Group-A

COTB031

Date

Experiment No-6

Title: 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.

Objective: Students should be able to data analysis using Naive Bayes classification algorithm using python for any open source dataset.

Prerequisite:
1. Basic of Python programming
2. Concept of Joint & Marginal probability.

1. Concepts used in Naive Bayes classifier Naive Bayes classifier can be used for classification of categorical data.
- Let there be a 'J' number of classes $c = \{1, 2, \cdots J\}$
- Let, input observation is specified by 'p' features therefore input observation x is given, $x = \{F_1, F_2, \cdots F_p\}$
- The Naive Bayes classifier depends on Bayes rule from probability theory.

Conditional probabilities

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) \neq 0 \qquad \cdots (1)$$

$$P\left(\frac{B}{A}\right) = \frac{P(B \cap A)}{P(A)} \qquad \cdots (2)$$

Date

from eqn ① & ②

$$P(A \cap B) = P\left(\frac{A}{B}\right) \cdot P(B) = P\left(\frac{B}{A}\right) \cdot P(A)$$

$$\therefore P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) \cdot P(A)}{P(B)}$$

Conditional probability

Here, we are predicting the probability of class 1 and class 2 based on the given condition. If I try to write the same formula in terms of classes & features, we will get the following equation.

$$P(k/x) = \frac{P(x1(x) * P(k))}{P(x)}$$

$$P(C_i | x_1 \cap x_2 \cap x_3 \cap x_4) = \frac{P(x_1 \cap x_2 \cap x_3 \cap x_4 | C_i) * P(C_i)}{P(x_1 \cap x_2 \cap x_3 \cap x_4)}$$

Conclusion: In this way we have done data analysis using naive Bayes algorithm for iris dataset & evaluated the performance of the model.