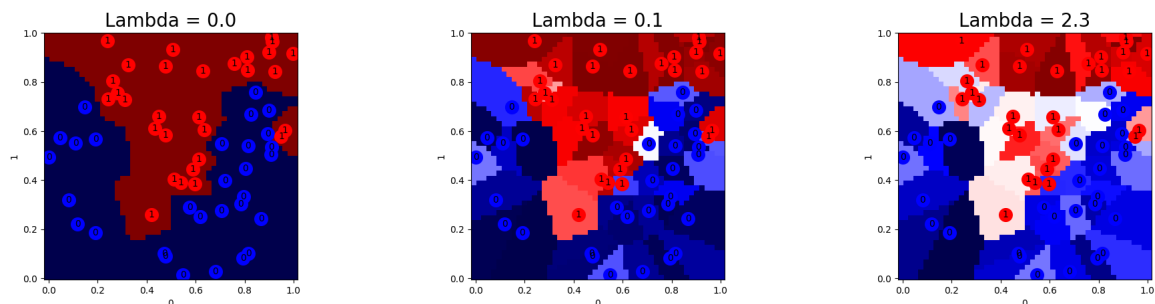


1 Uvod

Cilj tretje naloge pri predmetu Uvod v odkrivanje znanj iz podatkov je bil implementacija in evalvacija binarnega klasifikatorja z uporabo logistične regresije, ki je prikladen način za izračun verjetnosti pripadnosti vektorja značilk določenemu primeru. Pri implementaciji je bil preučen tudi vpliv različnih stopenj regularizacije na točnost, ki je bila ocenjena tako z deležem pravilnih napovedi (funkcija CA) kot tudi z ploščino pod krivuljo ROC (AUC).

2 Regularizacija

Vpeljava regularizacije nam omogoča, da preprečimo pretirano rast regresijskih koeficientov s tem, da velike koeficiente v kriterijski funkciji "kaznujemo". To "kazen" utežimo s koeficientom λ , katerega vrednost ima lahko velik vpliv na natančnost klasifikatorja. Rezultati klasifikacije z različno vrednostjo koeficienta λ so prikazani spodaj.



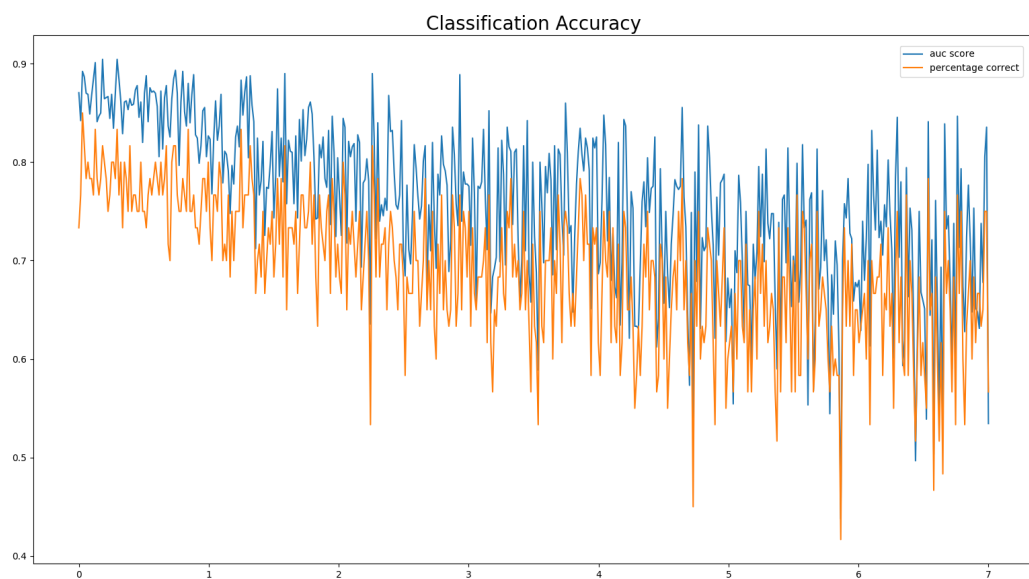
Slika 1: Vizualizacija klasifikacije z različnimi stopnjami regularizacije

3 Točnost

Točnost klasifikatorjev z različno stopnjo regularizacije je bila ocenjena tako z deležem pravilnih napovedi pri prečnem preverjanju kot tudi z mero AUC, ki predstavlja ploščino pod krivuljo ROC. Izkaže se, da je AUC enaka verjetnosti, da bo klasifikator (ki zna napovedovati verjetnosti) pravilno razločil med pozitivnim in negativnim primerom, t.j. pozitivnemu bo pripisal večjo verjetnost, da je pozitiven. (Kononenko in Šikonja, 2010)¹

Vpliv stopnje regularizacije lahko vizualiziramo z grafom, ki prikazuje odvisnost med stopnjo regularizacije in klasifikacijsko točnostjo. Takšen graf je prikazan spodaj.

¹Kononenko, I., Šikonja M. Inteligentni sistemi. Ljubljana: Založba FE in FRI, 2010.



Slika 2: Klasifikacijska točnost v odvisnosti od stopnje regularizacije

Rezultate lahko prikažemo tudi s tabelo.

stopnja regularizacije (λ)	0	0.1	0.1	0.3	0.4	0.5	0.6	0.7	0.9	1
AUC	0.89	0.82	0.89	0.9	0.85	0.84	0.89	0.88	0.8	0.84
Delež pravih klasifikacij	0.82	0.76	0.82	0.75	0.77	0.82	0.82	0.73	0.78	0.73

Tabela 1: Klasifikacijska točnost za različne stopnje regularizacije

4 Izjava o izdelavi domače naloge.

Domačo nalogo in pripadajoče programe sem izdelal sam.