

Treći domaći zadatak

Ljubica Vujović
Profesor: Predrag Tadić
Statistička klasifikacija signala

Oktobar 2017

1 Naivni Bayesov klasifikator

Naivni Bayes je algoritam za formiranje klasifikatora. Ovaj algoritam se oslanja na relativno jaku pretpostavku da su različita obeležja medjusobno nezavisna, uslovljena izlaznom promenljivom. Ova pretpostavka često nije legitimna, međutim algoritmi koji se oslanjaju na nju su efikasni za treniranje i ne zahtevaju mnogo odbiraka.

Zadatak je isprojektovati Bayes-ov klasifikator, pod pretpostavkom da su odbirci klasa normalno raspodeljeni. Na sledećoj slici prikazani odbirci klasa. Vidimo da imamo tri klase i dva obeležja, što implicira da je će vektor srednjih vrednosti koje treba da naučimo imati dimenzije 2×3 a vektor varijansi 2×3 .

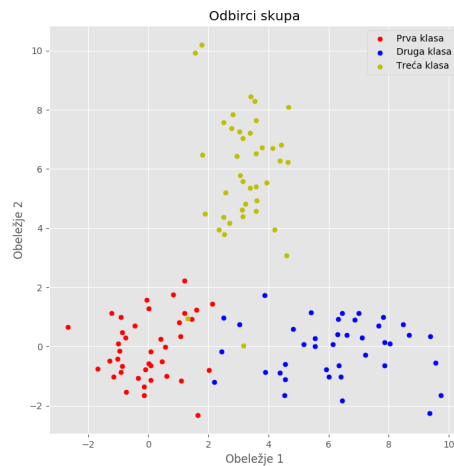


Figure 1: *Odbirci skupa*

Obučavanjem ovog modela na dati skup podataka dobijamo sledeće parametre:

μ	0	1	2
0	0.01	6.20	3.19
1	-0.04	-0.07	5.85

σ^2	0	1	2
0	1.15	3.81	0.70
1	1.12	0.88	4.17

Preciznost ove metode iznosi 0.97.

2 Gausova diskriminantna analiza

Gausova diskriminantna analiza je model koji pretpostavlja da su odbirci klase normalno rasporeeni. Parametri ovog modela su matrica srednjih vrednosti μ i jedna kovarijaciona matrica Σ . GDA model ima sličnosti sa linearnom regresijom. Ako verovatnoću modelovanu normalnom distribucijom, zamenimo sigmoid funkcijom, vidimo da oba modela imaju istu formu. Razlika je u tome što Gausova diskriminantna analiza pravi jaču pretpostavku time što modeluje verovatnoće kao normalnu raspodelu. Ukoliko klase zaista jesu normalno raspodeljene sa istom kovarijacionom matricom, GDA će naći bolje rešenje.

Parametri ovog modela su:

μ	0	1	2
0	0.01	6.20	3.19
1	-0.04	-0.07	5.85

Σ	0	1
0	1.88	0.04
1	0.04	2.06

Preciznost ove metode iznosi 0.95.

3 Gausova diskriminantna analiza sa različitim kovarijacionim matricama

Gausova diskriminantna analiza se može priminiti i sa pretpostavkom sa svaka klasa ima različitu kovarijacionu matricu. Očekivanje ovakvog modela je da će preciznost biti veća. Obučavanjem su dobijeni sledeći parametri:

μ	0	1	2
0	0.01	6.20	3.19
1	-0.04	-0.07	5.85

Σ_1	0	1
0	1.15	0.18
1	0.18	1.12

Σ_2	0	1
0	3.81	-0.11
1	-0.11	0.88

Σ_3	0	1
0	0.70	0.05
1	0.05	4.17

Tačnost ove metode je 0.97.

4 Uporedna analiza algoritama

Vidimo da iste performanse imaju GDA sa različitim kovarijacionim matricama i Naivni Bayes a da je GDA sa jedinstvenom matricom nešto lošija po performansama. Iako je možda očekivanje da Naivni Bayes daje lošije rezultate zato što uvodi dosta jaku pretpostavku o uslovnoj nezavisnosti obeležja, ako bolje pogledamo odbirke seta podataka, zaista deluje da je ova pretpostavka opravdana. Sa druge strane ako pogledamo kovarijacione matrice drugog GDA modela, vidimo da se glavna dijagonala gotovo poklapa sa varijansama kod Naivnog Bayes-a a da su vrednosti kovarijanse male što implicira da su obeležja slabo korelisana. To naravno ne implicira nezavisnost, ali na intuitivnom nivou daje objašnjenje zašto jednostavan model kao što je Naivni Bayes radi dobro.