

Assignment 3

SVM

1. Methodology: Details of the SVM package used

A) **Pandas** :- The pandas package is used for importing the CSV data from datasets.

B) **train_test_split** :- From sklearn.model_selection package, we use train_test_split to split the datasets in training data and testing data. In our case, we use 70% data for training and rest of 30% data as test set.

C) **confusion_matrix** :- From sklearn.matrix package, we use confusion_matrix to make confusion matrix of y_test and predicted output.

D) **Scale** :- from sklearn.preprocessing package, scale is used for standardising a dataset along any axis. In our case

$$X = x - \text{mean}(x)/\text{std}(x)$$

E) **GridSearchCV** :- The performance of any model is dependent on the value of hyper-parameters. The GridSearchCV is the process of performing hyper-parameter tuning in order to determine the optimal values for a given model.

2.

a) Linear

C value	Gamma = 1	Gamma = 0.1	Gamma = 0.01	Gamma = 0.001	Gamma = 0.0001
0.1	0.944	0.944	0.944	0.944	0.944
1	0.943	0.943	0.943	0.943	0.943
10	0.944	0.944	0.944	0.944	0.944
100	0.946	0.946	0.946	0.946	0.946
1000	0.944	0.944	0.944	0.944	0.944

Grid Best Estimator :- SVC(C=100, gamma=0.1, kernel='linear')

Grid Best Score :- 0.9298136645962731

Grid score :- 0.9217958001448225

B) Quadratic

C value	Gamma = 1	Gamma = 0.1	Gamma = 0.01	Gamma = 0.001	Gamma = 0.0001
0.1	0.921	0.887	0.630	0.602	0.602
1	0.904	0.932	0.705	0.604	0.602
10	0.902	0.939	0.797	0.607	0.602
100	0.894	0.921	0.887	0.630	0.602
1000	0.894	0.904	0.932	0.707	0.604

Grid Best Estimator :- SVC(C=10, gamma=0.1, kernel='poly')

Grid Best Score :- 0.9254658385093167

Grid score :- 0.9275887038377987

C) RBF

C value	Gamma = 1	Gamma = 0.1	Gamma = 0.01	Gamma = 0.001	Gamma = 0.0001
0.1	0.613	0.834	0.922	0.797	0.602
1	0.797	0.933	0.957	0.924	0.797
10	0.8	0.93	0.949	0.957	0.919
100	0.798	0.925	0.943	0.953	0.955
1000	0.8	0.911	0.933	0.950	0.952

Grid Best Estimator :- SVC(C=100, gamma=0.001 kernel='rbf')

Grid Best Score :- 0.9391304347826088

Grid score :- 0.9232440260680667

The Linear kernel method performs good in terms of Grid Best Estimator but takes too much time to compute the GridScores.

The RBF kernel method performs faster compared to other Linear and Quadratic method.