

# STROJNO UČENJE

## 3. Domaća Zadaća

Krešimir Špes

0036419866

ak. god. 2011. / 2012.

b) Model je diskriminativan, ali za razliku od generativnih modela, Logistička regresija izravno modelira aposteriornu vjerojatnost dok generativni

modeli tu vjerojatnost modeliraju posredno preko zajedničke gustoće.

c) što je stopa učenja veća, brže će algoritam konvergirati, ali kada dođe u okolinu minimuma funkcije, prevelika stopa učenja bi mogla uzrokovati velike oscilacije pa čak i divergirati. Dok mala stopa učenja povećava broj koraka potrebnih da se pronađe minimum funkcije. Kompromis rješenje je koristiti adaptivnu stopu učenja koja će se postepeno smanjivati kako se približava minimumu.

Što je veći N to je veća promjena težina, stoga bi stopu učenja trebalo korigirati ovisno o broju primjera, npr. podijeliti sa N.

Funkcija pogreške se može definirati kao očekivanje funkcije gubitka:

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

d)

pogreška unakrsne entropije:

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

$$E = -1 * \ln[h([1 \ 1 \ 1])] - \ln[1 - h([1 \ -1 \ -1])]$$

$$h([1 \ 1]) = \text{sigma}([1 \ 1]) = 0.5$$

$$E = 1.3863$$

gradijent:

$$\begin{aligned} \nabla E(\tilde{\mathbf{w}}) &= -\sum_{i=1}^N \left( \frac{y^{(i)}}{h(\mathbf{x}^{(i)})} - \frac{1 - y^{(i)}}{1 - h(\mathbf{x}^{(i)})} \right) h(\mathbf{x}^{(i)}) (1 - h(\mathbf{x}^{(i)})) \tilde{\mathbf{x}}^{(i)} \\ &= \sum_{i=1}^N \underbrace{(h(\mathbf{x}^{(i)}) - y^{(i)})}_{\nabla E_i(\tilde{\mathbf{w}})} \tilde{\mathbf{x}}^{(i)} \end{aligned}$$

$$\begin{aligned} \text{gradijent} &= (0.5 - 1) * [1 \ 1 \ 1] + (0.5 - 0) * [1 \ -1 \ -1] = [-0.5 \ -0.5 \ -0.5] + [0.5 \ -0.5 \ -0.5] \\ &= [0 \ -1 \ -1] \end{aligned}$$

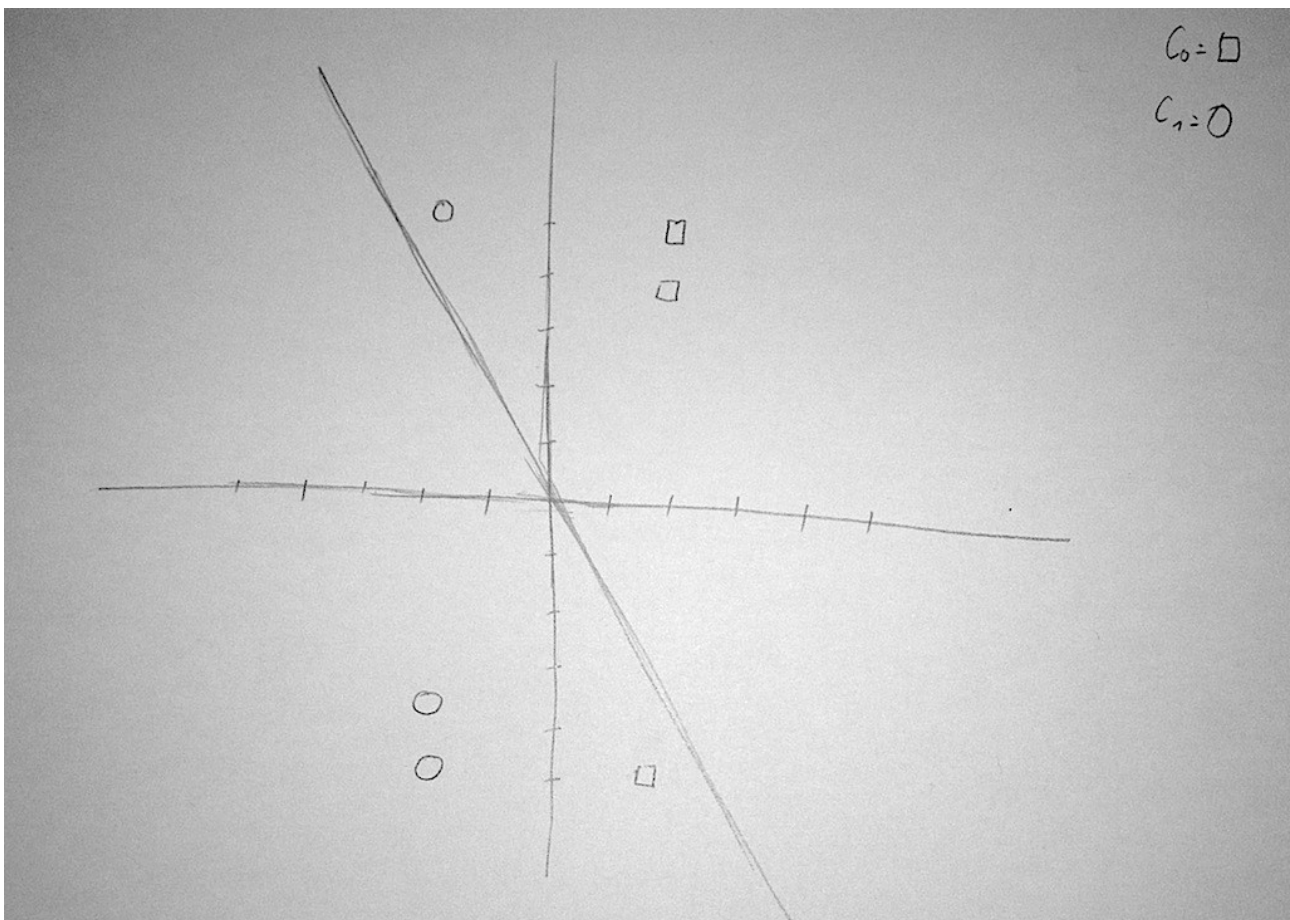
konvergirat će jer je već [0 -1 -1] rješenje koje odvađa uspješno te primjere.

( to je pravac  $y = -x$  )

e) hoće jer je u prvom koraku  $w == [0 \ 0 \ 0]$  te time utjecaj regularizacije u prvom koraku nema efekta a pošto nakon prvog koraka algoritam konvergira, nikad ni neće biti problem :)

f) empirijska pogreška će rasti porastom vrijednosti parametra regularizacije jer ograničavanjem složenosti modela smanjujemo i moć klasifikatora da se bolje prilagodi primjerima.

g) crtež:



logistička regresija u ovom slučaju nije uspjela odvojiti razrede iako su oni linearno odvojivi zbog veze logističke regresije sa generativnim modelom sa gausovom razdiobom i dijeljenom kovarijacijskom matricom. Dakle klasifikator pretpostavlja da će se distribucija primjera pokoravati gausovoj distribuciji što ovdje ili nije slučaj ili imamo premalo primjera da možemo točnije odrediti parametre gausovih razdioba.