Strojno učenje – domaća zadaća 4

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

Zadano: 31. 10. 2016. Rok: 4. 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 model logističke regresije. Razumjeti izvod funkcije pogreške unakrsne entropije i pripadne funkcije gubitka. Shvatiti zašto je ta funkcija gubitka unakrsne entropije prikladna za klasifikaciju, dok funkcija kvadratnog gubitka to nije.***]

   (a) Definirajte model logističke regresije. Zašto je sigmoidalna (logistička) funkcija prikladan odabir za aktivacijsku funkciju?

   (b) Izvedite pogrešku unakrsne entropije $E(\mathbf{w}|\mathcal{D})$ kao negativnu log-izglednost na skupu označenih primjera.

   (c) Napišite funkciju gubitka unakrsne entropije i nacrtajte njezin graf. Koliki je najveći a koliki najmanji mogući gubitak?

   (d) Pretpostavimo da su izlazne oznake $y \in \{-1, +1\}$ umjesto $y = \{0, 1\}$. Reformulirajte funkciju gubitka unakrsne entropije $L(h(\mathbf{x}), y)$ tako da koristi takve oznake te da vrijedi $L(0, y) = 1$ (kako bi funkcija bila kompatibilna s ostalim funkcijama gubitka koje smo radili).

   (e) Nacrtajte graf funkcije gubitka $L(\mathbf{x}, y)$ u ovisnosti o udjelu pogrešne klasifikacije $y\mathbf{w}^{\mathrm{T}}\mathbf{x}$, i to za: gubitak 0-1, kvadratni gubitak i logistički gubitak iz (d). Na temelju skice, odgovorite: (i) zašto je logistički gubitak dobar za klasifikaciju, a kvadratni gubitak to nije?; (ii) nanose li ispravno klasificirani primjeri ikakav gubitak?; (iii) možemo li reći da je logistički gubitak *konveksni surogat* gubitka 0-1, i što to znači?

2. [***Svrha: Prisjetiti se definicije konveksnosti funkcije. Razumjeti da konveksnost i unimodalnost nisu jedno te isto.***]

   (a) Formalno definirajte kada je funkcija $f : \mathbb{R}^n \to \mathbb{R}$ konveksna.

   (b) Funkcija f je *kvazikonveksna* (ili *unimodalna*) akko je njezina domena $\mathbf{dom}\, f$ konveksna te ako za svaki $\mathbf{x}, \mathbf{y} \in \mathbf{dom}\, f$ i $0 \leqslant \alpha \leqslant 1$ vrijedi

   $$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leqslant \max\{f(\mathbf{x}), f(\mathbf{y})\}.$$

   Kvazikonveksnost je poopćenje konveksnosti: svaka je konveksna funkcija unimodalna, ali obrat ne vrijedi. Pokažite primjerom da obrat ne vrijedi.

(c) Zašto u strojnom učenju volimo konkveksne funkcije pogreške? Koja je veza između konveksnosti funkcije pogreške i konveksnosti funkcije gubitka?

3. [***Svrha: Razumjeti gradijentni spust i potrebu za linijskim pretraživanjem. Znati izvesti gradijentni spust za logističku regresiju. Demonstirati upoznatost s prednostima i nedostatcima optimizacije drugog reda.***]

   (a) Objasnite ideju gradijentnog spusta i potrebu za linijskim pretraživanjem.

   (b) Objasnite razliku između grupnog (*batch*) i stohastičkog gradijentnog spusta. Koja je prednost ovog drugog?

   (c) Izrazite gradijent funkcije pogreške unakrsne entropije $\nabla E(\mathbf{w}|\mathcal{D})$ i napišite pseudokôd algoritma gradijentnog spusta (*batch* i stohastička izvedba).

   (d) Opišite prednost i nedostatak optimizacije drugog reda (Newtonov postupak) u kontekstu logističke regresije.

4. [***Svrha: Razumjeti kako regularizacija i linearna (ne)odvojivost utječu na gradijenti spust i na izgled funkcije pogreške u prostoru parametara.***] Koristimo model L2-regularizirane logističke regresije učene algoritmom gradijentnog spusta. Iskušavamo dvije vrijednosti regularizacijskog faktora $\lambda = 0$ i $\lambda = 100$. Razmatramo posebno linearno odvojiv i linearno neodvojiv problem.

   (a) Skicirajte pogreške učenja i ispitivanja $E(\mathbf{w}|\mathcal{D})$ u ovisnosti o broju iteracija za $\lambda = 0$ i $\lambda = 100$ te za slučaj (i) linearno odvojivih i (ii) linearno neodvojivih primjera (dva grafikona sa po četiri krivulje).

   (b) Načinite skice izokontura funkcije neregularizirane pogreške $E(\mathbf{w}|\mathcal{D})$ i L2-regularizacijskog izraza u ravnini w_1-w_2. Napravite dvije odvojene skice: za linearno odvojive i linearno neodvojive primjere.

   (c) Na grafikone iz prethodnoga zadatka docrtajte izokonture L2-regulariziranih funkcija pogreške za $\lambda = 100$ i naznačite gdje se nalazi točka minimuma (w_1^*, w_2^*). Gdje bi se nalazila točka minimuma za $\lambda = 0$?

5. [***Svrha: Znati izvesti algoritam multinomijalne logističke regresije.***]

   (a) Definirajte model multinomijalne logističke regresije.

   (b) Izvedite pogrešku modela multinomijalne logističke regresije kao negativnu log-izglednost na skupu označenih primjera.

6. [***Svrha: Uočiti zajedničkosti poopćenih linearnih modela.***]

   (a) Opišite veze između (i) modela linearne regresije, logističke regresije i multinomijalne logističke regresije, (ii) distribucija zavisne varijable y i (iii) aktivacijskih funkcija f. Što je zajedničko svim distribucijama s kojima smo dosada radili?

   (b) Objasnite riječima ovaj izraz:

$$\mathbf{w}^* = \operatorname*{argmax}_w \ln P(\mathcal{D}|\mathbf{w}) = \operatorname*{argmin}_w E(\mathbf{w}|\mathcal{D})$$

4) DOMAĆA ZADAĆA

1) a) DEF. MODEL LOG. REG. ZAŠTO JE SIGMOIDALNA F-JA PRIKLADNA KAO ODABIR ZA AKT. F-JU?

DEF. MODELA LOG. REG.: $h(x) = \sigma(\tilde{w}^T \tilde{x})$, gdje je $\sigma(\alpha) = \frac{1}{1+e^{-\alpha}}$

$$\boxed{h(x;w) = \frac{1}{1+e^{-w^T\phi}}}$$

⇒ Sigmoidalna f-ja je prikladna kao odabir za aktivacijsku funkciju zato što:

- Izlaz modela ima VJEROJATNOSNU INTERPRETACIJU, jer preslikava realne brojeve na interval $[0,1]$.
- Izlaz modela je Bernoullijeva varijabla i njegova f-ja gubitka slabo kažnjava predobro klasificirane primjere.
- Funkcija je DERIVABILNA $\sigma' = \sigma(1-\sigma)$ (Bitno za gradijentni spust)

b) IZVOD ZA POGR. UNAKRSNE ENTROPIJE $E(\vec{w}|D)$ KAO NEG. LOG. IZGLEDNOST NA SKUPU OZNAČENIH PR.

Funkcija pogreške = očekivanje funkcije gubitka! Odnosno definiramo ju kao NEGATIVNU LOG IZGLEDNOST NA SKUPU OZNAČENIH PR.

$$\Rightarrow \boxed{E(\vec{w}|D) = -\ln L(\vec{w}|D)}$$

S logističkom regr. gledamo kolika je izglednost param. $\vec{w}$ i želimo da ta izglednost bude što veća.

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

$$= \sum_{i=1}^{N} \ln p(y^{(i)}|x^{(i)},\vec{w}) + \underbrace{\sum_{i=1}^{N} \ln p(x^{(i)})}_{\text{konst.! Ne ovisi o } \vec{w} = 0}$$

$y^{(i)}$ – BERNOULLIJEVA VARIJABLA 0 ili 1, njezina distribucija: $\boxed{P(y^{(i)}) = \mu^{y^{(i)}}(1-\mu)^{1-y^{(i)}}}$

μ – vjerojatnost da je $y=1$

Nas zanima $p(y^{(i)}|x^{(i)},\vec{w})$, odnosno $p(y^{(i)}=1|x^{(i)},\vec{w})$ vjerojatnost da $x^{(i)}$ ima oznaku $y^{(i)}=1$:

$$\boxed{\mu = P(y^{(i)}=1|x^{(i)},\vec{w}) = h(x^{(i)},\vec{w})} \Rightarrow \boxed{P(y^{(i)}|x^{(i)},\vec{w}) = h(x^{(i)},\vec{w})^{y^{(i)}}(1-h(x^{(i)},\vec{w}))^{1-y^{(i)}}}$$

$$\Rightarrow \ln L(\vec{w}|D) = \sum_{i=1}^{N} \ln P(y^{(i)}|x^{(i)},\vec{w}) + \text{Konst.} = \sum_{i=1}^{N} \ln\left(h(x^{(i)},\vec{w})^{y^{(i)}}(1-h(x^{(i)},\vec{w}))^{1-y^{(i)}}\right) =$$

$$= \sum_{i=1}^{N}\left[y^{(i)} \ln h(x^{(i)},\vec{w}) + (1-y^{(i)}) \ln(1-h(x^{(i)},\vec{w}))\right] = \ln L(\vec{w}|D)$$

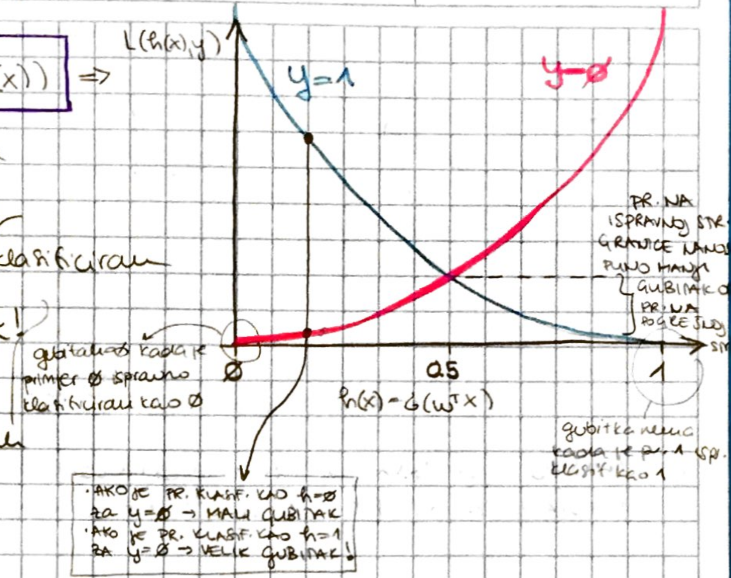
$$\Rightarrow \boxed{E(\vec{w}|D) = \sum_{i=1}^{N}\left(-y^{(i)} \ln h(x^{(i)}|\vec{w}) - (1-y^{(i)}) \ln(1-h(x^{(i)}|\vec{w}))\right)}$$

ALTERNATIVNO, KAKO POGR. NE BI OVISILA O BROJU PRIMJERA:

$$\boxed{E(\vec{w}|D) = \frac{1}{N}\sum_{i=1}^{N}\left(-y^{(i)} \ln h(x^{(i)}|\vec{w}) - (1-y^{(i)}) \ln(1-h(x^{(i)}|\vec{w}))\right)}$$

c) NAPIŠI FUNKCIJU GUBITKA UNAKRSNE ENTROPIJE I NACRTAJTE NJEZIN GRAF. KOLIKI JE NAJVEĆI, A KOLIKI NAJMANJI MOGUĆI GUBITAK?

$$\boxed{L(h(x),y) = -y \ln h(x) - (1-y)\ln(1-h(x))} \Rightarrow$$



↳ Izraz kompaktno definira gubitak za dva slučaja ($y=1$ i $y=\emptyset$)

↳ Gubitaka nema (SAMO)! kada je pr. savršeno (točno) klasificiran

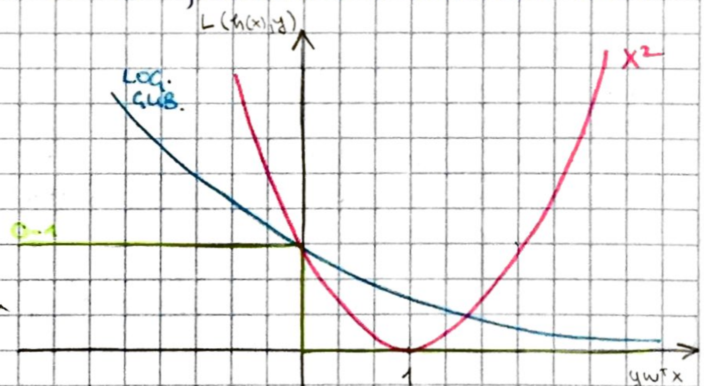
↓

U svim drugim slučajevima POSTOJI GUBITAK! S grafa možemo vidjeti da, čak i ako je pr. ispravno klasificiran, nalazi se na ispravnoj strani granice, postoji malen gubitak ovisno o pouzdanosti klasifikacije.

d) $y \in \{-1, 1\}$ umjesto $\{0, 1\}$. REFORMULIRAJTE $L(h(x), y)$ DA KORISTI TAKVE OZNAKE I DA VRIJEDI DA JE $L(0,y)=1$

$$\boxed{L(h(x),y) = \frac{1}{\ln 2}\ln\left(1 + e^{-y\vec{w}^T\vec{x}}\right)}$$

e) NACRTAJ GRAF FJE GUBITKA $L(x,y)$ ZA 0-1, KVADRATNI I LOGISTIČKI GUBITAK.



i) ZAŠTO JE LOGISTIČKI GUBITAK DOBAR ZA KLASIFIKACIJU A KVADRATNI GUBITAK NIJE?
- Zato što logistički gubitak puno manje kažnjava točno klasific. primjere.

ii) NANOSE LI ISPRAVNO KLASIFICIRANI PR. GUB.?
- Da, nanose. Ali kao što vidimo na grafu, gubitak je zanemarivo mal u odnosu na krivo klasificirane primjere.

iii) MOŽEMO LI REĆI DA JE LOG. GUB. KONVEKSNI SUROGAT GUBITKA 0-1 I ŠTO TO ZNAČI?
- Da možemo. To znači da ona za svaki $y\vec{w}^T\vec{x}$ prolazi IZNAD funkcije gubitka 0-1, odnosno kažemo da je ona njezina KONVEKSNA GORNJA MEĐA!

② a) FORMALNO DEF. KADA JE FJA $f: \mathbb{R}^n \to \mathbb{R}$ KONVEKSNA!

$f: \mathbb{R}^n \to \mathbb{R}$ je konveksna $\Longleftrightarrow$ ① je njezina DOMENA $\mathrm{dom}(f)$ KONVEKSAN SKUP

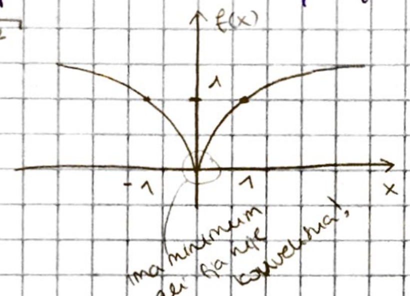
② za $\forall x_1, x_2 \in \mathrm{dom}(f)$ i $\forall \lambda \in (0,1)$ vrijedi:

$$\boxed{f(x) = f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)}$$

b) f je KVAZIKONVEKSNA (unimodalna) $\iff$ je dom(f) cvx te ako za $\forall x, y \in \text{dom}(f)$ i $\alpha \in (0,1)$ vrijedi $f(\alpha x + (1-\alpha)y) \leq \max\{f(x), f(y)\}$

KVAZIKONVEKSNOST JE POOPĆENJE KONVEKSNOSTI: Svaka je cvx f-ja unimodalna, ali obrat ne vrijedi! → Dokaži primjerom

npr. $f(x) = \sqrt[3]{x^2}$



$x_1 = 1$, $x_2 = 2$, $\alpha = 0.5$

i) KONVEKSNA? $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$

$f(0.5x_1 + 0.5x_2) \leq 0.5 f(x_1) + 0.5 f(x_2)$

$f(1.5) \leq 0.5 f(1) + 0.5 f(2)$

$1.31 \leq 0.5 + 0.79 = 1.29$

$1.31 \leq 1.29$? **NE** f NIJE CVX

ii) KVAZIKONVEKSNA? $f(\alpha x_1 + (1-\alpha)x_2) \leq \max(f(x_1), f(x_2))$

$1.31 \leq \max(1, 1.59) = 1.59$ **DA** fja je KVAZICVX

c) ZAŠTO U SU KORISTIMO KONVEKSNE FJE POGREŠKE? KOJA JE VEZA IZMEĐU KONVEKSNOSTI FJE POGR. I KONVEKSNOSTI FJE GUBITKA?

Svojstvo konveksnosti je poželjno jer garantira da će se rješenjem problema dobiti GLOBALNI, a ne LOKALNI MINIMUM! → Konveksne fje imaju JEDAN (GLOB) MIN!

Funkcija pogreške $= \sum$ fje gubitka. Stoga, ako je SVAKA POJEDINAČNA fja GUBITKA KONVEKSNA, TADA JE I ZBROJ TIH fja KONVEKSAN. Odnosno, tada je i fja POGREŠKE KONVEKSNA fja!

③ a) KOJA JE IDEJA GRAD. SPUSTA I LINIJSKOG PRETRAŽIVANJA?

GRADIJENTNI SPUST je jednostavan postupak heurističke optimizacije.

Gradijentni spust zasniva se na ideji da za funkciju f(x) u točci ekstrema vrijedi $\nabla f(x) = \emptyset$, dok u ostalim točkama funkcije vrijednost gradijenta $\nabla f(x)$ odgovara SMJERU PORASTA FUNKCIJE.

LINIJSKO PRETRAŽIVANJE – potrebno nam je s obzirom da ne možemo analitički odrediti grad. fje pogr. unakrsne entropije.

- odabire se η koji minimizira f(x) u smjeru spusta.

IDEJA: Krene se od neke početne točke x, minimum fje pronalazi se ažuriranjem vrijednosti x u smjeru koji je suprotan vektoru ∇f, dokle god se ∇f ne izjednači s $\emptyset$! Ako je fja konveksna, pronađeni minimum je ujedno i globalni minimum.

b) OBJASNI RAZLIKU IZMEĐU GRUPNOG (batch) I STOHASTIČKOG GRAD. SPUSTA. KOJA JE PREDNOST STOHASTIČKOG?

GRUPNI (batch) STOH. SPUST: – Računa ukupan $\nabla E(\tilde{w})$ za sve primjere iz skupa za učenje D

STOHASTIČKI GRAD. SPUST: – Ugađanje težina obavlja se na temelju svakog primjera pojedinačno, a to se onda ponavlja za svaki primjer iz skupa za učenje.

PREDNOST: – MANJE RAČUNALNO ZAHTJEVAN I MANJE PODLOŽAN ZAGLAVLJENJU U LOKALNOM OPTIMUMU, odnosno minimumu, kod fja koje nisu konveksne.
– Pogodan za on-line učenje

c) IZRAZI GRAD. FJ. POGR. UNAKRSNE ENTROPIJE $\nabla E(\vec{w}|D)$ I NAPIŠI PSEUDOKOD ALG. GRAD. SPUSTA (batch i on-line)

$$\boxed{\nabla E(\vec{w}|D) = \sum_{i=1}^{N} \left(h(\vec{x}^{(i)}) - y^{(i)}\right) \tilde{x}^{(i)}}$$

BATCH

$\tilde{w} = (0,0,0,\ldots,0)$
PONAVLJAJ DO KONVERGENCIJE
  $\Delta\tilde{w} = (0,0,0,\ldots,0)$
  za $i = 1,\ldots,N$
    $h = \sigma(\tilde{w}^T\tilde{x}^{(i)})$
    $\Delta\tilde{w} = \Delta\tilde{w} + (h - y^{(i)})\tilde{x}^{(i)}$
  $\tilde{w} = \tilde{w} - \eta \cdot \Delta\tilde{w}$

STOH

$\tilde{w} = (0,0,0,\ldots,0)$
PONAVLJAJ DO KONVERGENCIJE
  SLUČ. PERMUTIRAJ PR. U D
  ZA $i = 1,\ldots,N$
    $h = \sigma(\tilde{w}^T\tilde{x}^{(i)})$
    $\tilde{w} = \tilde{w} - \eta \cdot (h - y^{(i)})\tilde{x}^{(i)}$

d) PREDNOSTI / NEDOSTATCI OPTIMIZACIJE II REDA (NEWT.) U KONTEKSTU LOGISTIČKE REGRESIJE?

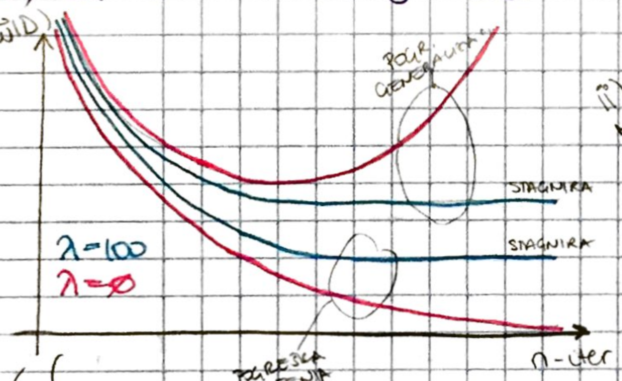
PREDNOSTI: MANJE ITERACIJA → BRŽI PRONALAZAK MINIMUMA

NEDOSTATCI: RAČUNSKA ZAHTJEVNOST → PROBLEM JE PRORAČUN HESSIAN MATRICE U SVAKOM KORAKU

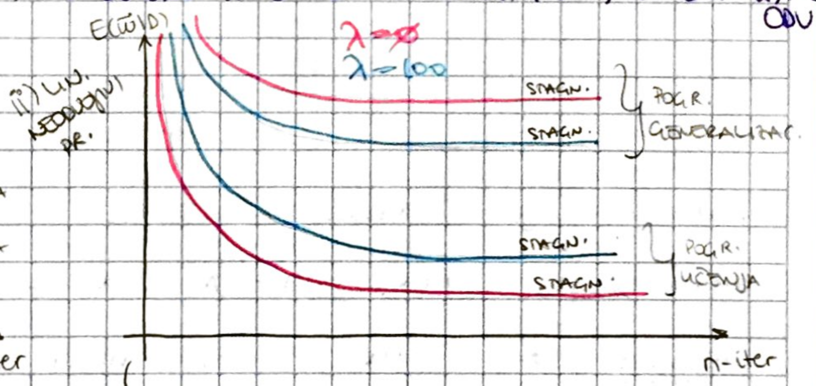
(4) L2-reg UČEN. ALG. GRAD. SPUSTA. $\lambda = 0$ (nema regular.) i $\lambda = 100$. RAZM. LIN. ODV. I NEODV. PROBLEM

a) POGR. UČ. i ISPITIVANJA $E(\vec{w}|D)$ U OVISNOSTI O BR. ITERAC. (n, za $\lambda=0$ i 100) i) LIN. ODV. ii) NEODV.

i) LINEARNO ODVOJIVI PRIMJERI



ii) LIN. NEODVOJIVI PR.



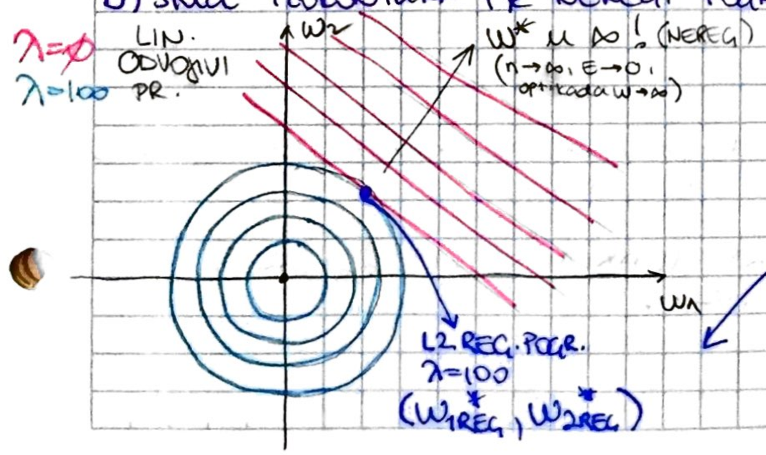
za $\lambda=\emptyset$, nema regularizacije, lin. odvojivi primjeri, standardni $E(\vec{w}|D)$. GRADIJENT POGREŠKE U OVOM SLUČ. NIKADA NEĆE BITI JEDNAK $\emptyset$ (Fja pogreške nema min), GRAD SPUST NEĆE KONVERGIRATI I TEŽINE RASTU U ∞. (SLOŽENIJI MODEL)

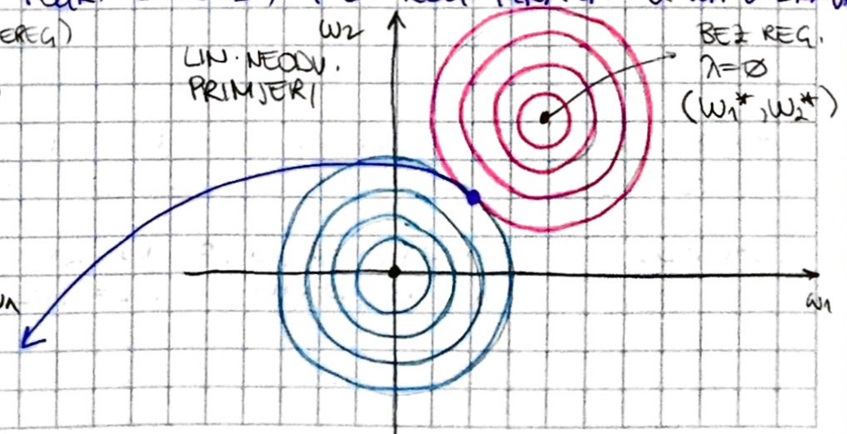
$\lambda=100$, UVODIMO REGULARIZACIJU kojom KAŽNJAVAMO SLOŽENE MODELE! VEĆI λ, VEĆA KAZNA SLOŽENIJIM MODELIMA, SPRJEČAVA SE PRENAUČENOST, DOLAZI DO STAGNACIJE U IZNOSU POGREŠKE.

OVAJ MODEL, za linearno neodvojive primjere, NIKADA NEĆEMO PRENAUČITI!! Zato što će uvijek biti nekih pogrešno klasificiranih primjera s obzirom da ih nemožemo linearno odvojiti. Iz tog razloga i za $\lambda=\emptyset$ i za $\lambda=100$ dolazi do stagnacije nakon određenog broja iteracija. Za $\lambda=100$ greška je manja jer se regularizacijom kažnjavaju složeni modeli.

c) (w_1^*, w_2^*)

b) SKICE IZOKONTURA Fje NEREG. POGR. $E(\vec{w}|D)$ I L2 REG. IZRAZA U w_1-w_2 RAV.





5) a) DEF. MODEL MULTINOMIJALNE LOGISTIČKE REGRESIJE → MNR

$$\boxed{h_k(\vec{x}|\vec{w}) = \frac{\exp(\vec{w}_k^T \Phi(x))}{\sum_j \exp(w_j^T \Phi(x))} = P(y=j|\vec{x},\vec{w})} \rightarrow w = (w_1, \dots, w_k)$$

b) IZVOD POGREŠKE MNR MODELA KAO NEG. LOG IZGL. NA SKUPU OZNAČ. PR.

$$E(\vec{w}|D) = -\ln L(\vec{w}|D) = -\left(\ln P(D|w)\right) = -\left(\ln \prod_{i=1}^{N} P(y^{(i)}|x^{(i)})\right)$$

uz: $y = (y_1, y_2, \dots, y_K)^T$ s $P(y|\mu) = \prod_{k=1}^{K} \mu_k^{y_k}$

$$E(\vec{w}|D) = -\ln \prod_{i=1}^{N}\prod_{k=1}^{K} \mu_k^{y_k^{(i)}} = -\ln \prod_{i=1}^{N}\prod_{k=1}^{K} h_k(x^{(i)}|w)^{y_k^{(i)}} = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_k^{(i)} \ln h_k(x^{(i)}|w)$$

$$\Rightarrow \boxed{E(\vec{w}|D) = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_k^{(i)} \ln h_k(\vec{x}^{(i)}|\vec{w})}$$

6) a) OBJASNI VEZE IZMEĐU:

i) MODELA LINEARNE REGRESIJE, LOGISTIČKE REGRESIJE I MULTINOMIJALNE LOG. REG:

$\boxed{h(x;w) = f(w^T\Phi(x))}$ →
- LIN. REG. $f = x \rightarrow h_1 = w^T x$
- LOG. REG $f = \sigma(x) \rightarrow h_2 = 1 / 1 + \exp(-w^T\Phi(x))$
- MNR $f = \text{SOFTMAX} \rightarrow h_3 = \dfrac{\exp(w^T\Phi(x))}{\sum_j \exp(w_j^T \Phi(x))}$

SVA 3 SU POOPĆENI LINEARNI MODELI?

iii) AKTIVAC. F-ja

određena distribucijom

ii) DISTRIBUCIJA ZAVISNE VAR. y:

- → LIN. REG.: $P(y|x,w) = \mathcal{N}(h(x), \sigma^2)$ – NORMALNA DISTRIB.
- → LOG. REG: $P(y|x,w) = h(x)^y (1-h(x))^{1-y}$ – BERNOULLIJEVA DISTRIBUC.
- → MNR: $P(y|x,w) = \prod_{k=1}^{K} h_k(x)^{y_k}$ – MULTINULLIJEVA DISTRIBUC.

SVE TRI DISTRIBUC. PRIPADAJU EKSPONENCIJ. FAMILIJI DISTRIBUC.

b) OBJASNI RIJEČIMA OVAJ IZRAZ:

$$\underbrace{\vec{w}^*}_{\text{OPTIMALNI PARAM.}} = \underbrace{\underset{w}{\text{argmax}} \ln P(D|\vec{w})}_{\text{MAKSIMIZACIJA LOG IZGLEDNOSTI}} = \underbrace{\underset{w}{\text{argmin}}\, E(\vec{w}|D)}_{\text{MINIMIZACIJA POGREŠKE MODELA}}$$

→ Minimizacija pogreške modela odgovara maksimizaciji log izglednosti na skupu označenih primjera. Odnosno:

→ OPTIMALNI PARAMETRI $\vec{w}^*$ MAKSIMIZIRAJU LOG IZGLEDNOST, ODNOSNO MINIMIZIRAJU POGREŠKU!!