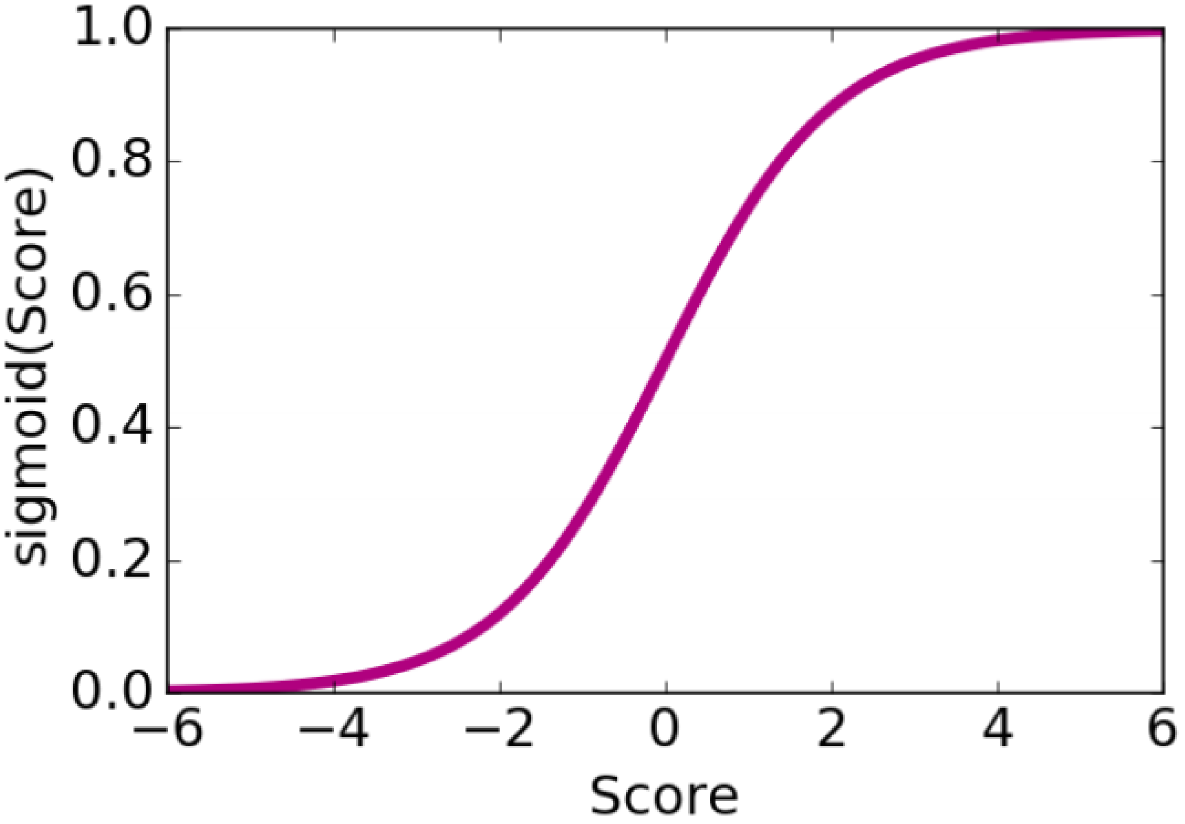
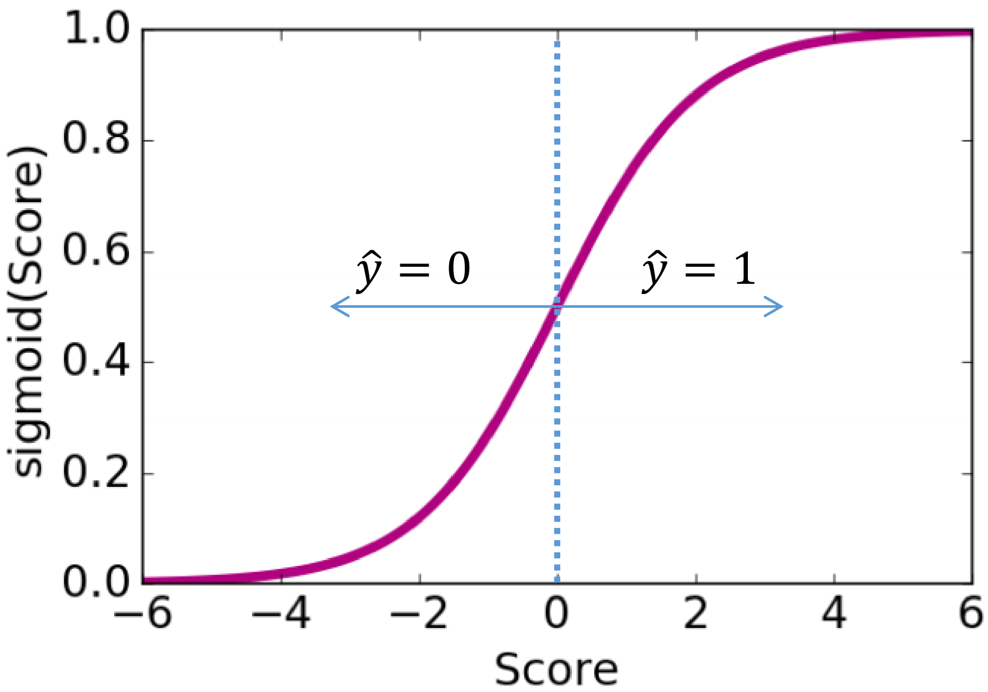


Logistička regresija

Logistička regresija

- Jedan specijalan slučaj GLM-a je *Logistic Regression Classifier*
- Kao funkcija *g* koristi se *logistic link function* (*sigmoid*, *logit*):

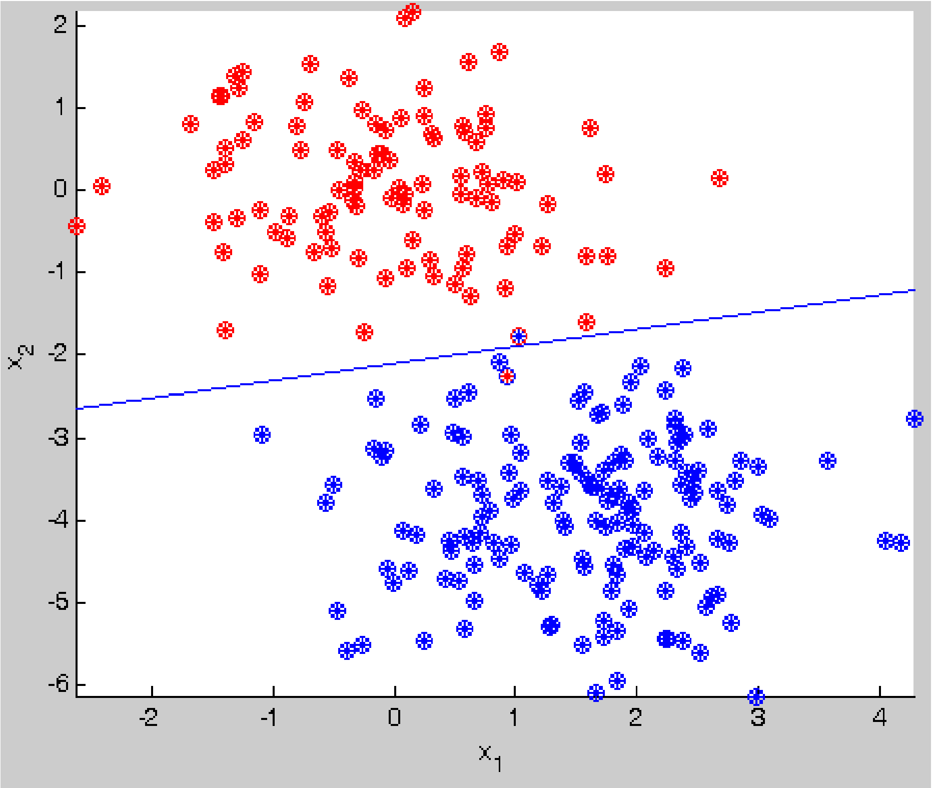
$$\sigma(score) = \frac{1}{1 + e^{-score}}$$

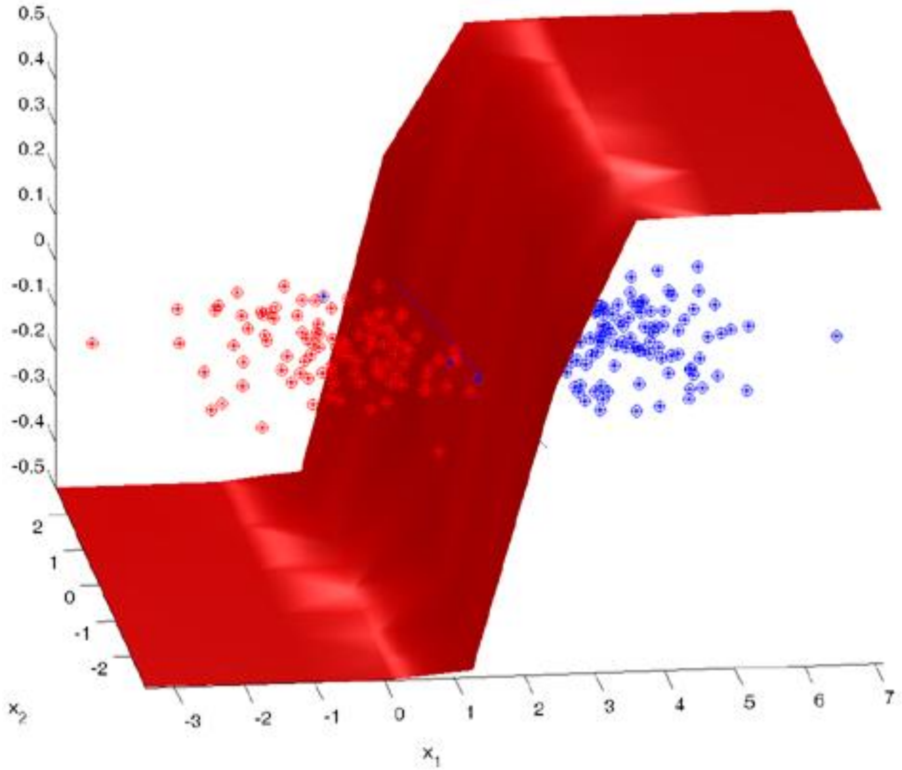


"soft threshold": reflektuje meru nesigurnosti

score	sigmoid(score)
$-\infty$	0
-2	0.12
0	0.5
2	0.88
$+\infty$	1

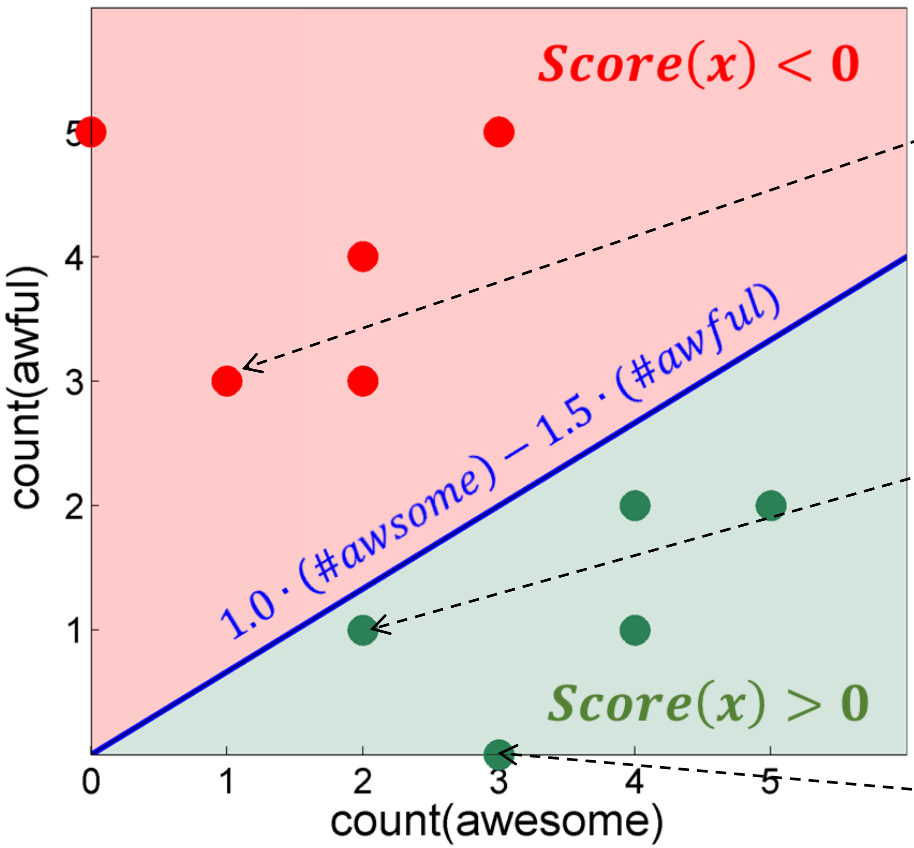
Logistička regresija – linearna granica odluke





Logistička regresija

$$P\left(y^{(i)} = 1 | x^{(i)}, \theta\right) = \sigma\left(\theta x^{(i)}\right) = \frac{1}{1 + e^{-\theta x^{(i)}}}$$



Primer iz klase 0
$score = -3.5$
$P\left(y = 1 | x^{(i)}\right) = 0.03$

Primer blizu granice odluke
$score = 0.5$
$P\left(y = 1 | x^{(i)}\right) = 0.62$

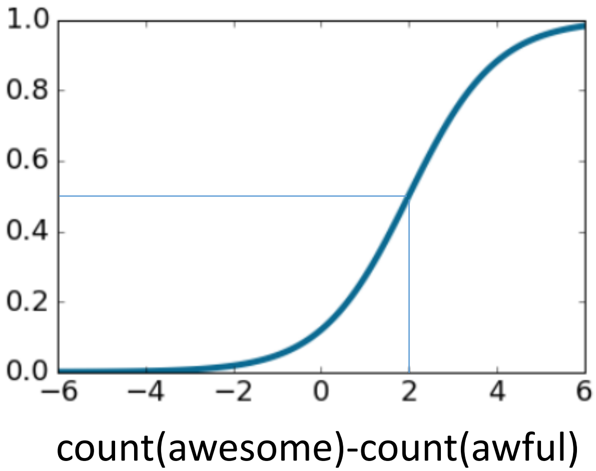
Primer daleko od granice odluke
$score = 3$
$P\left(y = 1 | x^{(i)}\right) = 0.95$

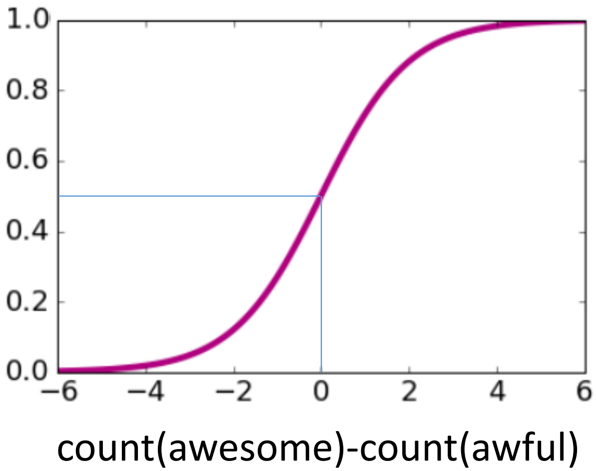
Uticaj θ na sigmoid

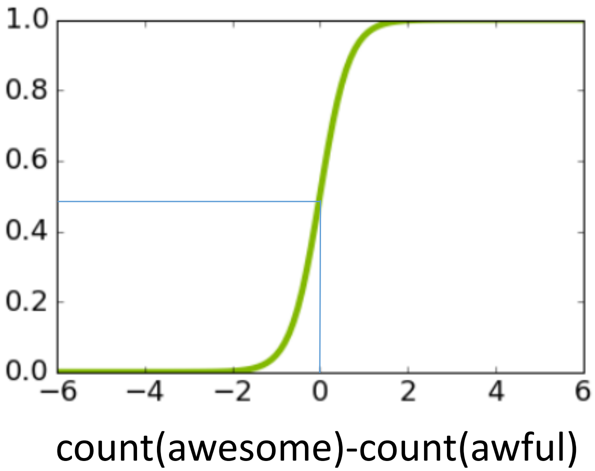
θ_0	-2
$\theta_{awesome}$	+1
θ_{awful}	-1

θ_0	0
$\theta_{awesome}$	+1
θ_{awful}	-1

θ_0	0
$\theta_{awesome}$	+3
θ_{awful}	-3







Sigmoid se pomerio u desno

Što je veća magnituda koeficijenata, kriva je strmija – brže postajemo sigurni u predikciju

Treniranje modela (određivanje θ)

- Trening skup $\mathrm{T} = \left\{\left(x^{(i)}, y^{(i)}\right)\right\}, i \in \{1, \dots, N\}$
- Svaki pimer opisan je sa *D* obeležja: $x = \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_D \end{bmatrix}, x_0 = 1$
- Ciljna varijabla $y \in \{1,0\}$
- Hipoteza:

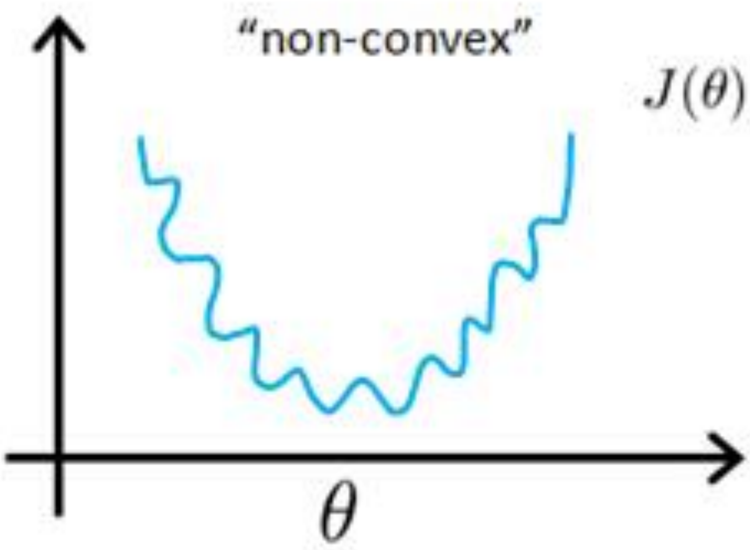
$$h_\theta(x) = \sigma(\theta x) = 1/\left(1 + e^{-\theta x}\right)$$

- Kako odrediti parametre θ?

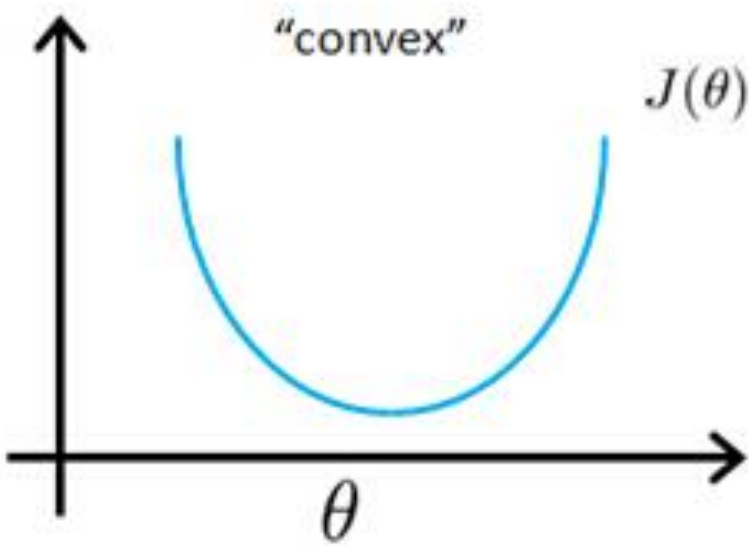
Treniranje modela (određivanje θ)

- Da li bismo mogli kao kod linearne regresije?

$$J(\theta) = \frac{1}{2N}\sum_{i=1}^{N}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$



Logistička regresija



Linearna regresija

Metod maksimalne verodostojnosti

- Treba da odaberemo θ tako da verovatnoća uočenog trening skupa bude najveća

- Idealno, želeli bismo da za svaki primer *x* iz skupa *T*:

$$h_\theta\left(x^{(i)}\right) = \begin{cases} 1, & \text{ako je } y^{(i)} = 1 \\ 0, & \text{ako je } y^{(i)} = 0 \end{cases}$$

- Željenu verovatnoću modelujemo kao

$$p_\theta(y|x) = h_\theta(x)^y \left(1 - h_\theta(x)\right)^{1-y}$$

Metod maksimalne verodostojnosti

- Funkcija verodostojnosti (pod pretpostavkom da su primeri generisani IID):

$$L(\theta) = \prod_{i=1}^{N} p_\theta\left(y^{(i)} | x^{(i)}, \theta\right)$$

- Cilj: naći θ za koje je verodostojnost $L(\theta)$ najveća

- Minimalizovaćemo negativan logaritam funkcije verodostojnosti

Metod maksimalne verodostojnosti

$$L(\theta) = \prod_{i=1}^{N} p_\theta\left(y^{(i)} | x^{(i)}, \theta\right)$$

- Negativna vrednost logaritma verodostojnosti:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{N} \log p_\theta\left(y^{(i)} | x^{(i)}, \theta\right)$$

$$= -\sum_{i=1}^{N} \log\left[h_\theta\left(x^{(i)}\right)^{y^{(i)}} \left(1 - h_\theta\left(x^{(i)}\right)\right)^{1-y^{(i)}}\right]$$

$$= -\sum_{i=1}^{N} \left[y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right]$$

- Optimizacioni problem: $\min_\theta \mathcal{L}(\theta)$

Cross-entropy error

Metod maksimalne verodostojnosti

- Optimizacioni problem: $\min_{\theta} \mathcal{L}(\theta)$
- Ne postoji closed-form solution
- Funkcija $\mathcal{L}(\theta)$ je konveksna pa ima jedinstven globalni minimum
- Za optimizaciju možemo upotrebiti metod gradijentnog spusta
- Obično se koristi Njutnova metoda

Gradijentni spust

- *Update rule*:

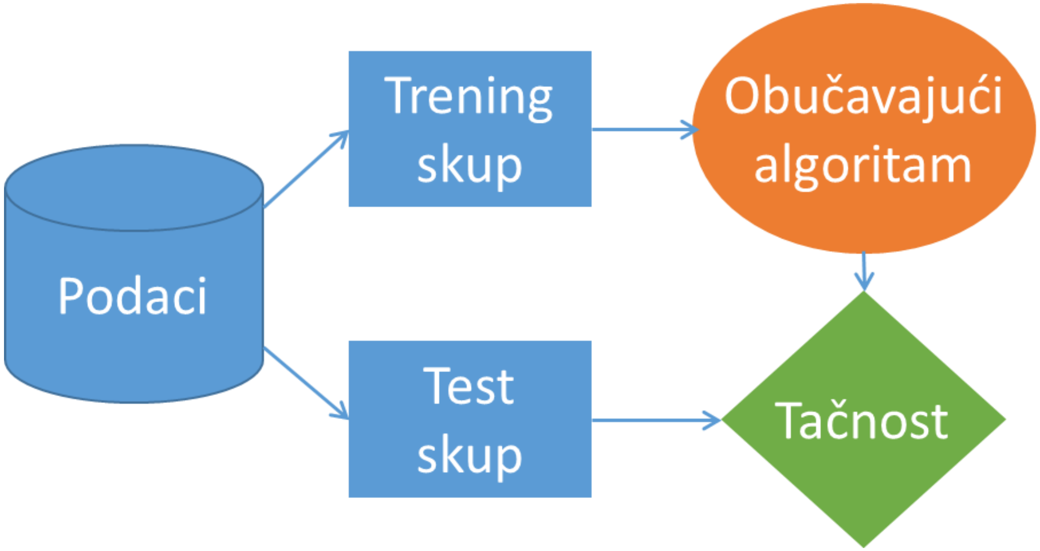
$$\frac{\partial}{\partial\theta}\mathcal{L}(\theta)$$

$$\theta_j^{(t+1)} = \theta_j^{(t)} - \alpha \sum_{i=1}^{N} \boxed{\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j{}^{(i)}}$$

- Izgleda identično linearnoj regresiji! Razlika je samo u obliku hipoteze $h_\theta\left(x^{(i)}\right)$

- Kao i kod linearne regresije
  - Konvergenciju možemo proveritu iscrtavanjem $\mathcal{L}(\theta)$ kao funcije broja iteracija – $\mathcal{L}(\theta)$ treba da se smanjuje sa svakom iteracijom
  - Bolje je normalizovati obeležja (brža konvergencija)

Evaluacija klasifikatora



- Tačnost (*accuracy*) na test skupu:

$$accuracy = \frac{count(correct)}{N_{test}}$$

(broj korektno klasifikovanih instanci test skupa podeljen sa ukupnim brojem instanci test skupa)