Naive Bayes

The version implemented here is Gaussian Naïve Bayes.

This algorithm assumes that the data is normally distributed and the features are conditionally independent. Let’s say there are ‘n’ features. Then,

In our case, there are two classes and two features.

During training we will find the mean and standard deviation of the samples belonging to different classes.

is the mean of samples belonging to class ‘i’.

is the standard deviation of the samples belonging to class ‘i’.

Both will be vectors of size ‘n’ where ‘n’ is the number of features

During prediction, we will calculate the probability , where is the ‘j’th feature, for all the features and classes.

Where,

is the ‘j’th feature of the ‘i’th sample.

is the mean of the ‘j’th feature of the samples belonging to class y=k

is the standard deviation of the ‘j’th feature of the samples belonging to class y=k

Then for each sample we will find

P(y) is called class prior, which can be specified or also can be calculated from the data. To calculate, it is just the fraction of the sample that belongs to the class y=i.

We will return the y for which the above expression is maximum.