## Naive Bayes

### Način implementacije algoritma

Dataset je najpre podeljen na dataset-ove tako da u jednom dataset-u budu uzorci koji pripadaju samo jednoj klasi:

1. Korak – Računanje srednje vrednosti(promenljiva var\_means) i standardne devijacije(promenljiva variances) .Ove vrednosti se koriste u Gausovoj funkciji gustine verovatnoće. Ovo se radi zato što naivni Bajes pretpostavlja da podaci prate Gausovu distribuciju.
2. Korak – Definisanje Gausove funkcije gustine verovatnoće. Ova funkcija se koristi za procenu verovatnoće vrednosti nekon feature-a za datu klasu. U mojoj implementaciji algoritma naziv funkcije je log\_gaussian\_pds
3. Korak – Računanje uslovne verovatnoće klase. Za novu instancu, izračunava se proizvod verovatnoće za sve vrednosti feature-a. To se radi za svaku klasu. Klasne verovatnoće su predstavljene promenljivom test\_set\_probabilities.
4. Korak – Predviđanje(predictions) – Da bi se napravilo predviđanje za novu instancu, sabiraju se verovatnoće za svaku klasu, umesto množenja jer množenje mnogo malih vrednosti može dati rezultat 0.

### Skup podataka

Skup podataka koji je korišćen je iz biblioteke sklearn. Namenjen je za binarnu klasifikaciju. Labela govori da li je rak dojke magni ili benigni. Sadrži 455 uzoraka i 30 atributa.

### Zaključak

Tačnost predviđanja sa mojom implementacijom algoritma je 76.31%.

Za klasu 0 preciznost je 0.77 što znači da od svih negativno predviđenih uzoraka, 77% je zaista negativno. Recall je 0.53, što znači da je od svih negativnih uzoraka 53% predviđeno ispravno (Figure 1).

Za klasu 1 preciznost je 0.76% što znači da je od svih pozitivno predviđenih uzoraka, 76% zaista pozitivno. Recal je 0.90, što znači da je od svih stvarno pozitivnih uzoraka 90% ispravno kalsifikovano (Figure 2).

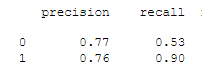


Figure 1

