Machine Learning
Assignment - 3

Name: Mansi Goel
Roll Number: 2014062

Procedure

Preprocessing:

1) I read all the training and testing data using a standard template from https://gist.github.com/akesling/5358964
2) Then I normalised the data by dividing by 255.0 and further converted it into numpy array.
3) Then I created a smaller dataset by choosing 2000 samples from each class and for classes 3 and 8 as required in the first part. Similar was done for testing dataset.
4) Then I created a dictionary for grid parameters to store the kernel and C values to be passed in the GridSearchCV argument.
5) Then I called GridSEarchCV as per each required part with 5 cross validation. Then I trained the training dataset by clf.fit
6) Then through attributes best_params_ and grid_results_, we get the best parameters and mean scores for each parameter respectively.
7) By using these values, I use SVC model to create a new classifier which I train using my training dataset. I further dump it using sklearn.joblib [for multiclass I dump 10 models].
8) I then test it with my testing dataset and using decision_function, I get the decision matrix which I reshape into a numpy array. I binarize my Testing labels using inbuilt function and pass both matrices as arguments to the ROC function which plots the ROC curve.