**Lab - 4**

**PRML**

AY 2020-21 Trimester - III

Boosting and Bayes Classification

**Name: Akshaykumar Kanani(B19EE008)**

**Q 1 :**
**ANS:**

We did the preprocessing of data after reading the csv file. And then we study the distribution of different class of data and their amount on data where we get
All the class has same amount and their are three class
Iris-setosa = 33.33%,50 sample
Iris-virginica = 33.33%,50 sample
Iris-setosa = 33.33%,50 sample

And we also analyze every feature with respect to the target class and plot their graphs.

And after that, we perform boosting on it with a decision tree as a base classifier.

And after that we Perform cross-validation over the data and got an accuracy as =0.9333333333333332

And from adaboost we got accuracy and std as below for a different number of decision trees

```
>1 0.938 (0.038)

>2 0.956 (0.040)

>3 0.949 (0.036)

>4 0.940 (0.033)

>5 0.953 (0.034)
```

**Q 2 :**

**ANS :**

We did all the steps required to fit ou model as preprocessing and then we got results as follows **:**

Gaussian Naive Bayes accuracy = 0.92381 , standard deviation = 0.03810 , Variance = 0.00145
Multinomial Naive Bayes accuracy = 0.73333 with standard deviation = 0.03810, Variance = 0.00145
Complement Naive Bayes accuracy = 0.63810 with standard deviation = 0.02333 , Variance = 0.00054
Bernoulli Naive Bayes accuracy = 0.36190 with standard deviation = 0.02333 , Variance = 0.00054

Using logarithmic discriminant function to compute accuracy we get accuracy = 1.0
And we also got Bayes risk for

Given: $\lambda =$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Bayes risk for Gaussian = : [ 93.3902827 228.3902827 363.3902827]

Bayes risk for Multinomial = : [ 93.49073385 228.49073385 363.49073385]

Bayes risk for Categorical = : [ 90.27496081 225.27496081 360.27496081]

Bayes risk for Bernoulli = : [ 88.58039417 223.58039417 358.58039417]

Bayes risk for Linear Discriminant = : [ 92.99622855 227.99622855 362.99622855]

**Q 3 :**

**ANS:**

**For data.csv :**

a,b) We create the labels and plot the histogram from given data.

c) After that we get prior probability as :

Prior Class 1: 0.8518518518518519

Prior Class 2: 0.14814814814814814

d) Posterior Probability we got :

1.0: 0.22222222222222224,

1.5: 0.1851851851851852,

2.0: 0.14814814814814814,

2.5: 0.1111111111111111,

3.0: 0.25925925925925924,

4.0: 0.037037037037037035,

5.0: 0.037037037037037035

For C1 And C2 data set :

a,b) We create the labels and plot the histogram from given data.

c) After that we get prior probability as :

Prior Class 1: 0.8040201005025126

Prior Class 2: 0.19597989949748743

d) Posterior Probability we got in code file.

**REAL data:**

Done on code file.

Remaining details can be found on output of code file.