Page No.:
Date 30/04/2021

Name - Rubana
Roll no - 17EAYC6079
Lab - ML Lab (GCS4-22)
Branch - CSE

Q1: Write a program to implement the naive Bayesian classifier for a sample training data set stored as a CSV file. Compute the accuracy of the classifier, considering few test data sets.

Ans:-

```
import csv
import random
import math

def loadCSV(filename):
    lines = csv.reader(open(filename, "r"));
    dataset = list(lines)

    for i in range(len(dataset)):
        dataset[i] = [float(x) for x in dataset[i]]

    return dataset

def splitdatdata(dataset, splitratio):

    trainsize = int(len(dataset) * splitration);
```

Teacher's Sign

```
trainset = []
Copy = list (dataset);
while (len(trainset) < trainsize):
    index = random.randrom(len(copy));
    trainset.append (Copy.pop(index));
    return [trainset, copy]

def SeparateByClass (dataset):
    Separated = {}
for i in range (len (dataset)):
    Vector = data [i]
        if (vector [-1] not in Separated):
            Seperated [Vector [-1]] = []
    Separated [Vector [-1]] : append (Vector)

return Seperated

def mean (numbers):
    avg = main (numbers)
    Variance = sum ([pow (x-avg),2)
for x in numbers])

    return: math.sqrt (Variance)

def: Summarise (dataset):
    Summaries = [(mean (attribute), stdev
    (attribute) for attribute in zip (*dataset)];
    del. Summaries. [-1]
    return Summaries
```

Summaries = {}
for classvalue, instance in separated.items():
Summaries[classvalue] = summarize(instance)
(9) to mean and std
return Summaries.

def calculateclassprobabilities(summaries, inputVector):
probabilities = {}

test data
for classvalue, classsummaries in Summaries.items():
information as means and sd.
probabilities[classvalue] = 1
for i in range(len(classsummaries)):

stdev = classmates[i]
attribute for class 0 and 1 seperately
x = inputvector[i]
probabilities[classvalue] = calculateprob
mean, stdv);

return probabilities

def predict(Summaries, inputVector):

Teacher's Sign

bestLabel, bestProb = None, -1

which has the height prob

if bestLabel is None or probabilities > bestProb:

bestProb = probability

bestLabel = classValue

return bestLabel

def getPredictions (summaries, testSet):

predictions = []

for i in range (len (testSet)):

result = predict (summaries, testSet(i))

predictions.append (result)

return predictions

def getAccuracy (testSet = splitDataset (dataset

print ('Split {0} rows into: train = {1} and test

summaries = summarizeByClass (trainSet);

predictions = getPrediction (summaries, testSet)

test data with the trainig data.

accuracy = getAccuracy (testSet, prediction)

print('Accuracy of the classifier is {0} %', format(accuracy))

main()

Output :-

Confusion matrix is as follows : [[17 0 0]
[0 17 0
[0 0 11]]

Accuracy matrix

	precision	recall	f1-score	Support
0	1.00	1.00	1.00	17
1	1.00	1.00	1.00	17
2	1.00	1.00	1.00	17
avg/total	1.00	1.00	1.00	45

(21) 1-nearest neighbor is a Consistent estimator.

Ans :- (b) False

Ans (c) ~~More than One Independent Variable~~

(22) A multiple regression model has

Ans (b) ~~polynomial in D~~.

Ans : more than One Independent Variable

(23) Computational Complexity of gradient decent is -

Teacher's Sign

Ans (c) polynomial in D.

Q(24) Selecting data so as to assure that each class is properly represented in both the training and test set is.

Ans:- (a) Cross Validation

Q(25) A dataset with two class values that is significantly skewed (more than 90%) towards one of those class value is known as ___ dataset.

Ans (c) An Imbalanced

Q(26) The method of imputation that fills in missing values using similar instances from the same dataset is known as ___ imputation.

Ans (c) Hot-Deck

Q(27) In Association rules, a collection of one or more items is known as ___

Ans (a) An Itemset.

Q(28) K-means clustering is useful in creating non-spherical clusters.

Ans (b) False.

Q(29) Which of the following is true about Naive Bayes?

Ans:- Both A and B

Q(30) T/D items Bought

T-1 bread, milk, beer
I2 bread, diper, beer, eggs.
T3. milk, diper, beer, Coke
Ty. bread, milk, diaper, beer
T5. bread, milk, diaper, coke

What is the support of the itemset (beer, coke) in the dataset by ?

Ans (b) Assumes that all the features in a dataset are independent.

Teacher's Sign