Total number of examples : 768
Training examples : 537
Test examples : 231
Accuracy of the model : 100.0

Expt, No.

Page No. 25
Date 09/05/2024

Write a python code to implement Naive Baye's classifier.

```
import math
import random
import pandas as pd
import numpy as np

def encode_class(mydata):
    classes = []
    for i in range(len(mydata)):
        if mydata[i][-1] not in classes:
            classes.append(mydata[i][-1])
    for i in range(len(classes)):
        for j in range(len(mydata)):
            if mydata[j][-1] == classes[i]:
                mydata[j][-1] = i
    return mydata

def splitting(mydata, ratio):
    train_num = int(len(mydata) * ratio)
    train = []
    test = list(mydata)
    while len(train) < train_num:
        index = random.randrange(len(test))
        train.append(test.pop(index))
    return train, test
```

Avijit®

Teacher's Signature :.....

```
def groupUnderClass(mydata):
    data_dict = {}
    for i in range(len(mydata)):
        if mydata[i][-1] not in data_dict:
            data_dict[mydata[i][-1]] = []
        data_dict[mydata[i][-1]].append(mydata[i])
    return data_dict

def MeanAndStdDev(numbers):
    avg = np.mean(numbers)
    stddev = np.std(numbers)
    return avg, stddev

def MeanAndStdDevForClass(mydata):
    info = {}
    data_dict = groupUnderClass(mydata)
    for classValue, instances in data_dict.items():
        info[classValue] = [MeanAndStdDev(attribute) for attribute in
                            zip(*instances)]
    return info

def calculateGaussianProbability(x, mean, stdev):
    epsilon = 1e-10
    expo = math.exp(-(math.pow(x-mean,2)/(2*math.pow(stdev+
                                          epsilon,2))))
    return (1/(math.sqrt(2*math.pi)*(stdev+epsilon)))*expo
```

Teacher's Signature :.....

```
def calculateClassProbabilities(info, test):
    probabilities = {}
    for classValue, classSummaries in info.items():
        probilities[classValue] = 1
        for i in range(len(classSummaries)):
            mean, std_dev = classSummaries[i]
            x = test[i]
            probabilities[classValue] *= calculateGaussianProbability
                                          (x, mean, std_dev)
    return probabilities

def predict(info, test):
    probabilities = calculateClassProbabilities(info, test)
    bestLabel = max(probabilities, key = probabilities.get)
    return bestLabel

def getPredictions(info, test):
    predictions = [predict(info, instance) for instance in test]
    return predictions

def accuracy_rate(test, predictions):
    correct = sum(1 for i in range(len(test)) if test[i][-1] ==
                                                predictions[i])
    return (correct / float(len(test))) * 100.0
```

Teacher's Signature :.....

Expt, No.

Page No. 28
Date

```
filename = 'pima-indians-diabetes.csv'
df = pd.read_csv(filename)
mydata = df.values.tolist()

mydata = encode_class(mydata)
for i in range(len(mydata)):
    for j in range(len(mydata[i])-1):
        mydata[i][j] = float(mydata[i][j])

ratio = 0.7
train_data, test_data = splitting(mydata, ratio)
print('Total number of examples:', len(mydata))
print('Training examples:', len(train_data))
print('Test examples:', len(test_data))

info = MeanAndStdDevForClass(train_data)

predictions = getPredictions(info, test_data)
accuracy = accuracy_rate(test_data, predictions)
print('Accuracy of the model:', accuracy)
```

Avijit®

Teacher's Signature :.....