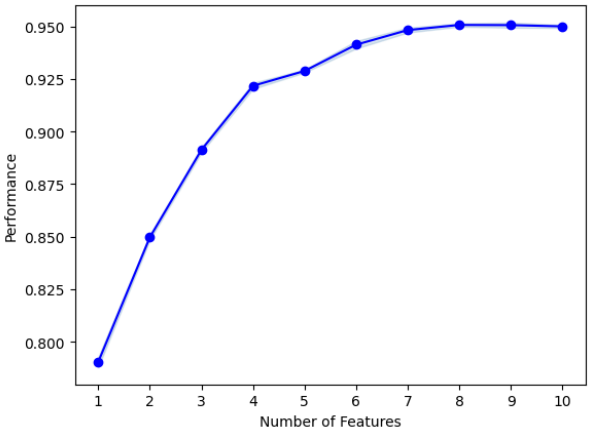
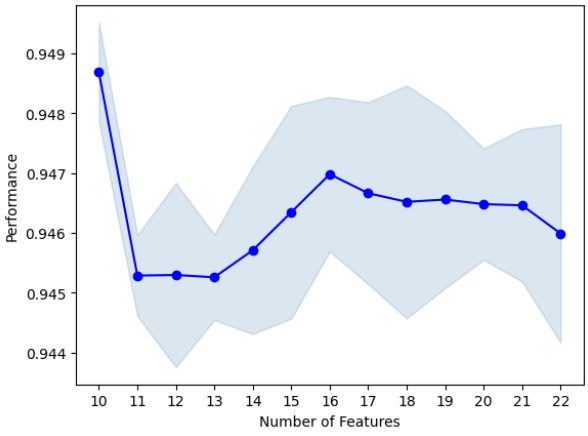
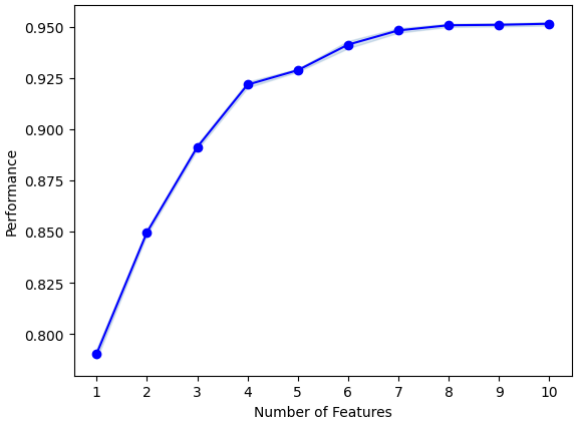
LAB ASSIGNMENT 8
Lab Report
Shubh Goyal (B21CS073)

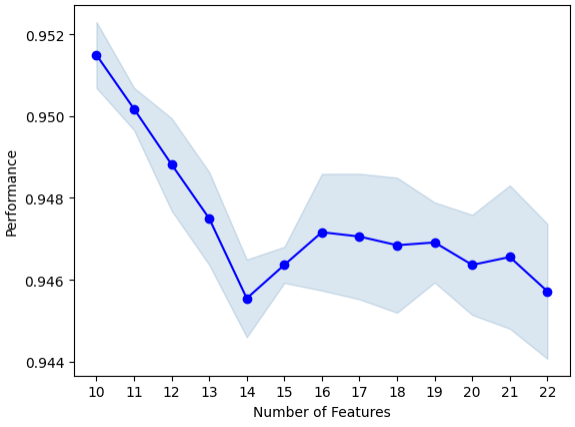
QUESTION 1-

Label encoding was applied on various features and the missing values in 'Embarked' was replaced by the features mean. Then, SFS, SBS, SFFS, SBFS were applied on the data and the following results were obtained for the four respectively.







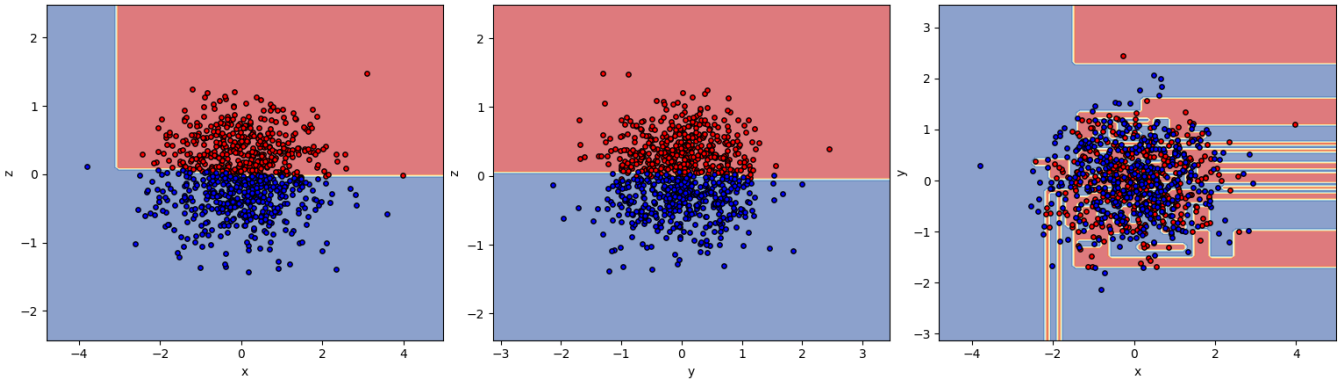


A class was implemented to perform bidirectional feature selection from scratch. From the features selected using city block distance, angular distance and euclidean distance, a decision tree classifier was giving around 90% accuracy.

QUESTION 2-

Created the dataset. The dataset is easily separable along the z-axis.

Task 3: A decision tree was trained taking x and y axis values only. It overfitted and gave an accuracy of 100% on the train set while an accuracy of 53.5% was achieved on the testing data. Through the decision boundaries too, the decision tree appears to be overfitted.



Task 4: The data obtained after applying PCA resembles the x and y axis plot of the data. On this data, a train accuracy of 100% and a test accuracy of 50% was obtained.