Strojno učenje – domaća zadaća 2

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

Zadano: 17. 10. 2016. Rok: 21. 10. 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: Razumjeti matrično rješenje za neregulariziranu i regulariziranu regresiju i izvježbati potrebnu matematku. Razumjeti kako loša kondicija matrice utječe na stabilnost rješenja i kaok regularizacija to popravlja.***]

   (a) Izvedite u matričnom obliku rješenje za vektor $\mathbf{w}$ za neregularizirani linearni model regresije uz kvadratnu funkciju gubitka.

   (b) Izvedite to rješenje za L2-regularizirani model.

   (c) Raspolažemo sljedećim skupom primjera za učenje:

   $$\mathcal{D} = \left\{(x^{(i)}, y^{(i)})\right\}_{i=1}^{4} = \left\{(0,4),(1,1),(2,2),(4,5)\right\}.$$

   Podatke želimo modelirati polinomijalnom regresijskom funkcijom $h(x) = w_0 + w_1x + w_2x^2$, uz regularizacijski faktor $\lambda = 10$. Napišite kako bi u ovome konkretnom slučaju izgleda jednadžba iz zadatka (b) (Ne morate ju izračunavati, samo ju napišite.)

   (d) Jednadžba iz zadatka (b) daje rješenje u zatvorenoj formi, međutim njezin izračun ponekad može biti računalno zahtjevan. Što konkretno predstavlja problem? Je li problem izražen kada imamo mnogo primjera za učenje ili kada imamo mnogo značajki? Obrazložite odgovor.

   (e) Rješenje jednadžbe iz zadatka (a) može biti numerički nestabilno. Što to znači i kada će to biti slučaj? Postoji li takav problem u slučaju formulacije problema iz zadatka (b)? Obrazložite.

2. [***Svrha: Shvatiti kako se nelinarna funkcija u ulaznom prostoru funkcija preslikava u linearnu funkciju odnosno (hiper)ravninu u prostoru značajki.***]

   (a) Regresijom želimo aproksimirati funkciju jedne varijable $y = 3 \cdot (x-2)^2 + 1$. Skicirajte tu funkciju u ulaznome prostoru. Definirajte linearan model $h(x)$ uz funkciju preslikavanja $\boldsymbol{\phi}(x) = (1, x, x^2)$. Odredite vektor težina $\mathbf{w} = (w_0, w_1, w_2)$ tog modela.

   (b) Skicirajte u prostoru (x_1, x_2) ~~(dakle u prostoru značajki, pri čemu zanemarite težinu w_0)~~ izokonture funkcije y. Naznačite u tom prostoru točke u koje se preslikavaju primjeri $x^{(1)} = 1$, $x^{(2)} = 2$ i $x^{(3)} = 3$. Koja je vrijednost od $h(x)$ za navedene primjere?

3. [***Svrha: Isprobati izračun regresijskog modela s različitim funkcijama preslikavanja i razviti intuiciju kako o tome kako funkcija preslikavanja određuje složenost hipoteze u ulaznome prostoru.***] Linearnim modelom univarijatne regresije želimo aproksimirati jednu periodu funkcije $f(x) = \sin(\pi x)$. Raspolažemo skupom primjera za učenje

$$\mathcal{D} = \{(0.25, 0.707), (0.5, 1), (1, 0), (1.5, -1), (2, 0)\}.$$

   (a) Izračunajte parametre linearnog modela regresije u izvornom prostoru primjera, tj. s funkcijom preslikavanja definiranom kao $\boldsymbol{\phi}(x) = (1, x)$. Skicirajte dobivenu regresijsku funkciju.

   (b) Izračunajte parametre modela jednostruke polinomijalne regresije drugog stupnja, tj. modela koji koristi funkciju preslikavanja $\boldsymbol{\phi}(x) = (1, x, x^2)$. Skicirajte dobivenu regresijsku funkciju.

   (c) Izračunajte parametre modela višestruke polinomijalne regresije četvrtog stupnja, tj. modela koji koristi funkciju preslikavanja $\boldsymbol{\phi}(x) = (1, x, x^2, x^3, x^4)$, uz L2-regularizaciju ($\lambda = 1$). Skicirajte dobivenu regresijsku funkciju.

   (d) Koji je model u ovom slučaju najprikladniji? Zašto?

   Napomena: Izračun možete načiniti u nekom alatu koji podržava izračun matričnih operacija. Skicu također možete načiniti u nekom alatu, ili je možete napraviti ručno, izračunom vrijednosti regresijske funkcije u nekoliko odabranih točaka.

4. [***Svrha: Uvjeriti se da, uz određene pretpostavke, kvadratna pogreška ima probabilističko tumačenje i opravdanje.***] Kod postupka najmanjih kvadrata empirijska je pogreška definirana kao:

$$E(\mathbf{w}|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}^{(i)}) - y^{(i)}\right)^2.$$

   Pokažite da je minimizacija gornjeg izraza istovjetna maksimizaciji log-izglednosti $\ln P(\mathcal{D}|\mathbf{w})$ (odnosno minimizaciji negativne log-izglednosti) uz pretpostavku normalno distribuiranog šuma $\mathcal{N}(h(\mathbf{x}|\mathbf{w}), \sigma^2)$.

5. [***Svrha: Razumjeti vezu između faktora regularizacije i složenosti modela.***] Neka $\mathcal{H}_{d,\lambda}$ označava model polinomijalne regresije stupnja d s L2-regularizacijskim faktorom λ. Razmatramo četiri modela: $\mathcal{H}_{2,0}$, $\mathcal{H}_{5,0}$, $\mathcal{H}_{5,100}$, $\mathcal{H}_{5,1000}$ u ulaznome prostoru $\mathcal{X} = \mathbb{R}$. Pretpostavimo da su podatci u stvarnosti generirani funkcijom koja je polinom trećeg stupnja ($d = 3$). Pretpostavite da imamo razmjerno malo podataka i da je šum u podatcima razmjerno velik. Na dva odvojena crteža skicirajte

   (a) regresijsku funkciju $h(x)$ za svih pet modela te

   (b) pogrešku učenja i ispitnu pogrešku za svih pet modela.

6. [***Svrha: Shvatiti kako regularizacija utječe na optimizaciju. Shvatiti geometrijski argument zašto L1-regularizacija rezultira rijetkim modelima.***]

   (a) Objasnite koja je svrha regularizacije i na kojoj se pretpostavci temelji.

   (b) Koja je prednost regulariziranog modela u odnosu na neregularizirani? Dolazi li ta prednost više do izražaja u slučajevima kada imamo puno primjera za učenje ili kada ih imamo malo?

(c) Razmatramo ~~jednostruku~~višestruku regresiju, $h(x) = w_0 + w_1 x_1 + w_2 x_2$. Skicirajte izokonture neregularizirane funkcije pogreške u ravnini $\mathbb{R}^2$ koju definiraju parametri $w_1 \cancel{w_0}$ i $w_2 \cancel{w_1}$ (napomena: funkcija pogreške je konveksna). Zatim skicirajte izokonture regularizacijskog izraza definiranog L2-normom vektora težina (i ova je funkcija konveksna). Pomoću ove skice objasnite na koji način regularizacija utječe na izbor optimalnih parametara $(w_1^*, w_2^*)\cancel{(w_0^*, w_1^*)}$. Skicirajte krivulju mogućih rješenja za $\lambda \in [0, \infty)$.

(d) Ponovite prethodnu skicu, ali ovog puta sa L1-regularizacijom. Na temelju ove skice pokušajte odgovoriti na pitanje zašto L1-regularizacija daje rjeđe modele od L2-regularizacije.

7. [***Svrha: Shvatiti vezu između težine značajki, važnosti značajki i složenosti modela.***]

(a) Treniramo model regresije uz nelinearnu funkciju preslikavanja $\boldsymbol{\phi} : \mathbb{R}^n \to \mathbb{R}^m$, gdje $m > n$, uz L2-regularizaciju. Optimalan regularizacijski faktor λ odredili smo unakrsnom provjerom. Kako, nakon treniranja modela, možemo provjeriti (1) koje su značajke nebitne i (2) je li izvorni model presložen?

(b) Kako bi se u ovom slučaju ponašao L1-regularizirani model?

(c) Pretpostavite da u skupu postoji skup multikolinearnih značajki koje su, osim što su redundantne, također i irelevantne, odnosno zavisna varijabla u stvarnosti uopće ne ovisi o tim varijablama. Ako model nije regulariziran, koje su očekivane težine tih značajki?

② DOMAĆA ZADAĆA 21.10.2016

① a) IZVOD U MATRIČNOM OBLIKU RJ. ZA VEKTOR $\vec{w}$ ZA NEREGULARIZIRANI LINEARNI MODEL REGRESIJE UZ KVADRATNU FUNKCIJU GUBITKA.

$$h(\vec{x}|w) = \vec{w}^T \vec{x}$$

$$E(h|D) = \frac{1}{2}\sum (h^{(i)}(x) - y^{(i)})^2 = \frac{1}{2}\sum(\vec{w}^T x^{(i)} - y^{(i)})^2 =$$

$$E(w|D) = \frac{1}{2}(\vec{w}^T \vec{x}^{(i)} - y^{(i)})^T(\vec{w}^T \vec{x}^{(i)} - y^{(i)}) = \frac{1}{2}(Xw - y)^T(Xw - y)$$

$\Rightarrow$ MINIMIZACIJA POGREŠKE $\nabla_w E(w|D) = \emptyset$ $\nabla_w \rightarrow$ parc. deriv. po w

$\nabla x^T A x = x^T(A + A^T)$

$\nabla A x = A$

$$E(w|D) = \frac{1}{2}((Xw)^T Xw - (Xw)^T y - y^T Xw + y^T y) =$$
$$= \frac{1}{2}((Xw)^T Xw - y^T(Xw) - y^T Xw + y^T y) =$$
$$= \frac{1}{2}((Xw)^T Xw - 2y^T(Xw) + y^T y) = \frac{1}{2}(w^T X^T X w) - 2y^T X w + y^T$$

$x^T A x$, $A = X^T X$

$$\nabla_w E(w|D) = \frac{1}{2}(w^T(A + A^T) - 2y^T X + \emptyset) =$$
$$= \frac{1}{2}(w^T(X^T X + X^T X) - 2y^T X) = w^T X^T X - y^T X = \emptyset$$

$w^T X^T X = y^T X \rightarrow X^T X w = X^T y \rightarrow w = (X^T X)^{-1} X^T y$

$(X^T X)^{-1} X^T$ PSEUDOINVERZ OD $X \rightarrow X^+$

$$\Rightarrow \boxed{w = X^+ y} \boxed{= (X^T X)^{-1} X^T y}$$

b) IZVEDI $\vec{w}$ ZA L2-REGULARIZIRANI MODEL

$$E(w|D) = \frac{1}{2}(\Phi w - y)^T(\Phi w - y) + \frac{\lambda}{2} w^T w =$$
$$= \frac{1}{2}((\Phi w)^T \Phi w - (\Phi w)^T y - y^T \Phi w + y^T y + \lambda w^T w)$$
$$= \frac{1}{2}(w^T(\Phi^T \Phi) w - 2y^T \Phi w + y^T y + \lambda w^T w)$$

$w^T w = w^T I w$

$$\nabla_w E(w|D) = \frac{1}{2}(w^T(\Phi^T\Phi + \Phi^T\Phi) - 2y^T\Phi + \emptyset + \lambda(w^T(I + I^T)))$$
$$= \frac{1}{2}(2w^T\Phi^T\Phi - 2y^T\Phi + 2\lambda w^T I) = w^T\Phi^T\Phi - y^T\Phi + \lambda w^T = \emptyset$$

$\Rightarrow \Phi^T\Phi w - \Phi^T y + \lambda w = \emptyset$ $(\Phi^T\Phi + \lambda I) w = \Phi^T y$

$$\boxed{w = (\Phi^T\Phi + \lambda I)^{-1} \Phi^T y}$$

c) $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^{4} = \{(0,4),(1,1),(2,2),(4,5)\}$. $\lambda = 10$, $h(x) = w_0 + w_1 x + w_2 x^2$. KAKO BI U OVOM SLUČAJU IZGLEDALA JEDNADŽBA IZ b)?

$$\Phi = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix} \quad \lambda I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{bmatrix} \quad y = \begin{bmatrix} 4 \\ 1 \\ 2 \\ 5 \end{bmatrix}$$

$w = (\Phi^T\Phi + \lambda I)^{-1} \Phi^T y$ (MATLAB)

$w_0 = 2.2945$

$w_1 = -0.1306$

$w_2 = 0.1779$

d) IZRAČUN JEDN. IZ b) ZNA BITI RAČUNALNO ZAHTJEVAN. ŠTO KONKRETNO PREDSTAVLJA PROBLEM? JE LI PROBLEM IZRAŽEN KAD IMAMO PUNO PR. ZA UČENJE ILI KADA IMAMO MNOGO ZNAČAJKI? OBRAZLOŽI.

- Računanje inverza matrica velikih dimenzija računski je izrazito zahtjevna.
- Problem je izražen kad imamo mnogo značajki m zbog:

$$\Phi = \text{dim. } N \times (m+1) \quad , \quad \Phi^T\Phi = [(m+1) \times N][N \times (m+1)] = (m+1) \times (m+1)$$

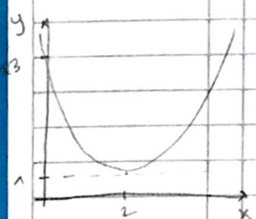
A kako od tog faktora $(+\lambda \mathbb{I})$ tražimo inverz, računska je složenost tim veća što imamo veći broj značajki.

e) RJ. JEDN. IZ a) MOŽE BITI NUMERIČKI NESTABILNO. ŠTO TO ZNAČI I KADA ĆE TO BITI SLUČAJ? POSTOJI LI TAKAV PROBLEM U SLUČAJU FORMULACIJE PROBLEMA IZ ZADATKA b)?

Rj. može biti nestabilno kada postoji singularitet. To će biti slučaj kada je matrica $\Phi^T\Phi$ loše kondicionirana, tj. kada dva stupca ili retka te matrice jako koreliraju.

U slučaju formulacije problema iz b) to neće biti problem jer matrici $\Phi^T\Phi$ koja može biti problematična, dodajemo faktor $\lambda \mathbb{I}$ (dijagonalnu matricu) što rješava problem. Na taj način smanjujemo korelaciju između stupaca/redaka, odnosno smanjujemo kondicijski broj.

(2) a) $y = 3(x-2)^2 + 1$. SKICIRAJ OVU FJU U ULAZNOM PROSTORU. DEF. LIN. MOD. $h(x)$ UZ FJU PRESLIKAVANJA $\Phi(x) = (1, x, x^2)$. ODREDI VEKTOR TEŽINA $\vec{w} = (w_0, w_1, w_2)$ TOG MODELA.



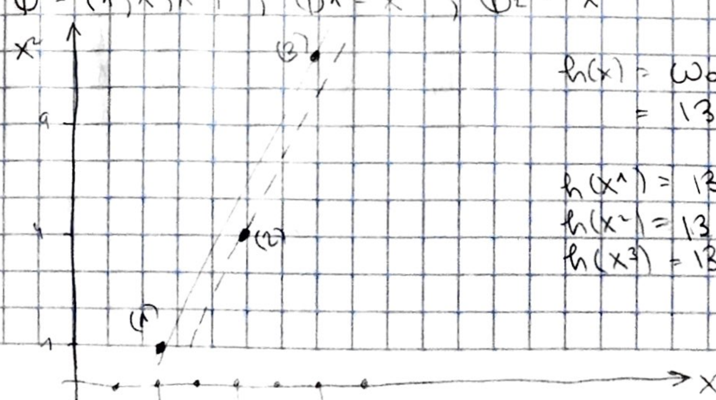
$$y = 3(x-2)^2 + 1 = 3(x^2 - 4x + 4) + 1 = 3x^2 - 12x + 12 + 1$$

$$y = \overset{w_2}{3}x^2 \overset{w_1}{-12}x + \overset{w_0}{13} \rightarrow \text{idealni } w\text{-ovi}$$

OPĆENITO: $\boxed{h(x) = w_0 + w_1 x + w_2 x^2}$

b) SKICIRAJ U PROSTORU (x_1, x_2) IZOKONTURNE FJE y. NAZNAČI U TOM PROSTORU TOČKE U KOJE SE PRESLIKAVAJU PR. $x^{(1)} = 1$, $x^{(2)} = 2$ i $x^{(3)} = 3$. KOJA JE VR. OD $h(x)$ ZA NAVEDENE PR.?

$\Phi = (1, x, x^2)$; $\Phi_1 = x$; $\Phi_2 = x^2$



$$h(x) = w_0 + w_1 x + w_2 x^2$$
$$= 13 - 12x + 3x^2$$

$$h(x^1) = 13 - 12 + 3 = 4$$
$$h(x^2) = 13 - 24 + 12 = 1$$
$$h(x^3) = 13 - 36 + 27 = 4$$

③ a) LINEARNIM MOD. UNIVAR. REGR. ŽELIMO APROX. JEDNU PERIODU F-je $f(x) = \sin(\pi x)$. RASPOLAŽEMO SKUPOM PR. ZA UČENJE

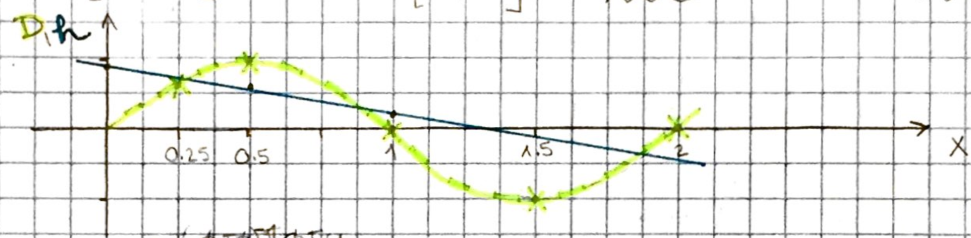
$$D = \{(0.25, 0.707), (0.5, 1), (1, 0), (1.5, -1), (2, 0)\}$$

w $\quad w = (\Phi^T\Phi)^{-1}\Phi^T y$

IZR. PARAM. LIN. MOD. REGRESIJE U IZVORNOM PROSTORU PR., S F-jom PRESLIKAVANJA $\Phi(x) = (1, x)$. SKICIRAJ DOBIVENU REGR. F-ju.

$$\Phi = \begin{bmatrix} 1 & 0.25 \\ 1 & 0.5 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 2 \end{bmatrix} \quad y = \begin{bmatrix} 0.707 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad \vec{w} = (\Phi^T\Phi)^{-1}\Phi^T y = \begin{bmatrix} 0.9433 \\ -0.7637 \end{bmatrix}$$

$$h(x) = w_0 + w_1 x = 0.9433 - 0.7637x$$

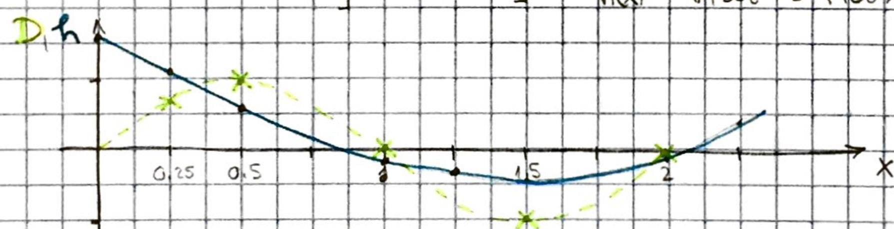


$w = (\Phi^T\Phi)^{-1}\Phi^T y$

b) IZR. PARAM. MODELA JEDNOSTRUKE POLINOMIJALNE REGR. II STUPNJA tj. MODELA koji KORISTI $\Phi(x) = (1, x, x^2)$. SKICIRAJ

$$\Phi = \begin{bmatrix} 1 & 0.25 & 0.0625 \\ 1 & 0.5 & 0.25 \\ 1 & 1 & 1 \\ 1 & 1.5 & 2.25 \\ 1 & 2 & 4 \end{bmatrix} \quad y = \begin{bmatrix} 0.707 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad w = (\Phi^T\Phi)^{-1}\Phi^T y = \begin{bmatrix} 1.7538 \\ -2.9408 \\ 0.9755 \end{bmatrix}$$

$$h(x) = 1.7538 - 2.9408x + 0.9755x^2$$



c) -||- $\Phi(x) = (1, x, x^2, x^3, x^4)$ uz L2-reg. $(\lambda = 1)$ + SKICA

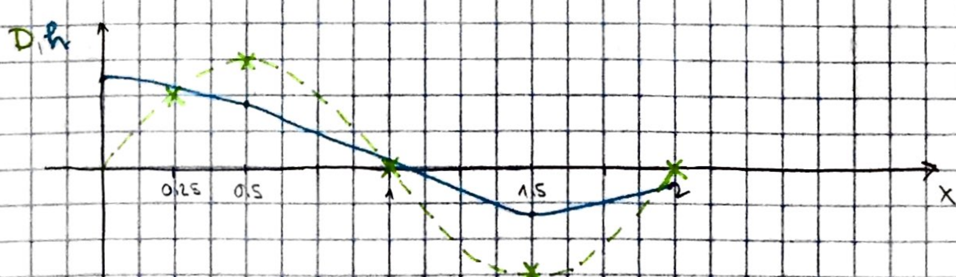
$\Phi = 5\times5 = \Phi^T$
$\Phi^T\Phi = 5\times5$
$\lambda I = 5\times5$

$$\Phi = \begin{bmatrix} 1 & 0.25 & 0.0625 & 0.0156 & 3.906\times10^{-3} \\ 1 & 0.5 & 0.25 & 0.125 & 0.0625 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1.5 & 2.25 & 3.375 & 5.0625 \\ 1 & 2 & 4 & 8 & 16 \end{bmatrix} \quad y = \begin{bmatrix} 0.707 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad \vec{w} = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T y$$

$$I = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Rightarrow \vec{w} = \begin{bmatrix} 0.8330 & -0.2818 & -0.4155 & -0.3461 & 0.2479 \end{bmatrix}^T$$

$$h(x) = 0.8330 - 0.2818x - 0.4155x^2 - 0.3461x^3 + 0.2479x^4$$



d) KOJI JE MODEL U OVOM SLUČAJU NAJPRIKLADNIJI I ZAŠTO?

Najprikladniji model je onaj pod c) s obzirom da ima najmanju kvadratnu pogrešku i oblikom najbolje prati željenu fju $f(x) = \sin(\pi x)$.

④

KOD POSTUPKA NAJMANJIH KVADRATA EMPIRIJSKA POGREŠKA DEFINIRANA JE KAO

$$E(\vec{w}|D) = \frac{1}{2}\sum_{i=1}^{N}\left(\vec{w}^T\phi(x^{(i)}) - y^{(i)}\right)^2$$

POKAŽITE DA JE MINIMIZACIJA GORNJEG IZRAZA ISTOVJETNA MAKSIMIZACIJI LOG IZGLEDNOSTI $\ln p(D|\vec{w})$ (ODNOSNO MINIMIZACIJI NEGATIVNE LOG IZGLEDNOSTI) UZ PRETPOSTAVKU NORMALNO DISTRIBUIRANOG ŠUMA $\mathcal{N}(h(x|w), \sigma^2)$

$\Rightarrow \quad E(w|D) = \frac{1}{2}\sum_{i=1}^{N}\left(\vec{w}^T\phi(x^{(i)}) - y^{(i)}\right)^2$

$\longrightarrow \quad p(D|w) = \prod_{i=1}^{N} p(x^{(i)}, y^{(i)}) = \prod_{i=1}^{N} p(y^{(i)}|x^{(i)})\, p(x^{(i)})$

$\longrightarrow \quad \ln \mathcal{L}(w|D) = \ln p(D|w) = \ln \prod_{i=1}^{N} p(y^{(i)}|x^{(i)})\, p(x^{(i)}) =$

$= \ln \prod_{i=1}^{N} p(y^{(i)}|x^{(i)}) + \ln \prod_{i=1}^{N} p(x^{(i)}) =$ NE OVISI O W!! Konst. je

$\ln \prod_{i=1}^{N} p(y^{(i)}|x^{(i)}) = \ln \prod_{i=1}^{N} \mathcal{N}(h(x^{(i)}|w), \sigma^2) =$

$= \ln \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(y^{(i)} - h(x^{(i)}|w))^2}{2\sigma^2}} = \ln \prod_{i=1}^{N} (\sqrt{2\pi}\,\sigma)^{-1} \cdot e^{-\left(\frac{(y^{(i)} - h(x|w))^2}{2\sigma^2}\right)} =$

$-(\ln(x \cdot x \cdot x \cdots)) =$
$-(\ln x + \ln x + \ln x \cdots) =$
$= -N \ln x$

$= \underbrace{-\ln \prod_{i=1}^{N} \sqrt{2\pi}\,\sigma}_{} + \ln \prod_{i=1}^{N} e^{-\frac{(y^{(i)} - h(x^{(i)}|w))^2}{2\sigma^2}} =$

$= \underbrace{-N \ln \sqrt{2\pi}\,\sigma}_{\text{KONST, ne ovisi o W}} - \sum_{i=1}^{N} \frac{(y^{(i)} - h(x^{(i)}|w))^2}{2\sigma^2}$

UZ PRETPOSTAVLJENU NORM. DISTR. GDJE JE GAUSSOV ŠUM $\mathcal{N}(0, \sigma)$, $\sigma = 1$!

$= -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y^{(i)} - h(x^{(i)}|w))^2 = -\frac{1}{2} \sum_{i=1}^{N} (y^{(i)} - h(x^{(i)}|w))^2$

MAKSIMIZACIJA LOG-IZGLEDNOSTI ODGOVARA MINIMIZACIJI FJE POGREŠKE DEFINIRANE KAO ZBROJ KVADRATNIH ODSTUPANJA (ako je pretpostavljeni šum Gaussov)

$\hookrightarrow \quad y^{(i)} = h(x^{(i)}) + \mathcal{N}(0, \sigma^2)$

5

$H_{d,\lambda}$ = MODEL POLINOM. REGR. STUPNJA d S L2 - REGULARIZAC. FAKTOROM λ

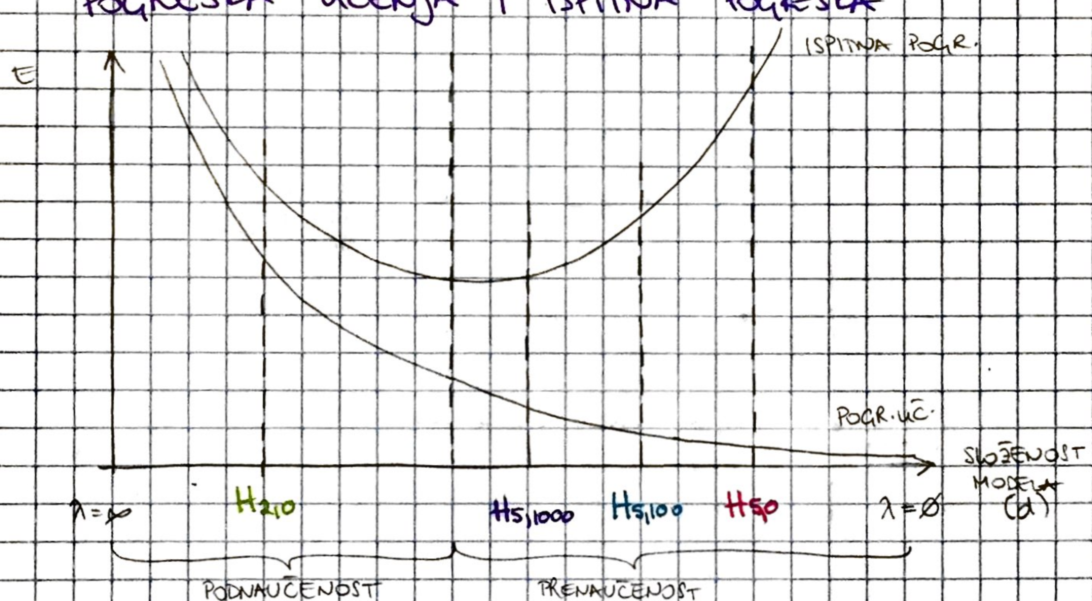
RAZMATRAMO 4 MODELA: $H_{2,0}$, $H_{5,0}$, $H_{5,100}$, $H_{5,1000}$ U UL. PROSTORU $X = \mathbb{R}$. PRETPOSTAVLJAMO DA SU PODACI U STVARNOSTI GENERIRANI F-JOM UZ $d=3$. PRETPOSTAVIMO DA IMAMO RAZMJERNO MALO PODATAKA I DA JE ŠUM U PODACIMA RAZMJERNO VELIK. NA DVA ODVOJENA CRTEŽA SKICIRAJTE:

a) REGRESIJSKU F-JU $h(x)$ ZA SVIH PET MODELA

$H_{2,0} \rightarrow d=2,\ \lambda=\emptyset$
$H_{5,0} \rightarrow d=5,\ \lambda=\emptyset$
$H_{5,100} \rightarrow d=5,\ \lambda=100$
$H_{5,1000} \rightarrow d=5,\ \lambda=1000$

$\lambda \uparrow$ f-ja postaje glađa
složenost modela $\downarrow$
smanjuje nelinearnost

b) POGREŠKA UČENJA I ISPITNA POGREŠKA



- $H_{2,0}$ će sigurno imati najveću pogrešku, stupanj d je premal + nema regularizacije
- $H_{5,0}$ će imati najmanju pogrešku učenja (model smo odlično istrenirali na primjerima) ali će imati najveću pogrešku generalizacije (jer nemamo uključenu regularizaciju i model se prilagodio šumu)
- $H_{5,100}$ ima veću pogr. generalizacije od $H_{5,1000}$ zbog manjeg λ, ali manju od $H_{5,0}$ jer ima veći λ. Pogr. učenja od $H_{5,100}$ je manja od $H_{5,1000}$ ali veća od $H_{5,0}$ jer se manje prilagođava na šum zbog regularizacije (od $H_{5,0}$)

⑥ a) SVRHA REGULARIZACIJE? NA KOJOJ SE PRETPOSTAVCI TEMELJI?

SVRHA regularizacije je SPRJEČAVANJE PRENAUČENOSTI MODELA ograničavanjem rasta vrijednosti parametara.

TEMELJI se na pretpostavci da što je LINEARNI MODEL SLOŽENIJI, to ima VEĆE VRIJEDNOSTI PARAMETARA. Iz tog razloga kažnjavamo hipoteze s visokim vrijednostima parametara! (u fju pogreške ugrađujemo mjeru složenosti modela).

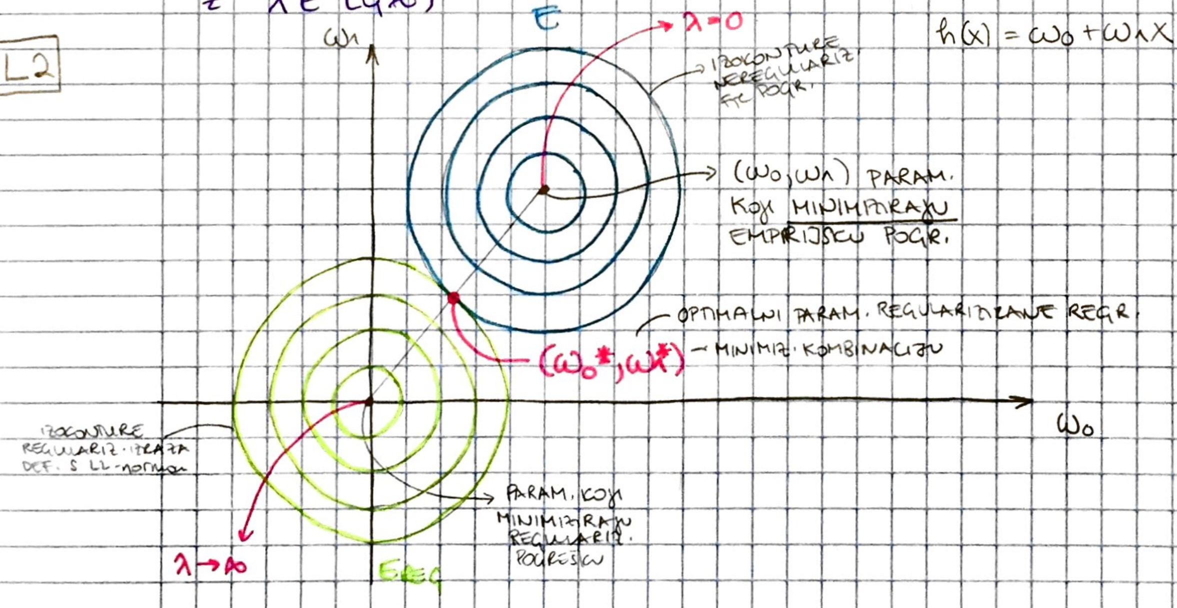
b) PREDNOST REGULARIZIRANOG MODELA U ODNOSU NA NEREGUL.? DOLAZI LI TA PREDNOST DO IZRAŽAJA U SLUČAJEVIMA KADA IMAMO PUNO ILI MALO PR. ZA UČENJE?

Prednost REGULARIZIRANOG modela u odnosu na neregularizirani je ta da je REG. MODEL TEŽE PRENAUČITI!! To dolazi do izražaja kad imamo MALO PR. ZA UČENJE!!

Druga prednost je poboljšanje numeričke stabilnosti.

c) RAZMOTRIMO JEDN. REGRESIJU $h(x) = w_0 + w_1 x$. SKICIRAJ IZOKONTURE NEREG. FJE POGR. U RAVNINI $\mathbb{R}^2$ KOJU DEF. PARAM. w_0 i w_1, ZATIM SKI. IZOKONT. REGULARIZ. IZRAZA DEF. S L2-normom VEKTORA TEŽINA. POMOĆU OVE SKICE OBJASNI NA KOJI NAČIN REGULARIZACIJA UTJEČE NA IZBOR OPTIMALNIH PARAM. (w_0^*, w_1^*). SKICIRAJ TE KRIVULJU MOGUĆIH RJ. Z $\lambda \in [0, \infty)$

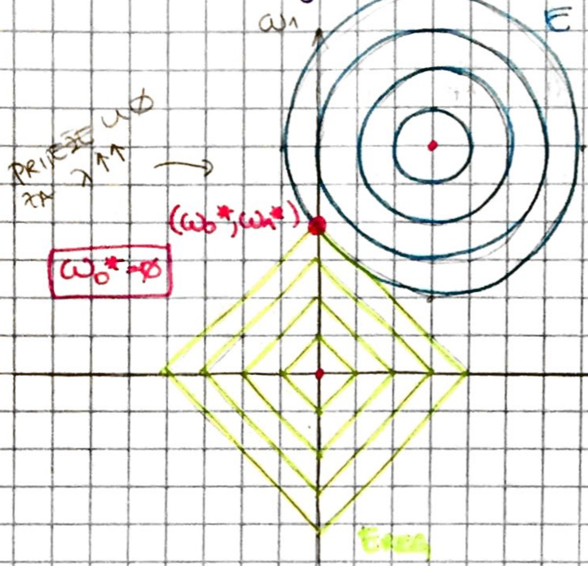
L2



- OPTIMALNI PARAM. UVIJEK SE NALAZE NA SPOJNICI IZOKONTURA
- OPT. PARAM. SU MANJI OD PARAM. KOJI MINIMIZIRAJU FJU POGREŠKE (KOD NEREG. REGR.).
- L2-REG. NE MOŽE REZULTIRATI RIJETKIM MODELIMA, ZA TO BI SE OPT. PARAM MORALI NALAZITI NA w_1 OSI. PLUS, L2 KAŽNJAVA TEŽINE PROPORCIONALNO NJIHOVOM IZNOSU
- L2 DAJE RJ. U ZATVORENOJ FORMI

d) PONOVI c) S L1 - reg. NA TEMELJU OVE SKICE POKUŠAJTE ODG NA PITANJE: ZAŠTO L1 REG. DAJE RJEĐE MODELE OD L2?

L1



IZOKONTURE REGULARIZACIJSKOG IZRAZA KOD L1 - REG. IMAJU KVADRATAN OBLIK TE JE VJEROJATNOST DA ĆE SE S IZOKONTUROM NEREG. FJE POGR. SJEĆI BLIZU (ILI NA) OSI KVADRATNOG SUSTAVA (w_0, w_1) PUNO VEĆA NEGO KOD IZOKONTURE L2 REG. IZRAZA, REZULTIRAJUĆI TIME RJEĐIM MODELIMA.

7 a) TRENIRAMO MODEL REGRESIJE UZ NELIN. FJU PRESLIKAVANJA $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$, GDJE JE $m > n$, UZ L2 REG. OPT. REG. FAKTOR λ ODREDILI SMO UNAKRSNOM PROVJEROM. KAKO, NAKON TRENIRANJA MODELA, MOŽEMO PROVJERITI:

L2 kažnjava složenost modela. što je model složeniji to će ga L2 više prigušiti, odnosno L2 nelinearne modele linearizira.

1) KOJE SU ZNAČAJKE NEBITNE

One značajke čije su vrijednosti nakon treniranja modela regresije uz L2 reg. JAKO PRIGUŠENE (odnosno param. w_j uz te značajke su $\approx Ø$) su NEBITNE! Odnosno manje su bitne od značajki koje nisu jako prigušene. (zato što L2 teže kažnjava značajke većeg iznosa) ($w_j >> 0$)

2) JE LI IZVORNI MODEL PRESLOŽEN

Izvorni je model PRESLOŽEN ako su parametri w_j uz "najsloženije" značajke (najviše potencije kod polinoma) blizu Ø! → To znači da je L2 kaznio, odnosno prigušio, te značajke.

b) KAKO BI SE U OVOM SLUČAJU PONAŠAO L1 - REG. MODEL?

L1 - regularizacijski model bi vjerojatno te značajke u potpunosti pritegnuo na Ø, rezultirajući rjeđim modelom. Ili bi ih u najmanju ruku još više prigušio.

c) →