Machine Learning
Experiment 8

SAP ID: 60004200107 Name: Kartik Jolapara
Division: B Batch: B1

`AIM
To implement Bayesian Classification.

THEORY
Bayesian classification is a probabilistic approach to machine learning that is used for classification tasks. It is based on Bayes' theorem, which describes the probability of an event occurring based on prior knowledge of conditions that might be related to the event. In Bayesian classification, the goal is to assign a given data point to one of several predefined classes. The classification is based on a probabilistic model that is learned from a training set of labeled data. The model assigns a probability to each class for a given data point, and the class with the highest probability is chosen as the predicted class for the data point.
The probabilistic model used in Bayesian classification is typically a Bayesian network, which is a graphical model that represents the dependencies between variables in a probabilistic way. Each node in the network represents a random variable, and the edges between nodes represent the conditional dependencies between variables. The network is learned from the training data using a maximum likelihood or maximum a posteriori estimation approach. Once the network is learned, Bayesian classification works by computing the posterior probability of each class for a given data point using Bayes' theorem:

$$P(C|X) = P(X|C) * P(C) / P(X)$$

where P(C|X) is the posterior probability of class C given the data point X, P(X|C) is the likelihood of the data point X given class C, P(C) is the prior probability of class C, and P(X) is the evidence, which is the probability of observing the data point X.

The likelihood of the data point X given class C is typically modeled using a probability density function, such as a Gaussian distribution. The prior probability of class C is the probability of observing class C in the training data. The evidence P(X) is a normalizing constant that ensures that the probabilities sum to 1 over all classes.
Once the posterior probabilities are computed for each class, the class with the highest probability is chosen as the predicted class for the data point.
Bayesian classification has several advantages over other classification methods, such as decision trees and neural networks. It is a probabilistic method, which means that it provides a measure of uncertainty in the classification results. It also handles missing data and noisy data well, and can be updated easily as new data becomes available. However, Bayesian classification can be computationally expensive, especially when dealing with highdimensional data, and it requires a significant amount of training data to learn an accurate model.

CODE AND OUTPUT import pandas as pd import numpy as np from sklearn.model_selection import train_test_split from sklearn.naive_bayes import GaussianNB from sklearn.metrics import confusion_matrix, roc_curve, roc_auc_score import matplotlib.pyplot as plt

df = pd.read_csv("BankLoan.csv") df.drop(df.columns[[0, 1, 2, 4, 5, 11]], axis=1, inplace=True) df = df.dropna()

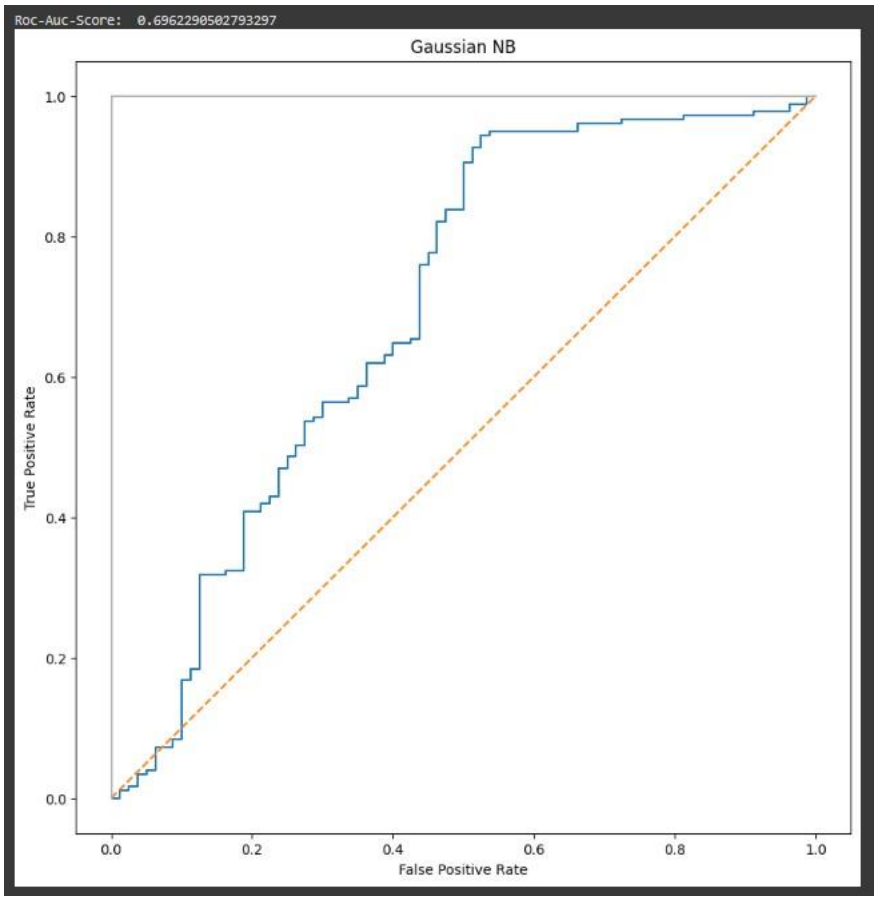
X = df.iloc[:, 1 : 6]
y = df.iloc[: , -1]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5)

gnb = GaussianNB() gnb.fit(X_train, y_train) y_pred = gnb.predict(X_test)
nb_loan = gnb.score(X_train, y_train) * 100

print("Bank Loan") print("Naive Baeyes Classifier") print(f"Accuracy - ", gnb.score(X_train, y_train)) print(f"Confusion Matrix : ")
print(confusion_matrix(y_test, y_pred))

```
Bank Loan
Naive Baeyes Classifier
Accuracy -  0.813953488372093
Confusion Matrix :
[[ 38  42]
 [ 13 166]]
```

y_score = gnb.predict_proba(X_test)[: , 1] false_pos, true_pos, threshold = roc_curve(y_test, y_score, pos_label = 'Y') print("\nRoc-Auc-Score: ", roc_auc_score(y_test, y_score)) plt.subplots(1, figsize = (10, 10)) plt.title("Gaussian NB") plt.plot(false_pos, true_pos) plt.plot([0, 1], ls = "--") plt.plot([0,0], [1,0], c="0.7"), plt.plot([1,1], c="0.7") plt.ylabel("True Positive Rate") plt.xlabel("False Positive Rate") plt.show()



CONCLUSION

Hence, we have successfully implemented Bayesian Classification.