**PRACTICAL 8**

**Aim:** Implementation and analysis of Classification algorithms like Naive Bayesian,K-Nearest Neighbor.

**Theory**:

**Classification**

Implementation and analysis of Classification algorithms like

1. Naive Bayesian,

2. K-Nearest Neighbor

**Naive Bayes • Based on the Bayes theorem**

1. It is a classification technique based on Bayes’ Theorem with an assumption of independence

among predictors.

2. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

3. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in

diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as ‘Naive’.

4. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with

simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

• Predicts based on probabilities from training data

P(B|A) = P(A|B) P(B)/P(A)

Gives posterior probability of ‘B’ given ‘A’ using

prior probability of ‘B’

prior probability of ‘A’

and conditional probability of ‘A’ given ‘B’

• Takes two step approach

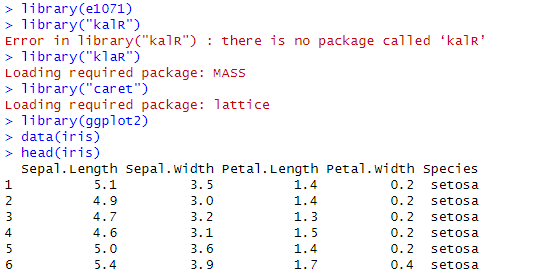
– Calculates the posterior probability of the Class given the input - for every class

– Assigns the class with higher posterior probability

• More suited when dimensionality of input is high the - widely used for document classification

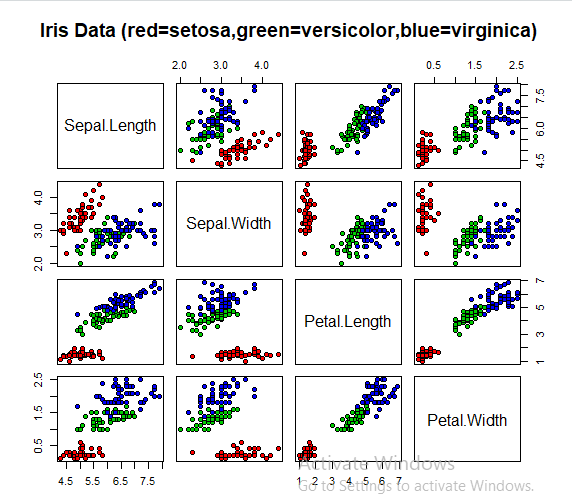
• Also good for the multiclass classifications

• Works well with less datasets also, but the assumption that predictor variables are independent should hold





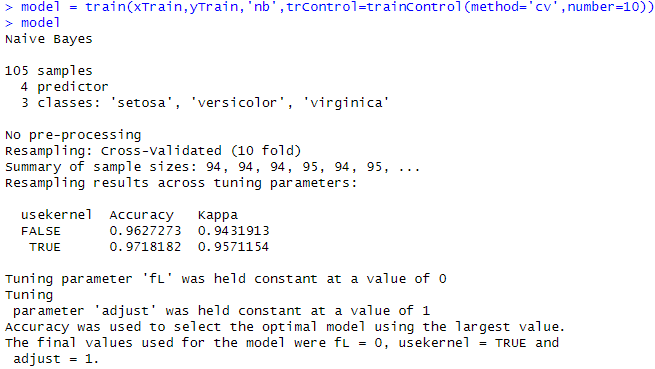


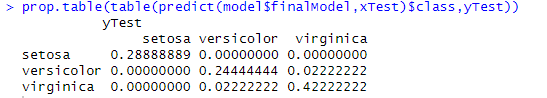












**K nearest Neighbour**

1. K-NN algorithm assumes the similarity between the new case/data and available cases and put the

new case into the category that is most similar to the available categories.

2. K-NN algorithm stores all the available data and classifies a new data point based on the

similarity. This means when new data appears then it can be easily classified into a well suite

category by using K- NN algorithm.

3. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for

the Classification problems.

4. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying

data.

5. It is also called a lazy learner algorithm because it does not learn from the training set immediately

instead it stores the dataset and at the time of classification, it performs an action on the dataset.

6. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it

classifies that data into a category that is much similar to the new data.

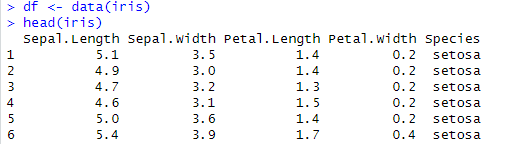
7. Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want

to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it

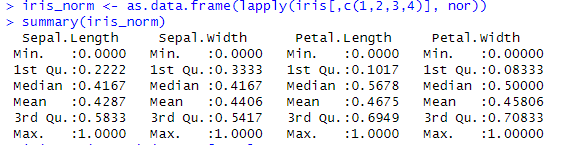
works on a similarity measure. Our KNN model will find the similar features of the new data set

to the cats and dogs images and based on the most similar features it will put it in either cat or dog

Category.

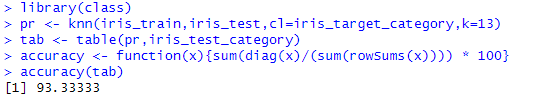
****

****

****

****

****

****

**Conclusion:** Successfully implemented classification algorithms like Naïve Bayes and K Nearest neighbor algorithms.