

Drugi domaći zadatak

Ljubica Vujović
Profesor: Predrag Tadić
Statistička klasifikacija signala

Oktobar 2017

1 Logistička regresija

Logistička regresija je model pomoću koga se može vršiti klasifikacija. Ovaj model meri vezu između kategoričke zavisne promenljive i jedne ili više nezavisnih promenljivih tako što procenjuje verovatnoće koristeći logističku funkciju. Hipoteza ovog modela je:

$$h_{\theta}(x) = g(\theta x) = \frac{1}{1 + e^{-\theta x}} \quad (1)$$

Obučavanjem ovog modela datim setom podataka dobijena je sledeća klasifikaciona linija:

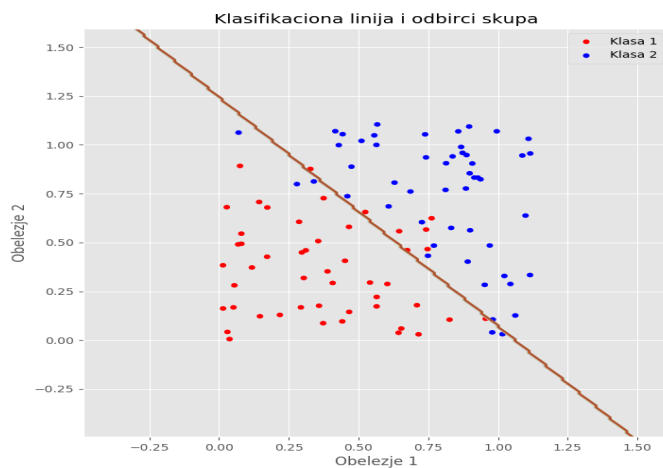


Figure 1: *Klasifikaciona linija dobijena logističkom regresijom*

Vrednost parametara za koje se postiže ova kriva su nakon 20000 iteracija korakom $\alpha = 0.001$:

$$\theta_1 = -10.34965468$$

$$\theta_2 = 9.72168654$$

$$\theta_3 = 8.30658873$$

Preciznost ovog klasifikatora je 0.88 za vrednost praga od 0.5, a broj lažnih jedinica iznosi 7 dok broj lažnih nula iznosi 5.

Promenom praga τ od 0.1 do 0.9 dobijena je i sledeća ROC kriva:

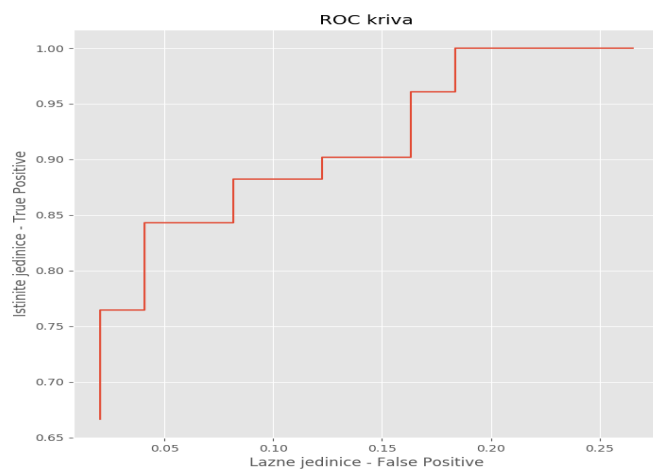


Figure 2: *ROC kriva*

Na sledećoj slici prikazana je kriterijumska funkcija kroz iteracije.

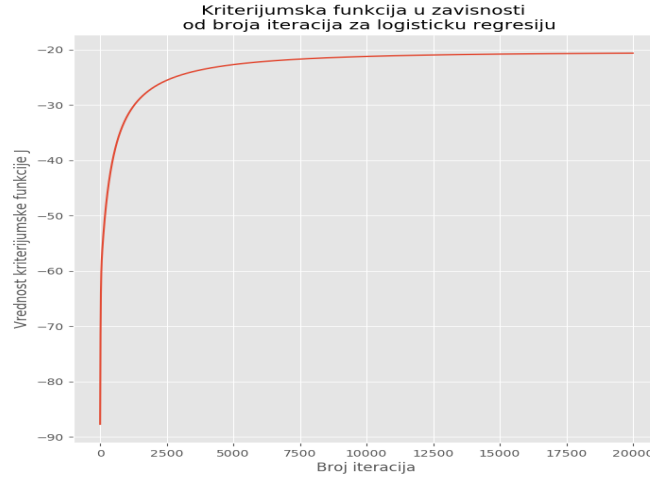


Figure 3: *Kriterijumska funkcija*

Zanimljivo je primetiti da u zavisnosti od broja iteracija dobijamo parametre iste parametre θ samo skalirane. U sledećoj tabeli prikazano je kako se menjaju spomenuti parametri i preciznost u zavisnosti od broja iteracija.

Broj iteracija	θ_1	θ_2	θ_3	Preciznost
1000	-3.5	3.4	2.9	0.89
10000	-8.8	8.4	7.1	0.88
100000	-11.8	11.1	9.5	0.88

2 Softmax regresija

Softmax regresija predstavlja generaliziju logističke regresije u smislu povećavanja broja klasa. U logističkoj regresiji podrazumevalo da je izlazna promenljiva binarna, dok je softmax model sposoban da klasifikuje i veći broj klasa. Neka je k broj klasa koje želimo da klasifikujemo i neka su $\theta^{(j)}$ parametri koji su dodeljeni klasi j . Tada hipoteza vezana za klasu k izgleda:

$$h_{\theta^{(j)}}(x) = \frac{\exp(\theta^{(j)}x)}{\sum_i \exp(\theta^{(i)}x)} \quad (2)$$

Obučavanjem softmax modela dobijamo preciznost od 0.983. Na sledećoj slici su prikazani odbirci sve tri klase i odbirci koji su loše klasifikovani.

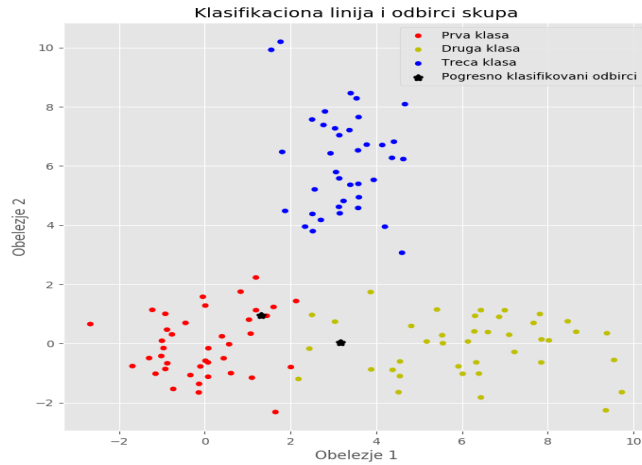


Figure 4: *Odbirci različitih klasa*

Ovi rezultati dobijeni su korakom od $\alpha = 0.001$ sa 1000 iteracija gradijentnom metodom. Na sledećoj slici prikazana je vrednost kriterijumske funkcije kroz iteracije.

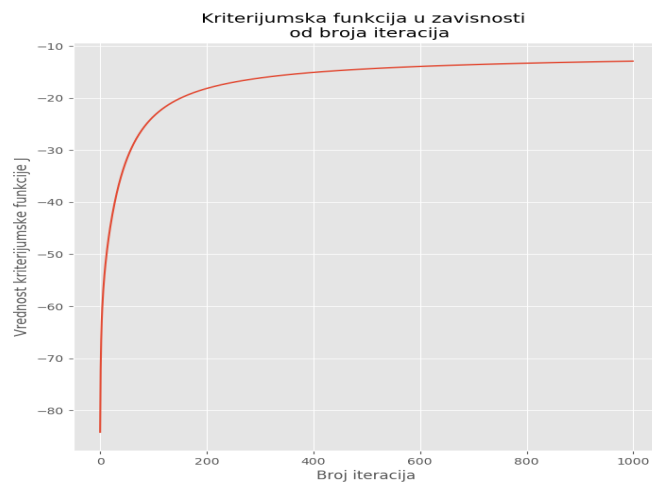


Figure 5: *Kriterijumska funkcija*

Ovom metodom direktno idemo ka minimumu funkcije, što se može videti i na sledećem grafiku.