Assignment No.6

Name- Kudharan Sumit Dattatraya

Class - TE

Div - 4

Subject - DSBDAL

Problem statement-

1. Implement simple Naive Bayes classification algorithm using Python/R on iris.csv dataset.
2. Compute confusion matrix to find TP, FP, TN, FN, Accuracy, Error rate, Precision, Recall on the given dataset.

Theory -

1) Explain different classification algorithm.

→ ① Logistic Regression -

It is a machine learning algorithm for classification. In this algorithm, the probabilities describing the possible outcomes of a single trial are modelled using a logistic function.

Advantages -

It is designed for the classification & is most useful for understanding the influence of several independent variables on a single outcome variable.

Disadvantages -

Works only when the predicted variable is binary.

② Naive Bayes -

Naive Bayes algorithm based on Bayes theorem with the assumption of independence between every pair of features. Naive Bayes classifiers work well in many real world situations such as document classification & spam filtering

Advantages - Naive Bayes work very fast compared to more sophasticated methods.

Disadvantages - Naive Bayes is know to be abad estimator.

③ K-Nearest Neighbours -

Neighbours based classification is a type of lazy learning as it does not attempt to construct a general internal model, but simply store instances of the training data. classification is computed from a simple majority vote of the k nearest neighbours of each point.

Advantages - This algorithm is simple to implement & it effective if training data is large.

Disadvantages - Need to determine the value of K & the computation cost is high

④ Support Vector Machine (SVM) -

SVM is a representation of the training data as points in space sperated into categories by a clear gap that is as wide as possible.

Advantages - Effective in high dimensional spaces & uses a subset of training points in the decision function.

Disadvantages - The algorithm does not directly provide probability estimates.

⑤ Decision Tree -

Decision tree produces a sequence of rules that can be used to classify the data.

Advantages - simple to understand & visualize

Disadvantages - It can create complex trees that do not generalise well.

2) explain bayes theorem.

→ Baye's Theorem is the basic foundation of probobility. It is the determination of the conditional probability of an event. This conditional probability is knocon as a hypothesis. This hypothesis is calculated through previous evidence or knowledge. The conditional probability is the probobility of the occurrance of an event given that some other event has already happened.

The formula of Bayes Theorem involves the posterior the probability P(H|E) as the product of the probability of hypothesis P(E|H), multiplied by the probability of the hypothesis P(H) and divided by the probability of the evidence P(E).

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)}$$

Here,

P(H|E) – This is referred to as the posterior probability.

P(E|H) - Denotes the likelihood.

P(H) - Referred as the prior probability.

P(E) - This is the probability of the occurrance of evidence regardless of the hypothesis.

3) Explain Naive bayes theorem.

→ It is a family of probabilistic classifiers based on bayes theorem. It uses the relationship between the probabilities of event for classification.

Bayes Theorem -

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$