

4. domača naloga: Softmax regresija

Luka Krsnik (63110179)

20. april 2015

1 Uvod

Cilj četrte domače naloge je implementacija metod, ki se iz učnih podatkov čim bolj naučijo določiti razred na testnim primerom.

2 Metode

Logistična regresija Logistična regresija je postopek, s pomočjo katerega je na podlagi učnih podatkov zgrajena "funkcija", ki iz značilk čim bolj klasificira določen primer v nek razred oziroma določi s kakšno verjetnostjo nek primer spada v določen razred. Zanj je značilno, da lahko odloča le med dvema razredoma (je binarni klasifikator).

Eden proti vsem Eden proti vsem je eden izmed možnih postopkov, ki je potreben, če želimo logistično regresijo uporabiti na več kot dveh razredih. Deluje tako, da učne primere razdeli na tiste, ki pripadajo željenemu razredu in na določene ostale. Ta postopek se ponovi za vse razrede, nato pa se nad vsakim izvede logistična regresija. Tako pridobimo θ za vsak razred. Nazadnje zmnožek s θ in testnimi primeri vstavimo v sigmoidno funkcijo ter tako dobimo napovedi.

Eden proti enemu Ta postopek je druga možna izbira, če želimo uporabiti logistično regresijo. Deluje tako, da vedno primerjamo dva razreda. Tisti, ki ima večjo verjetnost žmaga, zato se mu prišteje en glas. Na ta način preverjamo vsak razred z vsakim. Nazadnje preštajemo glasove, ter na njihovi podlagi predvidimo kam bo padel določen primer.

Softmax regresija Softmax regresija je postopek, ki je posplošitev logistične regresije. Posplošitev v smislu, da omogoča klasifikacijo med večimi razredi hkrati, torej brez uporabe metod kot sta eden proti vsem in eden proti enemu.

3 Rezultati

Rezultati so za določene metode (softmax in one vs one) precej slabi. V primeru softmaxa se mi zdi da so takšni zato, ker nisem izbral zelo dobre λ (prečnega preverjanja za iskanje parametrov namreč tokrat nisem implementiral). V mojem primeru je ta namreč tako pri

logistični, kot tudi pri softamxu enaka 0.01. Razlog zakaj so slabši rezultati tudi pri one vs one pa bi pripisal temu, da se pri učenju podatkov učimo iz različno velikih razredov.

Tabela 1: Tabela z rezultati metode cross validation.

metoda	rezultat
Logistična Regresija + Eden proti vsem	0.68
Logistična Regresija + Eden proti enemu	2.19
Softmax Regresija	1.61

Končna oddaja je oddaja s Softmax regresijo.

4 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.