Lab 4 – Assignment

(Logistic Regression)

1. Consider the two datasets given MPG - training dataset and test dataset. Preprocess the dataset appropriately for Logistic regression to be applied.

   Attribute “Origin” should be taken as the Class Label.

a. There are 3 distinct values in “Origin” attribute. Convert those values into 2 categories as USA and Non-USA
b. Apply Logistic regression on the Preprocessed Dataset with regularization parameter C = 0.1.
c. Vary C = 0.001, 0.01, 0.1, 1, 10,100 and plot a graph for the cost function J(Θ) of Logistic regression for the training dataset
d. Predict the target feature for the test dataset
e. Plot the confusion matrix

2. Consider the dataset MPG with 3 categories in Origin target feature.
   a. Preprocess the dataset appropriately
   b. Apply Logistic regression on the preprocessed dataset with regularization parameter C = 0.1

```
# define model

model = LogisticRegression(multi_class='ovr')
# fit model
model.fit(X, y)
# make predictions
yhat = model.predict(X)
```

   c. Classify the test instances into 1(usa),2(Europe) or 3(Japan)
   d. Find how many of the test instances were predicted correctly.
   e. Visualize the classified dataset with each class coloured differently