Lab-3
PRML 2023
Akshat Jain (B21CS005)

Question 1

A. We first drop the columns that are noisy or not useful.
   Since the cabin column has 687 out of 891, 77.104% of the entries missing, we can safely drop this column from our dataset without any major repercussions. Similarly, we can also drop names since all of them are different, and thus the complete column is noisy and would adversely affect our dataset. We can also drop the passenger Id Column as it is just a column which gives us the serial no., which we already have from loading our dataset. Similarly, we can drop the ticket column as we already have a separate passenger class Column.
   We then encode the remaining columns and calculate the covariance matrix.
   We then Split the Dataset into training and testing.
B. We Have Selected the Gaussian naive Bayes as it is the only one of the three naive bayes classifiers which work with continuous data.
C. As We have Selected the Gaussian Naive Bayes, We then implement it with the help of the sklearn library. We also plot different Metrics Such as AUC, ROC, MSE error etc.
D. We have performed the K-fold cross-validation with metrics accuracy, precision and recall.
   Accuracy:- [[0.74719101 0.79775281 0.80898876 0.75280899 0.8079096]
   Precision:- [0.60344828 0.69620253 0.72463768 0.66666667 0.74193548]
   Recall:- [0.61403509 0.82089552 0.76923077 0.70588235 0.71875]]
   We then print the probability of the top class.
E. We plot the different contour plots with the help of the seaborn library and see their correlation with each other. This implies that our naive Bayes assumption is somewhat flawed as the features are not perfectly independent of each other
F. We see that the performance metrics of DTC are better than the naive Bayes classifier. As the features here are correlated to each other thus the DTC model here outperforms naive Bayes as it takes in consideration the correlation between the different features.

Question 2

A. We plot the Histograms with the help of the Seaborn library
B. We find that the prior probability is 0.33 or ⅓ for each class of Y
C. We Discretize the data into 42 bins each of size 5. We use the mean function to discretise the data into bins.
D.
E.
F. We have Fount the posterior probability for every class concerning every feature.