There are 2 python file in project

**For part 1** In assignment main.py is used

In program we use the dataset of titanic having training and testing dataset. We use model Logistic regression to see the confusion matrix for which we use the sklearn

We first take the data set and clean them. After that we drop some of the variables in test and training data sets using drop function.

Then we load the datasets to logistic regression model to get the accuracy, auc and log loss. After using 10 folds for cross validation in logistic regression we get the score and print that. At the end we found the best parameter and best accuracy.

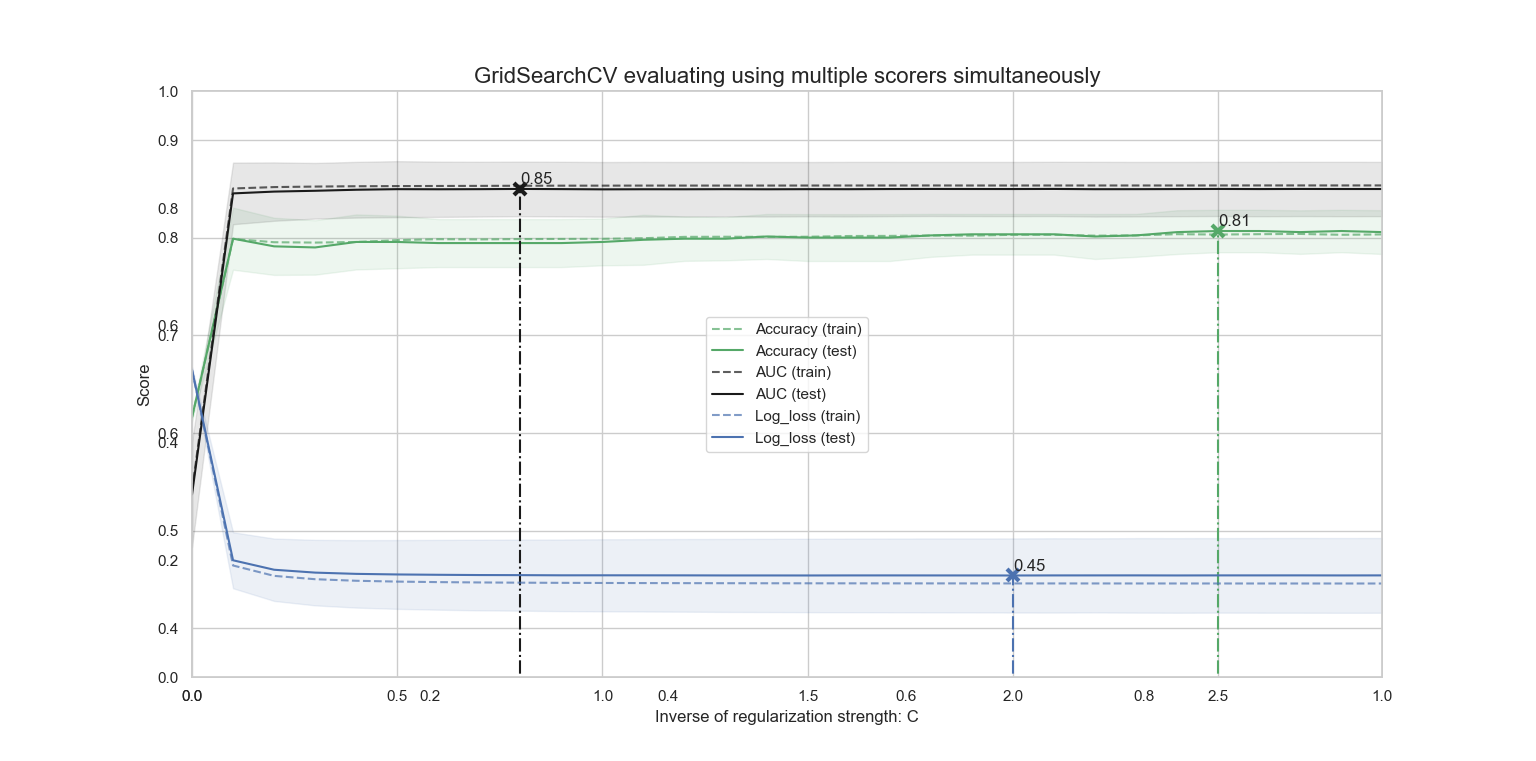
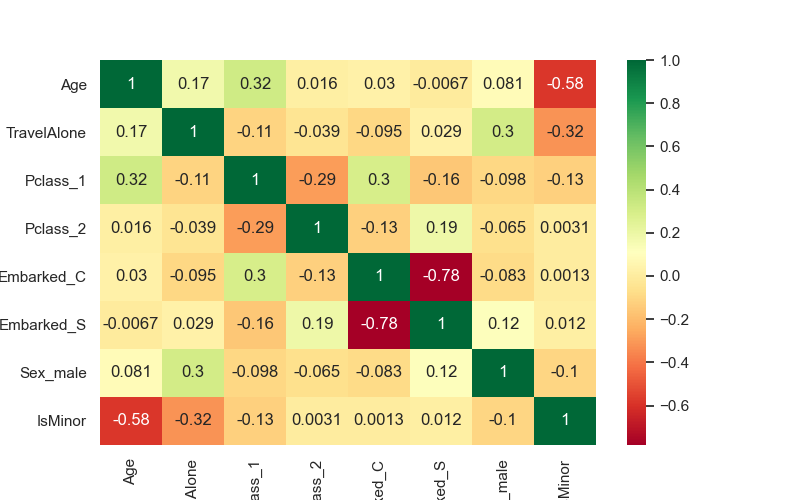
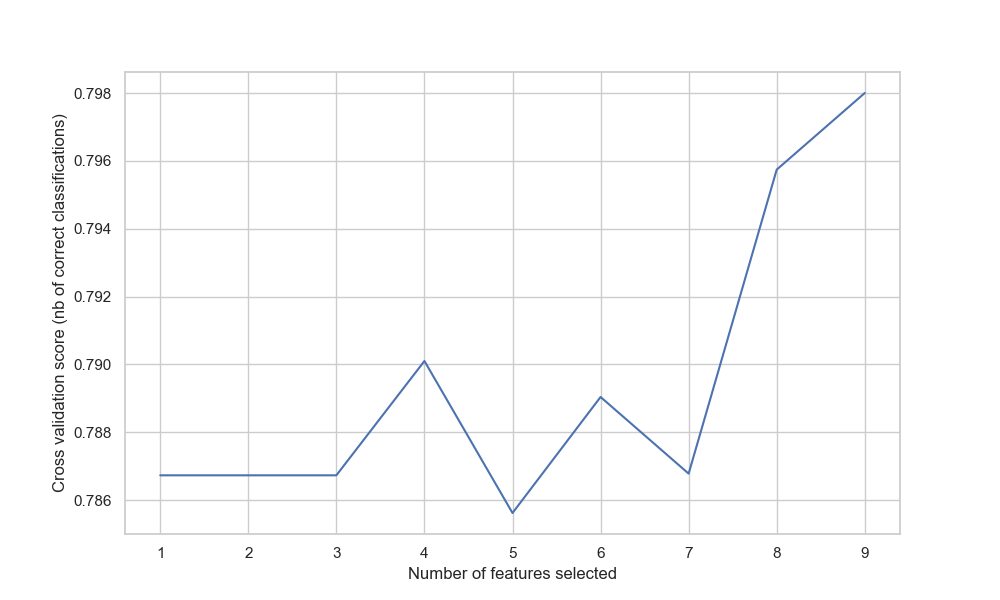
Final results are

best params: LogisticRegression(C=2.50001)

best params: {'C': 2.50001}

best score: 0.8069662921348316

Output of program



**For Point 2** we use the Neural network MLPClassifier. The dataset used for NN is makemoon and and makecircules. The program was copy pasted and try to understand. I was unable to use my datasets which I use in part 1 due to limited knowledge. But the basic theme was it take different values for regularization parameter 'alpha' on synthetic datasets. The plot shows that different alphas yield different decision functions. Alpha is a parameter for regularization term and Increasing alpha may fix high variance (a sign of overfitting) by encouraging smaller weights, resulting in a decision boundary plot that appears with lesser curvatures. Similarly, decreasing alpha may fix high bias (a sign of underfitting) by encouraging larger weights, potentially resulting in a more complicated decision boundary.