Strojno učenje

9. Logistička regresija

prof. dr. sc. Bojana Dalbelo Bašić
doc. dr. sc. Jan Šnajder

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Ak. god. 2012/13.

Creative Commons Imenovanje–Nekomercijalno–Bez prerada 3.0 v1.0

Danas. . .

Danas. . .

1. Model logističke regresije
2. Probabilistička interpretacija
3. Učenje modela
4. Usporedba linearnih modela

Podsjetnik: poopćeni linearni modeli

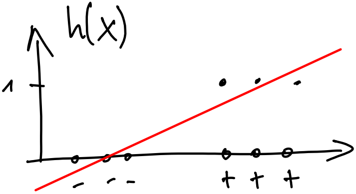
$$\boxed{h(\mathbf{x}) = f\big(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}}\big)}$$

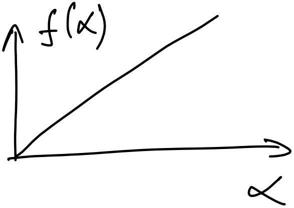
$f : \mathbb{R} \to [0,1]$ ili $f : \mathbb{R} \to [-1,+1]$ je aktivacijska funkcija

- Linearna granica u ulaznom prostoru (premda je f nelinearna)
- Model je nelinearan u parametrima (jer je f nelinearna)

Podsjetnik: klasifikacija regresijom

$$\boxed{h(\mathbf{x}) = \tilde{\mathbf{w}}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}) \qquad (f(\alpha) = \alpha)}$$



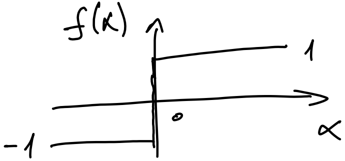


(+) uvijek dobivamo rješenje

(−) nerobusnost: ispravno klasificirani primjeri utječu na granicu
⇒ pogrešna klasifikacija čak i kod linearno odvojivih problema

Podsjetnik: perceptron

$$h(\mathbf{x}) = f\big(\tilde{\mathbf{w}}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x})\big) \qquad f(\alpha) = \begin{cases} +1 & \text{ako } \alpha \geqslant 0 \\ -1 & \text{inače} \end{cases}$$



(+) ispravno klasificirani primjeri ne utječu na granicu
$\Rightarrow$ ispravna klasifikacija linearno odvojivih problema

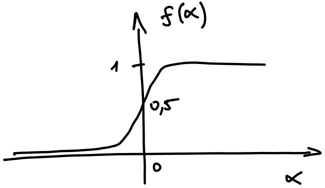
(−) aktivacijska funkcija nije derivabilna
$\Rightarrow$ funkcija gubitka nije derivabilna
$\Rightarrow$ gradijent funkcije pogreške nije nula u točki minimuma
$\Rightarrow$ postupak ne konvergira ako primjeri nisu linearno odvojivi

Logistička regresija

Aktivacijska funkcija s izlazima $[0, 1]$ i koja je derivabilna!

Logistička (sigmoidalna) funkcija:

$$\boxed{\sigma(\alpha) = \frac{1}{1 + \exp(-\alpha)}}$$



Model logističke regresije:

$$h(\mathbf{x}|\tilde{\mathbf{w}}) = \sigma\big(\tilde{\mathbf{w}}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x})\big) = \frac{1}{1 + \exp\big(-\tilde{\mathbf{w}}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x})\big)}$$

NB: Logistička regresija je klasifikacijski model (unatoč nazivu)!

Danas. . .

Probabilistički izlaz

$h(\mathbf{x}) \in [0,1]$, pa $h(\mathbf{x})$ možemo tumačiti kao vjerojatnost da primjer pripada klasi $\mathcal{C}_1$:

$$\boxed{h(\mathbf{x}) = \sigma\big(\tilde{\mathbf{w}}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x})\big) = P(\mathcal{C}_1|\mathbf{x})}$$

Koje je zapravo opravdanje za to?!

Moramo imati neku pretpostavku o distribuciji primjera i klasa!

Aposteriorna vjerojatnost klase

(Odsada nadalje: $\phi(\mathbf{x}) = \tilde{\mathbf{x}}$)

Aposteriorna vjerojatnost klase modelirana poopćenim linearnim modelom?

$$\boxed{p(\mathcal{C}_j|\mathbf{x}) = f\left(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}}\right)} \; ?$$

$$P(\mathcal{C}_1|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)} = \frac{1}{1 + \frac{p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)}{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}}$$

$$= \frac{1}{1 + \exp\left(\ln \frac{p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)}{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}\right)} = \frac{1}{1 + \exp(-\alpha)} = \sigma(\alpha)$$

$$\alpha = \ln \frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)} = \ln \frac{P(\mathcal{C}_1|\mathbf{x})}{P(\mathcal{C}_2|\mathbf{x})} = \ln \frac{P(\mathcal{C}_1|\mathbf{x})}{1 - P(\mathcal{C}_1|\mathbf{x})}$$

Digresija: logit-funkcija

Općenito, inverz logističke funkcije

$$\boxed{\sigma(\alpha) = \frac{1}{1+\exp(-\alpha)} = p}$$

naziva se logit-funkcija ili logaritam omjera šansi (engl. *log-odds*):

$$\mathrm{logit}(p) = \sigma^{-1}(\alpha) = \ln\frac{p}{1-p} = \alpha$$

Za poopćeni linearni model

$$P(\mathcal{C}_1|\mathbf{x}) = \frac{1}{1+\exp(-\alpha)} = \sigma(\alpha) = \sigma(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})$$

$=$

mora vrijediti

$$\alpha = \ln\frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)} = \ln p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1) - \ln p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2) = \tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}}$$

Prisjetite se Bayesovog klasifikatora za kontinuirane ulaze:

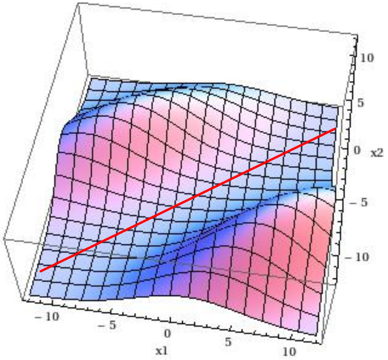
$$\boxed{h_j(\mathbf{x}) = \ln p(\mathbf{x}|\mathcal{C}_j)P(\mathcal{C}_j)}$$

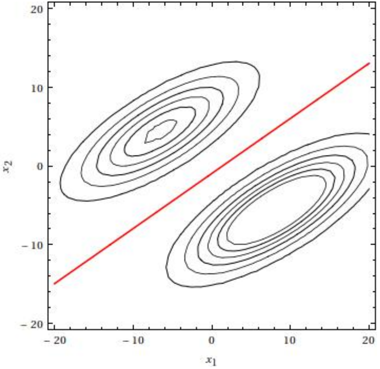
Q: Uz koje pretpostavke ovaj model degenerira u linearan model?

Aposteriorna vjerojatnost klase

Dijeljena kovarijacijska matrica $\mathbf{\Sigma}$:

$$h_j(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_j - \frac{1}{2}\boldsymbol{\mu}_j^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_j + \ln P(\mathcal{C}_j) \qquad \text{(3.29) u skripti}$$





Aposteriorna vjerojatnost klase

Diskriminacijska funkcija za $K = 2$:

$$
\begin{aligned}
h_{12}(\mathbf{x}) &= h_1(\mathbf{x}) - h_2(\mathbf{x}) \\
&= \ln p(x|\mathcal{C}_1)P(\mathcal{C}_1) - \ln p(\mathbf{x}|C_2)P(\mathcal{C}_2) = \ln \frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)} = \alpha \\
&= +\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \frac{1}{2}\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \ln P(\mathcal{C}_1) \quad \Leftarrow C_1 \\
&\quad\; - \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \frac{1}{2}\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 - \ln P(\mathcal{C}_2) \quad \leftarrow C_2 \\
&= \underset{x^T}{\mathbf{x}^{\mathrm{T}}}\underbrace{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}_{w_1} \underbrace{- \frac{1}{2}\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln \frac{P(\mathcal{C}_1)}{P(\mathcal{C}_2)}}_{w_0}
\end{aligned}
$$

Model je linearan:

$$\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)-\frac{1}{2}\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1+\frac{1}{2}\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2+\ln\frac{P(\mathcal{C}_1)}{P(\mathcal{C}_2)}=\underline{\mathbf{w}^{\mathrm{T}}\mathbf{x}+w_0}$$

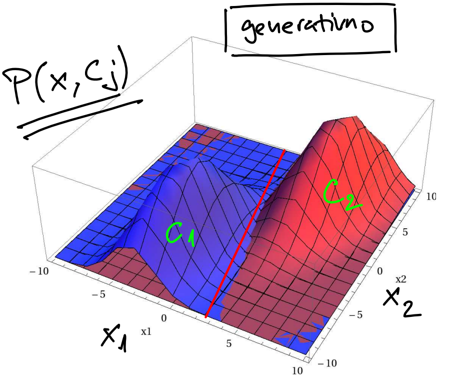
$$\begin{aligned}\mathbf{w}&=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\ w_0&=-\frac{1}{2}\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1+\frac{1}{2}\boldsymbol{\mu}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2+\ln\frac{P(\mathcal{C}_1)}{P(\mathcal{C}_2)}\end{aligned}$$

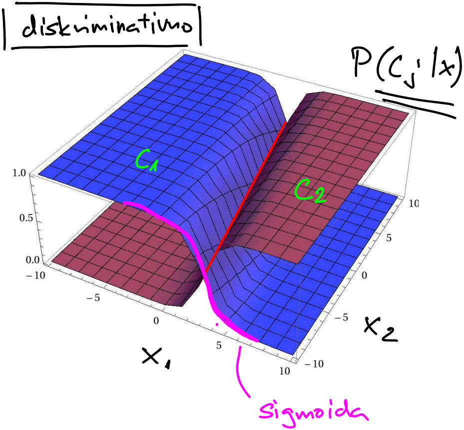
Prema tome:

$$\boxed{P(\mathcal{C}_1|\mathbf{x})=\sigma(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})}$$

Probabilistička interpretacija

Ako pretpostavimo da su primjeri $\mathbf{x}$ generirani iz klasa $\mathcal{C}_1$ i $\mathcal{C}_2$ s normalno distribuiranim izglednostima $p(\mathbf{x}|\mathcal{C}_1)$ odnosno $p(\mathbf{x}|\mathcal{C}_2)$ i s dijeljenom kovarijacijskom matricom $\mathbf{\Sigma}$, izlaz modela logističke regresije $h(\mathbf{x}|\tilde{\mathbf{w}}) = \sigma(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})$ odgovara aposteriornoj vjerojatnosti klase $P(\mathcal{C}_1|\mathbf{x})$.





Diskriminativni vs. generativni model

Logistička regresija – diskriminativan model:

$$P(\mathcal{C}_1|\mathbf{x}) = \sigma(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})$$ (conditional model)

⇒ modeliramo izravno aposteriornu vjerojatnost!

Bayesov klasifikator – generativan model:

$$P(\mathcal{C}_1|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_2)}$$ (joint model)

⇒ modeliramo izglednosti i apriorne vjerojatnosti klasa

Q: Koliko parametara imaju ovi modeli za n-dimenzijski ulazni prostor?

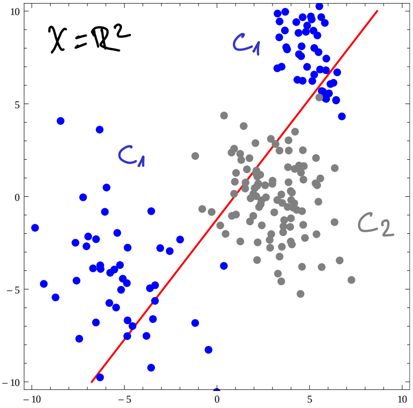
Logistička regresija: $n+1 \Rightarrow O(n)$

Bayesov klasifikator: $\frac{n}{2}(n+1) + 2n + 1 \Rightarrow O(n^2)$

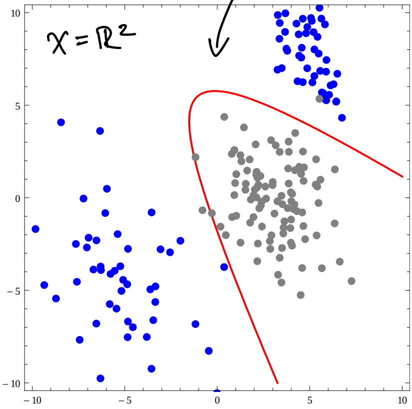
Nelinearan model?

Preslikavanje u prostor značajki: $\phi : \mathbb{R}^n \to \mathbb{R}^m$

nelinearna granica
u ulaznom prostoru



$$\phi(\mathbf{x}) = \tilde{\mathbf{x}}$$



$$\phi(\mathbf{x}) = (\underbrace{1, x_1, x_2}_{\tilde{\mathbf{x}}}, x_1 x_2, x_1^2, x_2^2)$$

Danas. . .

Plan za dalje...

Definirali smo (1) model logističke regresije. $\longrightarrow h(x) = \sigma(\tilde{w}^T \phi(x))$

Trebamo još definirati (2) funkciju gubitka i (3) optimizacijski postupak.

Funkciju gubitka izvest ćemo iz funkcije pogreške. $\longrightarrow \frac{1}{N}\sum_{i}^{N} L(x^i)$
(Podsjetnik: funkcija pogreške = očekivanje funkcije gubitka.)

Funkcija pogreške je negativna log-izglednost skupa primjera $\mathcal{D}$
$\Rightarrow$ $\tilde{\mathbf{w}}$ koji minimizira pogrešku je onaj koji primjere čine najvjerojatnijima.

$\downarrow$

MLE

Funkcija pogreške

Log-izglednost skupa primjera $\mathcal{D}$ jednaka je log-izglednosti skupa $\underline{\underline{\{y^{(i)}\}_{i=1}^N}}$ (jer su $\mathbf{x}^{(i)}$ fiksirani):

vjerojatnost oznake

$$\ln \mathcal{L}(\tilde{\mathbf{w}}|\mathcal{D}) = \ln \prod_{i=1}^N P(y^{(i)}|\mathbf{x}^{(i)}) = \ln \prod_{i=1}^N \underbrace{h(\mathbf{x}^{(i)})^{y^{(i)}} \big(1 - h(\mathbf{x}^{(i)})\big)^{1-y^{(i)}}}_{\text{Bernoullijeva distr.}}$$

$= \ln P(\mathcal{D}|\tilde{\mathbf{w}})$

ne modeliramo $P(x,y)$ => diskriminativan model!

Funkcija pogreške:

$$E(\tilde{\mathbf{w}}|\mathcal{D}) = -\ln \mathcal{L}(\tilde{\mathbf{w}}|\mathcal{D}) = -\sum_{i=1}^N \Big\{ y^{(i)} \ln h(\mathbf{x}^{(i)}) + (1-y^{(i)}) \ln \big(1 - h(\mathbf{x}^{(i)})\big) \Big\}$$

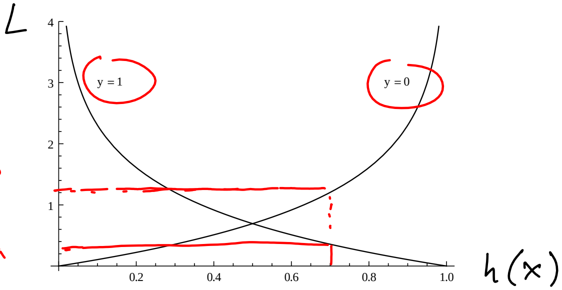
$\Rightarrow$ pogreška unakrsne entropije (engl. *cross-entropy error*)

Alternativno:

$$E(\tilde{\mathbf{w}}|\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} -\Big\{\underbrace{y^{(i)}\ln h(\mathbf{x}^{(i)}) + (1-y^{(i)})\ln\big(1-h(\mathbf{x}^{(i)})\big)}\Big\}$$

Funkcija gubitka:

$$L(h(\mathbf{x}), y) = -y\ln h(\mathbf{x}) - (1-y)\ln\big(1-h(\mathbf{x})\big)$$



Samo za savršeno točnu klasifikaciju nema gubitka

Minimizacija pogreške

$$L(h(\mathbf{x}), y) = -y \ln h(\mathbf{x}) - (1-y) \ln \big(1 - h(\mathbf{x})\big)$$

$$\boxed{E(\tilde{\mathbf{w}}) = \sum_{i=1}^{N} L\big(h(\mathbf{x}^{(i)}|\tilde{\mathbf{w}}), y^{(i)}\big)}$$

Konveksne funkcije!

$\Rightarrow$ Nema rješenje u z.f. Minimiziramo gradijentnim spustom:

$$\nabla E(\tilde{\mathbf{w}}) = \sum_{i=1}^{N} \nabla L\big(h(\mathbf{x}^{(i)}|\tilde{\mathbf{w}}), y^{(i)}\big)$$

$$\frac{\partial}{\partial x}\sigma(x) = \sigma(x)(1-\sigma(x))$$

$$\nabla L\big(h(\mathbf{x}), y\big) = \Big(-\frac{y}{h(\mathbf{x})} + \frac{1-y}{1-h(\mathbf{x})}\Big) h(\mathbf{x})\big(1-h(\mathbf{x})\big)\tilde{\mathbf{x}} = \big(h(\mathbf{x}) - y\big)\tilde{\mathbf{x}}$$

DZ

$$\boxed{\nabla E(\tilde{\mathbf{w}}) = \sum_{i=1}^{N} \big(h(\mathbf{x}^{(i)}) - y^{(i)}\big)\tilde{\mathbf{x}}^{(i)}}$$

Gradijentni spust

Logistička regresija (gradijentni spust)

1: $\tilde{\mathbf{w}} \leftarrow (0, 0, \ldots, 0)$
2: **ponavljaj** do konvergencije
3: $\quad \Delta\tilde{\mathbf{w}} \leftarrow (0, 0, \ldots, 0)$
4: $\quad$ **za** $i = 1, \ldots, N$
5: $\quad\quad h \leftarrow \sigma(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}}^{(i)})$
6: $\quad\quad \Delta\tilde{\mathbf{w}} \leftarrow \Delta\tilde{\mathbf{w}} + (h - y^{(i)})\,\tilde{\mathbf{x}}^{(i)}$
7: $\quad \tilde{\mathbf{w}} \leftarrow \tilde{\mathbf{w}} - \eta\Delta\tilde{\mathbf{w}}$

korekcija

faktor učenja
(npr. $\eta = 0{,}001$)

Gradijentni spust

Logistička regresija (stohastički gradijentni spust)

$\tilde{\mathbf{w}} \leftarrow (0, 0, \ldots, 0)$
ponavljaj do konvergencije
  (slučajno permutiraj primjere u $\mathcal{D}$)
  za $i = 1, \ldots, N$
    $h \leftarrow \sigma(\tilde{\mathbf{w}}^{\mathrm{T}} \tilde{\mathbf{x}}^{(i)})$
    $\tilde{\mathbf{w}} \leftarrow \tilde{\mathbf{w}} - \eta(h - y^{(i)})\tilde{\mathbf{x}}^{(i)}$

Regularizacija

Regularizacija sprečava (smanjuje mogućnost) prenaučenosti.

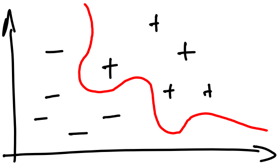
(1) Ako je model nelinearan, regularizacijom sprečavamo prenaučenost

(2) Ako imamo puno značajki, regularizacijom efektivno smanjujemo broj značajki jer težine potiskujemo prema nuli

(3) Ako je problem linearno odvojiv, sprječavamo "otvrdnjivanje" sigmoide

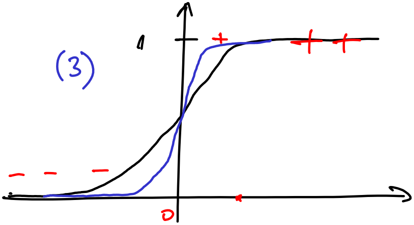
→ Svojstveno log. regresiji!

(2)



$\phi(\vec{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$

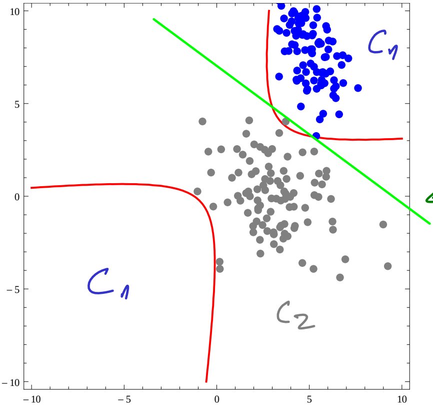
(3)



$\|w\|$ raste!

Bez regularizacije:

na linearno odvojivom problemu lako je prenaučiti nelinearan model



granica jednostavnog linearnog modela

$$\phi(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

L2-regularizacija

Regularizacijski izraz

$$E(\tilde{\mathbf{w}}|\mathcal{D}) = \sum_{i=1}^{N} -\Big\{y^{(i)} \ln h(\mathbf{x}^{(i)}) + (1-y^{(i)}) \ln\big(1-h(\mathbf{x}^{(i)})\big)\Big\} + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

$= \frac{\lambda}{2}\|w\|^2$

Korekcija težina:

$\underline{w_0}$ ne regularizirati!

$$\mathbf{w} \leftarrow \mathbf{w} - \eta\Big(\sum_{i=1}^{N}\big(h(\mathbf{x}^{(i)}) - y^{(i)}\big)\mathbf{x}^{(i)} + \lambda\mathbf{w}\Big)$$

Ekvivalentno:

$$\mathbf{w} \leftarrow \mathbf{w}(1-\eta\lambda) - \eta\sum_{i=1}^{N}\big(h(\mathbf{x}^{(i)}) - y^{(i)}\big)\mathbf{x}^{(i)}$$

L2-regularizacija

L2-regularizirana logistička regresija (gradijentni spust)

$\tilde{\mathbf{w}} \leftarrow (0, 0, \ldots, 0)$
ponavljaj do konvergencije
  $\Delta w_0 \leftarrow 0$
  $\Delta \mathbf{w} \leftarrow (0, 0, \ldots, 0)$
  za $i = 1, \ldots, N$
    $h \leftarrow \sigma(\tilde{\mathbf{w}}^{\mathrm{T}} \tilde{\mathbf{x}}^{(i)})$
    $\Delta w_0 \leftarrow \Delta \mathbf{w}_0 + h - y^{(i)}$
    $\Delta \mathbf{w} \leftarrow \Delta \mathbf{w} + (h - y^{(i)}) \mathbf{x}^{(i)}$
  $w_0 \leftarrow w_0 - \eta \Delta w_0$
  $\mathbf{w} \leftarrow \mathbf{w}(1 - \eta\lambda) - \eta \Delta \mathbf{w}$

L2-regularizacija

L2-regularizirana logistička regresija (stohastički gradijentni spust)

$\tilde{\mathbf{w}} \leftarrow (0, 0, \dots, 0)$
ponavljaj do konvergencije:
    (slučajno permutiraj primjere u $\mathcal{D}$)
    za $i = 1, \dots, N$
        $h \leftarrow \sigma(\tilde{\mathbf{w}}^{\mathrm{T}} \tilde{\mathbf{x}}^{(i)})$
        $w_0 \leftarrow w_0 - \eta(h - y^{(i)})$
        $\mathbf{w} \leftarrow \mathbf{w}(1 - \eta\lambda) - \eta(h - y^{(i)})\mathbf{x}^{(i)}$

Danas. . .

Funkcija gubitka logističke regresije

Elegantnija formulacija funkcije gubitka:

$$h(\mathbf{x}) = \frac{1}{1+\exp(-\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})} \qquad \boxed{1-h(\mathbf{x})} = \frac{1}{1+\exp(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})}$$

$-$ $+$

$$\begin{aligned} L(h(\mathbf{x}),y) &= -y\ln h(\mathbf{x}) - (1-y)\ln(1-h(\mathbf{x})) \\ &= y\ln\big(1+\exp(-\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})\big) + (1-y)\ln\big(1+\exp(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})\big) \end{aligned}$$

Uz $y \in \{-1,+1\}$, dobivamo:

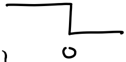
$$\boxed{L(h(\mathbf{x}),y) = \ln\big(1+\exp(-y\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})\big)}$$

$\Rightarrow$ logistička funkcija gubitka (engl. *logistic loss*)

Usporedba funkcija gubitka

$\boxed{y \in \{-1, +1\}.}$ Za pogrešnu klasifikaciju vrijedi $y\tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x} < 0$

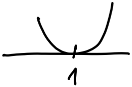
- Gubitak 0-1 (engl. *0-1 loss, misclassification loss*):



$$L(h(\mathbf{x}), y) = \mathbf{1}\{\mathrm{sgn}(\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}}) \neq y^{(i)}\} = \mathbf{1}\{y\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}} < 0\}$$

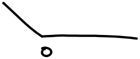
- Kvadratni gubitak (engl. *quadratic loss*):

$$L(h(\mathbf{x}), y) = (1 - y\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})^2$$



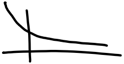
- Gubitak perceptrona (engl. *perceptron loss*):

$$L(h(\mathbf{x}), y) = \max\left(0, -y\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}}\right)$$

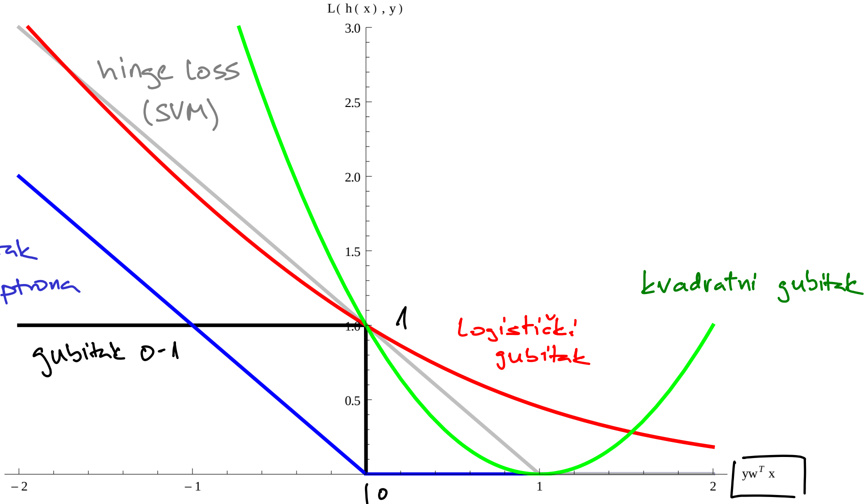


- Logistički gubitak (engl. *logistic loss*):

$$L(h(\mathbf{x}), y) = \ln\left(1 + \exp(-y\tilde{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{x}})\right)$$



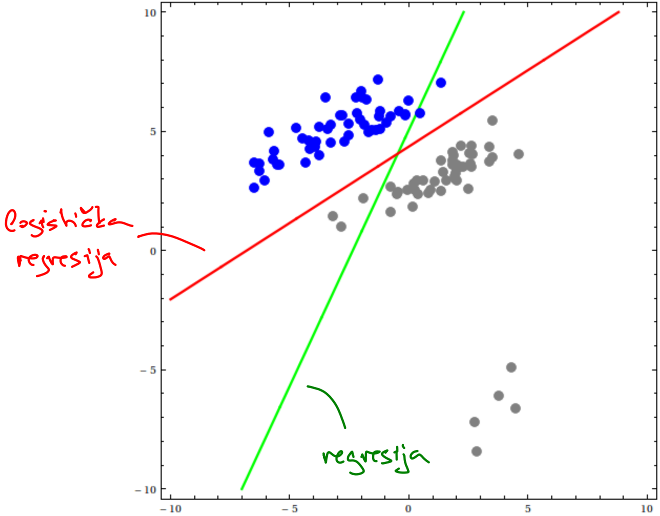
Usporedba funkcija gubitka



posrešna klasifikacija ← → točna klasifikacija

$y\tilde{w}^T\tilde{x} < 0$ $\quad$ $y\tilde{w}^T\tilde{x} > 0$

Usporedba hipoteza



Sažetak

- Logistička regresija je **diskriminativan klasifikacijski model** s probabilističkim izlazom
- Model **odgovara generativnom modelu** s normalno distribuiranim izglednostima i dijeljenom kovarijacijskom matricom, ali je broj parametara logističke regresije manji
- Koristi se **logistička funkcija gubitka** odnosno **pogreška unakrsne entropije**
- Optimizacija se provodi **gradijentnim spustom**, a prenaučenost se može spriječiti **regularizacijom**
- Logistička regresija je vrlo dobar algoritam koji **nema nedostatke** koje imaju klasifikacija regresijom i perceptron

Sljedeća tema: Neparametarski modeli