

Logistična regeresija

Tomaž Tomažič (63100281))

19. november 2012

1 Uvod

Cilj naloge je bil implementirati logistično regresijo in izgraditi model na določenih podatkih.

2 Podatki

Primer podatkov je v datotetki z 28 atributi in 60 primeri. Zadnji stolpec predstavlja skupino kateri posamezen primer pripada. Možni vrednosti sta le 0 in 1. Testiral sem tudi obsežnejše podatke z imenom GDS1059.

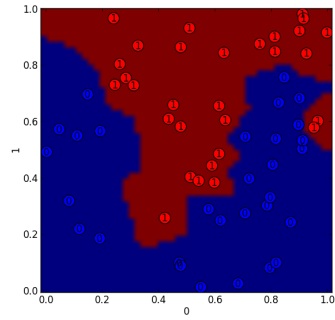
3 Metode

Nalogo sem rešil tako da sem uporabil priloženo ogrodje in ga smiselno dopolnil. Test sem pognal na različnih vrednostih λ , da sem lahko raziskal vpliv regularizacije na napovedi. Zanimive stopnje regularizacije so mi bile pri $\lambda=0.0005$ 1, $\lambda=0.05$ 2 in pri $\lambda=0.5$ 3. Pri majhnem λ je razvidno, da se prilagodimo podatkom v preveliki meri. Taka funkcija nam na prvi pogled izgleda optimalna, vendar ni nujno najbolj primerna za napovedovanje. Nekoliko boljše izgleda 2 slika 2, sej je tu dovoljeno manjše odstopanje. Preveč regularizirana funkcija pa je pri $\lambda=0.5$ 3, saj niti ne ohranja intuitivne oblike.

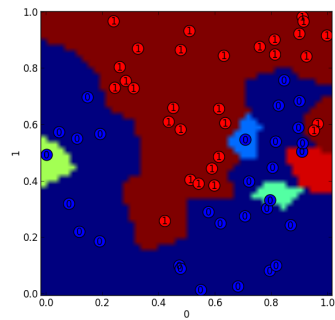
4 Rezultati

S K-kratnim prečnim preverjanjem, sem testiral gradnjo modela in napovedovanje istih primerov, ki so v vhodnih podatkih. Preverjanje sem izvedel za različen nabor vrednosti λ . Rezultati so prikazani v tabeli 1. Razvidno je, da točnost napovedovanja pri različnih λ niha. Nekje ima najvišjo vrednost, nato pa začne točnost spet upadati.

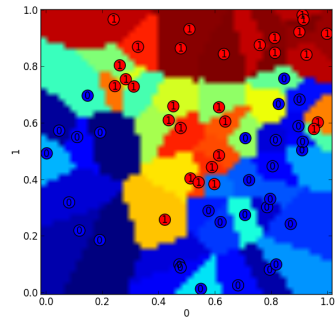
Za podatke GDS1059 sem dobil nalednje točnosti, ki so prikazane v tabeli 2. Stopnjo regularizacije ki bi bila najbolj primerna, nisem uspel najti, saj pri različnih λ (0.0005 4, 0.005 5, 0.5 6) dobim zelo podobne rezultate. Smiselno bi bilo opraviti permutacijski test in preveriti ali gre za naključne podatke.



Slika 1: Izris podatkov v 2 dimenzijah pri določenem lambda



Slika 2: Izris podatkov v 2 dimenzijah pri določenem lambda



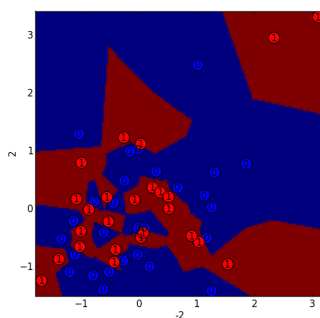
Slika 3: Izris podatkov v 2 dimenzijah pri določenem lambda

Tabela 1: K kratno preevrjanje točnosti.

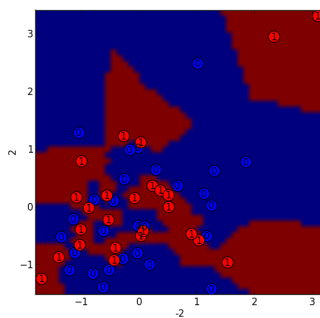
točnost	vrednost lambda
0.74442447876	1e-05
0.733398432863	5e-05
0.783135711911	0.00025
0.730327654345	0.00125
0.734638081205	0.00625
0.750608334012	0.03125
0.727649151841	0.15625
0.663830170228	0.78125

Tabela 2: K kratno preevrjanje točnosti v podatkih GDS1059.

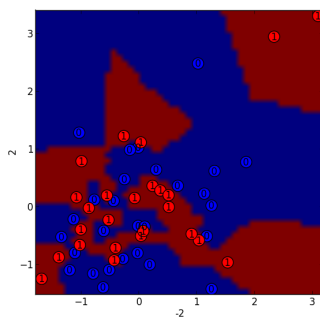
točnost	vrednost lambda
0.768021461362	1e-05
0.767515771074	5e-05
0.767690926141	0.00025
0.768451761935	0.00125
0.759956150213	0.00625
0.760401419028	0.03125
0.741364988033	0.15625
0.718926511778	0.78125



Slika 4: Izris podatkov v 2 dimenzijah pri določenem lambda



Slika 5: Izris podatkov v 2 dimenzijah pri določenem lambda



Slika 6: Izris podatkov v 2 dimenzijah pri določenem lambda

5 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.