

Assignment No. 6.



Date

Title:-

1. Implement simple naive Bayes classification algo using python on iris.csv dataset.
2. compute confusion matrix to find TP, FP, TN, FN accuracy, error, rate precision, recall on the given dataset.

Objective:-

student should be able to data analysis using Naive Bayes algorithm using python for any open source dataset.

Prerequisite:-

- 1) Basic of python programming.
- 2) concept of joint & marginal probability.

Theory:-

Naive Bayes classifier:-

Naive Bayes classifier are a family of simple 'probabilistic classifier' based on applying Bayes theorem with independence assumption between the features.

$$P\left(\frac{Y}{x_1, x_2, \dots, x_n}\right) = \frac{P(Y) \cdot \frac{P(x_1, x_2, \dots, x_n)}{Y}}{P(x_1, x_2, \dots, x_n)}$$

$$P\left(\frac{x_i}{Y} \mid x_1, x_2, \dots, x_n\right) = P\left(\frac{x_i}{Y}\right), \forall i$$

$$P\left(\frac{y}{x_1, x_2, \dots, x_n}\right) = \frac{P(y) \cdot \prod_{i=1}^n P\left(\frac{x_i}{y}\right)}{P(x_1, x_2, \dots, x_n)}$$

$$P\left(\frac{y}{x_1, x_2, \dots, x_n}\right) \propto P(y) \prod_{i=1}^n P\left(\frac{x_i}{y}\right)$$

= confusion matrix:-

It is used to judge performance of classifier on test dataset. confusion matrix is also termed as error matrix. it contains count of correct & incorrect values.

= TP:- model correctly predicts positive class.

TN:- correctly predict negative class.

FP:- wrong predictⁿ of negative class

FN:- wrong predictⁿ of positive class.

= conclusion:- In this manner we have successfully implemented Naive Bayes classifiers algorithm & computed confusion matrix.