_{24} = 27$ $d_{25} = 33$ $d_{26} = 59$

↳ 4-NN: X_1, X_2, X_4, X_5 s oznakama $y = [1 \ 1 \ 0 \ 0]$ -||-

b) PONOVI KLASIFIKACIJU S TEŽINSKIM 4-NN PRIMJENOM INVERZNE KVADRATNE JEZGRE

$\boxed{h(x) = \underset{j \in \{0,1\}}{\mathrm{argmax}} \sum_{i=1}^{N} \mathcal{K}(x^{(i)}, x)\, 1\{j = y^{(i)}\}}$ $\boxed{\mathcal{K}(x^{(i)}, x) = \frac{1}{1 + \|x - x'\|^2}}$

$$h(x) = \mathrm{argmax}\left\{\underbrace{\mathcal{K}(x^{(1)}, x_1) + \mathcal{K}(x^{(1)}, x_2)}_{\text{KLASA 1}}, \underbrace{\mathcal{K}(x^{(1)}, x_4) + \mathcal{K}(x^{(1)}, x_5)}_{\text{KLASA 0}}\right\} =$$

$$= \mathrm{argmax}\left\{(1+d_{11})^{-1} + (1+d_{12})^{-1}, (1+d_{14})^{-1} + (1+d_{15})^{-1}\right\} =$$

$h(x^1) = \mathrm{argmax}\{0.67, 0.67\}$ → Potrebno definirati neki dodatni kriterij

$h(x^2) = \mathrm{argmax}\{0.085, 0.065\}$ → $h(x^{(2)}) = 1$

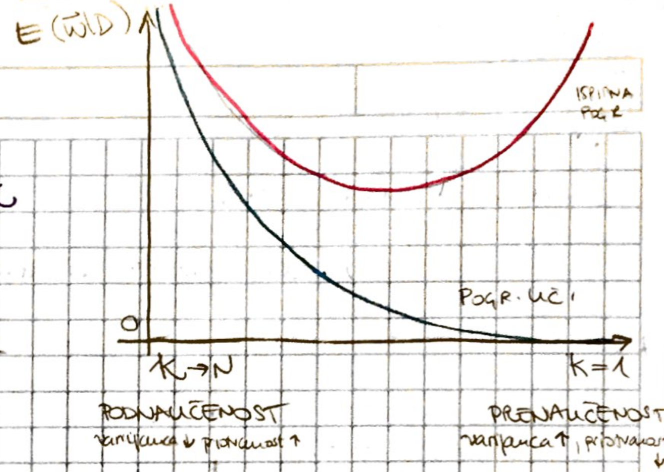
c) POGR. UČ/GEN KAO F-JA OD K

!!!

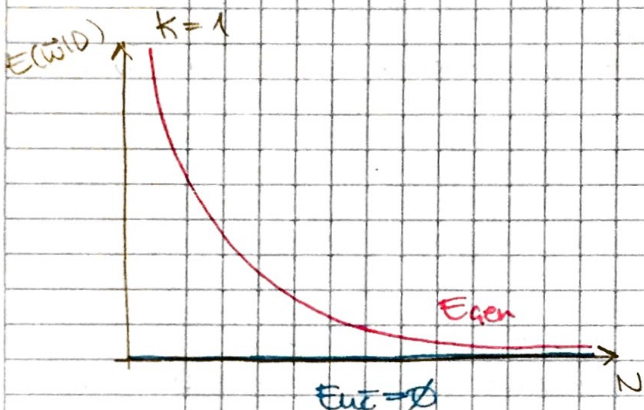
k=1 $E_{uč} = \emptyset$!

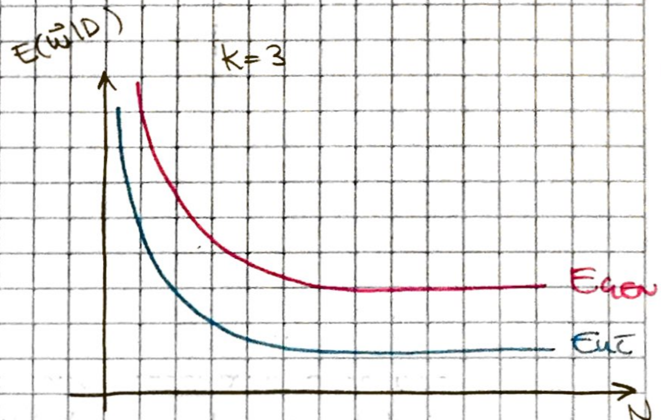
k↑ $E_{uč}$ ↑ !

s k=N se gubi mogućnost klasifikacije u bilo koju klasu osim najbrojnije



d) POGR UČ/GEN ZA k=1 i k=3 KAO F-JA OD N





6) ODNOSI IZMEĐU i) SLOŽENOSTI MODELA, ii) BR. PARAM. MODELA, iii) DIM. UL. PROSTORA iv) BR. PR. N. (ZASEBNO PARAM. M., NEPARAM. M.)

PARAMETARSKI MODELI:
- Broj parametara modela ($\vec{w}$) raste s dimenzijom ulaznog prostora (n) (odnosno s dimenzijom prostora značajki), ali broj parametara modela ne ovisi o broju primjera za učenje N!
- Složenost modela raste s povećanjem broja parametara, a ne ovisi o broju primjera.

PARAMETRI w-ovi
$\vec{w} = [w_0, w_1, \dots, w_n]$
n-parametara = dimenzija

NEPARAMETARSKI MODELI:
- Broj parametara modela raste s brojem primjera za učenje N, a ne ovisi o dimenziji ulaznog prostora n.
- Složenost modela (ovisi o) raste s brojem primjera za učenje, a ne ovisi o dimenziji ulaznog prostora.