BUAN 6341.501 - Applied Machine Learning
Assignment 2 - Pavan Gorantla (PCG180000)

Objective

Experimentation of classification algorithms (Support Vector Machine, Decision Trees and Boosting) on two datasets. Both datasets have been downloaded from archives of UCI Machine Learning Repository.

Dataset 1 - Appliances Energy Prediction (Link: appliances+energy+prediction)

Dataset consists of 19735 observations on a total of 29 variables. More information about these variables can be found in the above given weblink. The response variable is created using Appliances variable where appliances greater than median are classified as Heavy Appliances and the goal is to predict this classifier.

Dataset 2 - Online Shoppers Purchasing Intention (Link: online+shoppers+purchasing+intention)

Dataset consists of 12330 observations on a total of 18 variables. More information about these variables can be found in the above given weblink. The response variable Revenue specifies if an online user on the shopping website has purchase intent or not based on the pages visited and duration spent on each page.

Test/Train and Normalization

Both datasets are split into train and test sets using 75-25 split percentage and features are normalized to a range of [-1,+1]. Response from Dataset1 is classified as +1 for Heavy Appliances and -1 for the rest. Response from Dataset2 is classified as +1 for users with purchase intent and -1 for the remaining users. All features in Dataset1 are numeric and hence they are just normalized for our classification algorithms. There are few categorical features in Dataset2 which are transformed into dummies before normalizing.

Methodology and Algorithms

By using 3 classification algorithms (SVM, Decision Tree, Boosting) on both datasets, response variable is predicted by experimenting on different parametric values. Cross validation is done on AdaBoost model.

Hyperparameter Tuning

For SVM, kernels like Linear, Radial Bias Function, Sigmoid are used and are tuned for hyperparameters.

Dataset	Kernel	C	Gamma
Appliances Energy	Linear	0.01,0.05,0.1,0.2,0.5,1,2,5,10,20,25	-
Appliances Energy	RBF	0.001,0.01,0.1,1,10	0.001,0.01,0.1,1,10
Appliances Energy	Sigmoid	0.001,0.01,0.1,1,10	0.001,0.01,0.1,1,10
Online Shoppers	Linear	0.0001,0.0005,0.001,0.005,0.01,0.05	-
Online Shoppers	RBF	0.001,0.01,0.1,1,10	0.001,0.01,0.1,1,10
Online Shoppers	Sigmoid	0.001,0.01,0.1,1,10	0.001,0.01,0.1,1,10

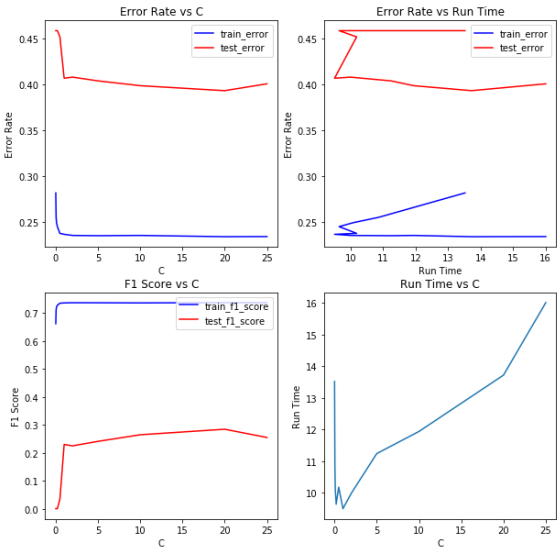
Decision trees are experimented by tuning maximum tree depth and maximum leaf nodes parameters. AdaBoost algorithm has been used for Boosting and tuned for Number of Estimators and Learning Rate.

Dataset	Algorithm	Depth / Number of Estimators	Nodes / Learning Rate
Appliances Energy	Decision Trees	5,10,15,20,25	5,10,15,20,25
Appliances Energy	AdaBoost on Trees	10,50,100,200,500,1000	0.005,0.01,0.05,0.1,0.2,0.5
Online Shoppers	Decision Trees	10,20,50,100,200	50,75,100,150,200
Online Shoppers	AdaBoost on Trees	5,10,20,50,100,200	0.02,0.05,0.1,0.2,0.5,1

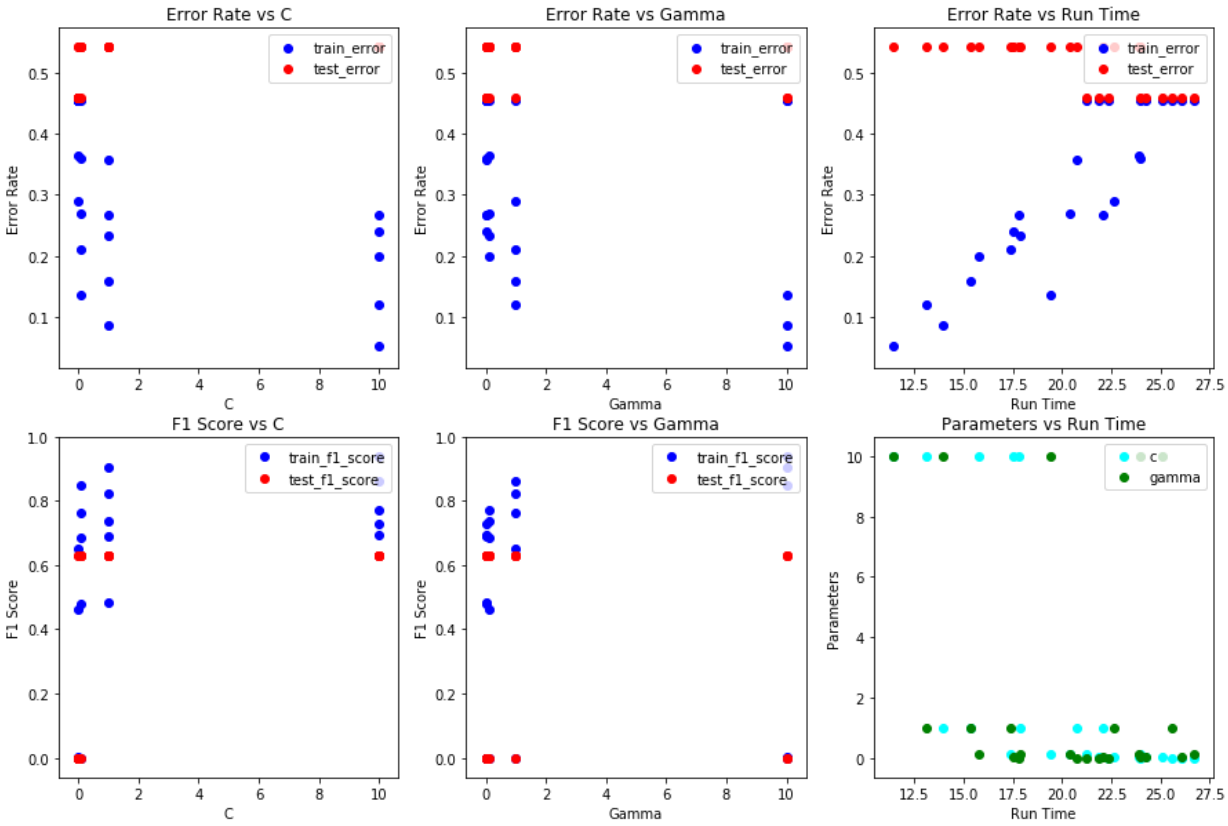
Support Vector Machines (Dataset 1 - Heavy Appliances Prediction)

a. Linear Kernel: When C=20, Linear kernel is giving the least error and best F1 Score for test set. As C increases, Run Time increases and there is non-linear relationship between Error Rate and Run Time.

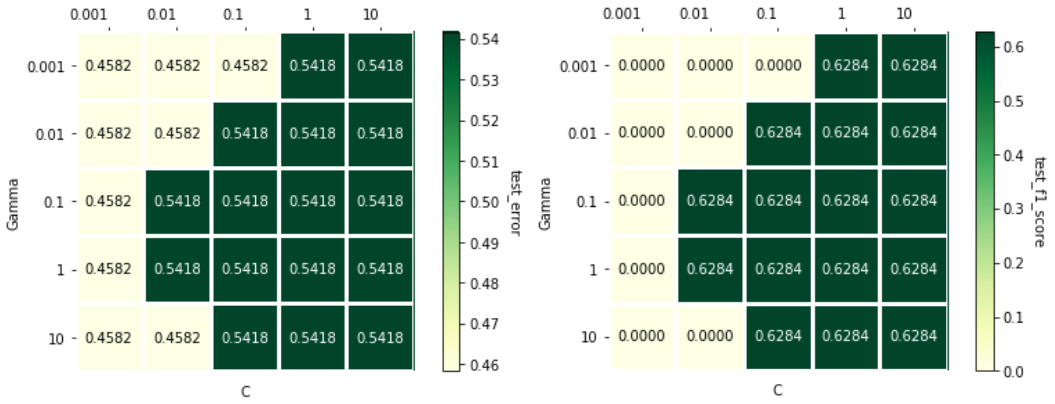
c	train_error	test_error	train_f1_score	test_f1_score
20.00	0.234327	0.393008	0.737235	0.284220
10.00	0.235703	0.398412	0.736185	0.264422
25.00	0.234472	0.400439	0.737030	0.254637
5.00	0.235413	0.403648	0.736680	0.240788
1.00	0.236934	0.406519	0.736197	0.229760
2.00	0.235703	0.407701	0.736868	0.224791
0.50	0.237947	0.451444	0.735154	0.033972
0.01	0.281888	0.458200	0.661920	0.000000
0.05	0.255393	0.458200	0.711953	0.000000
0.10	0.249602	0.458200	0.721621	0.000000
0.20	0.245258	0.458200	0.727653	0.000000



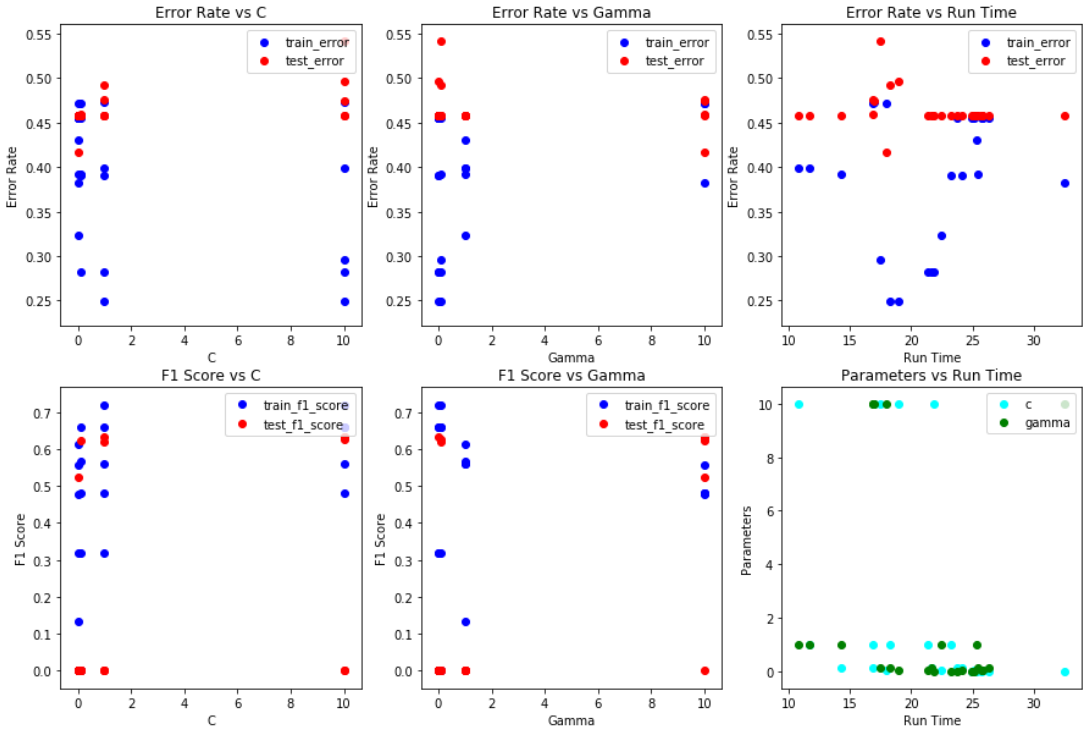
b. RBF Kernel: Following are various plots between the parameters, Error Rate, F1 Score and Run Time.



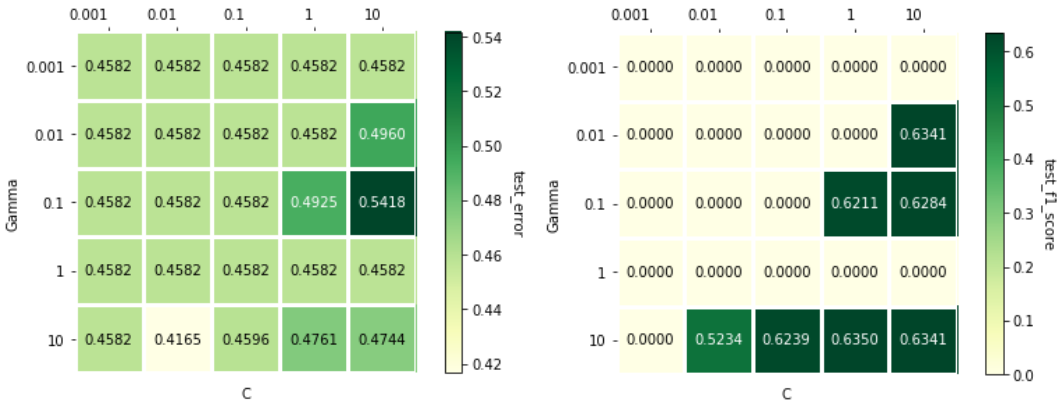
For all combinations of C and Gamma, the Error Rate and F1 Score are shown below. The combinations with higher F1 Score have slightly higher Error Rate. We can look at this as trade-off for better F1 Score.



c. Sigmoid Kernel: Following are some plots between parameters, Error Rate, F1 Score and Run Time.

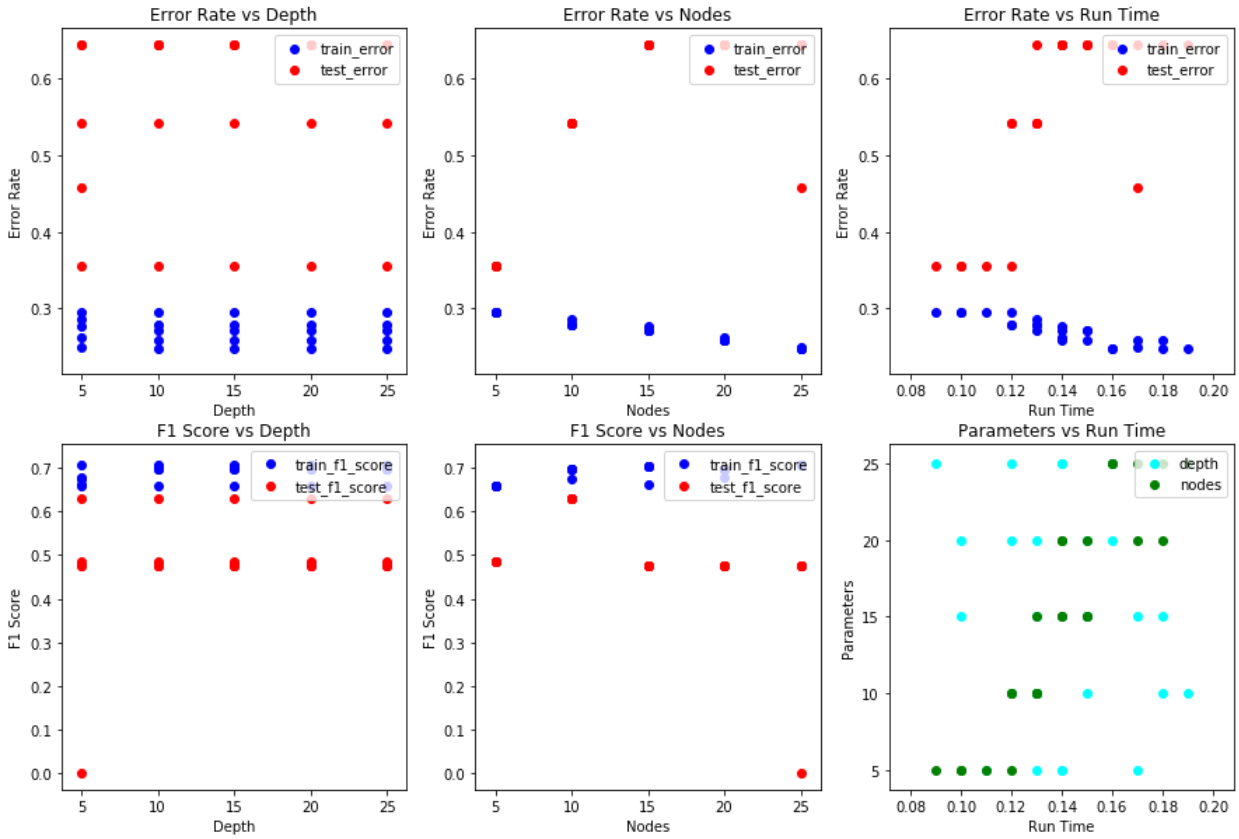


Again, a trade-off for least Error Rate, C = 0.1 and Gamma = 10, for best F1 Score, C = 1 and Gamma = 10.

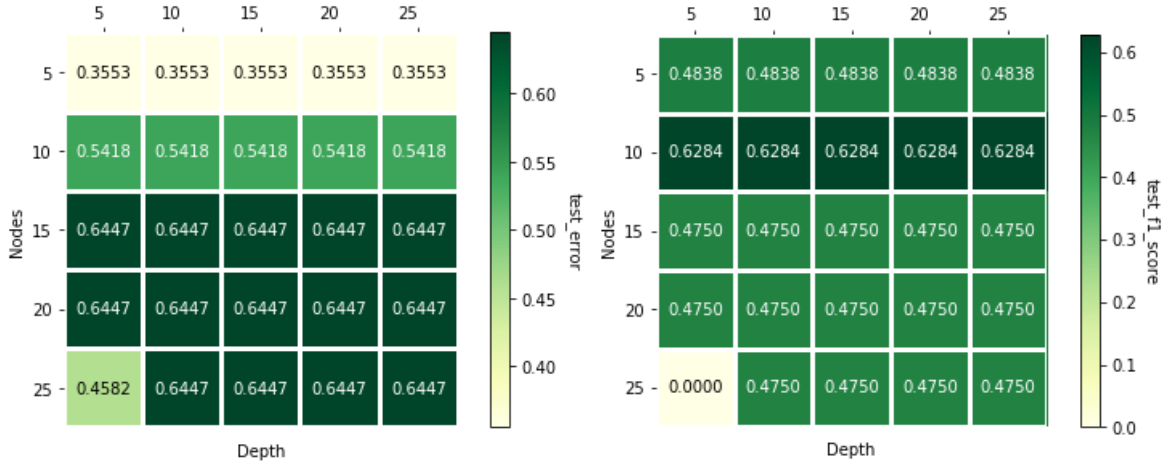


Decision Trees (Dataset 1 - Heavy Appliances Prediction)

Following are few plots between parameters, Error Rate, F1 Score and Run Time. As nodes number gets increasing, the train error is decreasing, the test error is increasing and Run Time is also increasing. From below, Error Rate and F1 Score seems independent of Depth and are non-linearly related to Leaf Nodes.

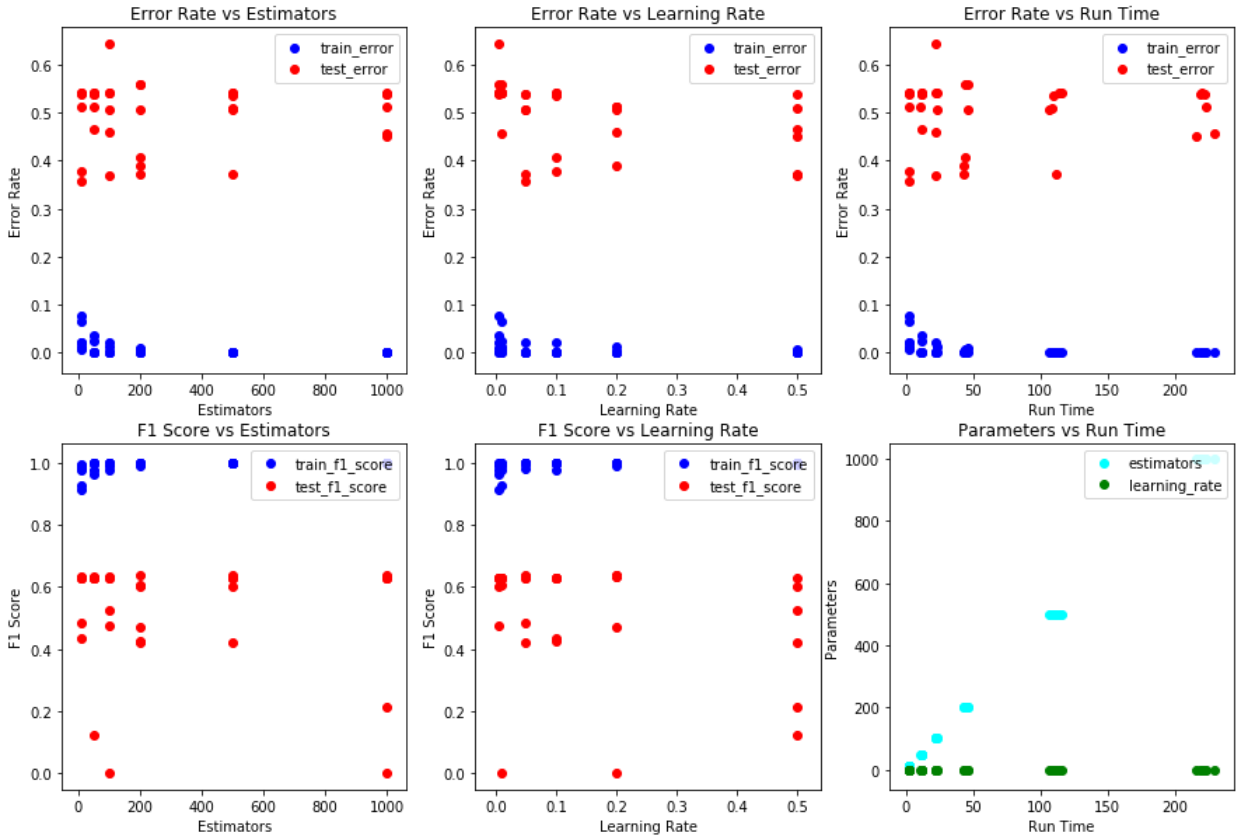


The least Error Rates are achieved for 5 and 10 Nodes, but highest F1 Score is achieved for Nodes = 10 irrespective of the Depth. For less complexity we can consider Depth = 5 as best parameter in this case.

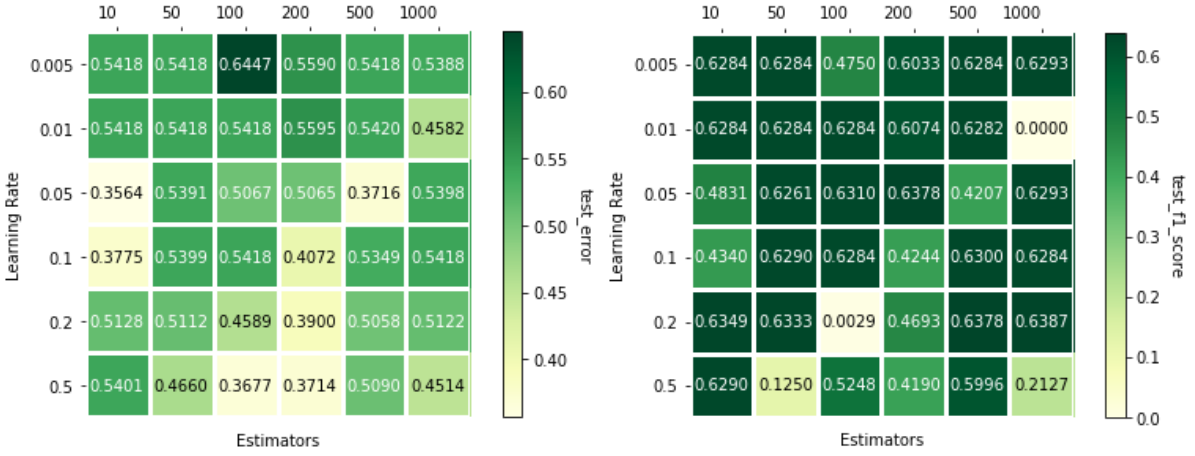


Adaptive Boosting on Decision Trees (Dataset 1 - Heavy Appliances Prediction)

Following are plots between parameters, Error Rate, F1 Score and Run Time. Estimators and Run Time have positive linear relation. Error Rate has low variance for Learning Rate = 0.2 and Estimators = 500.



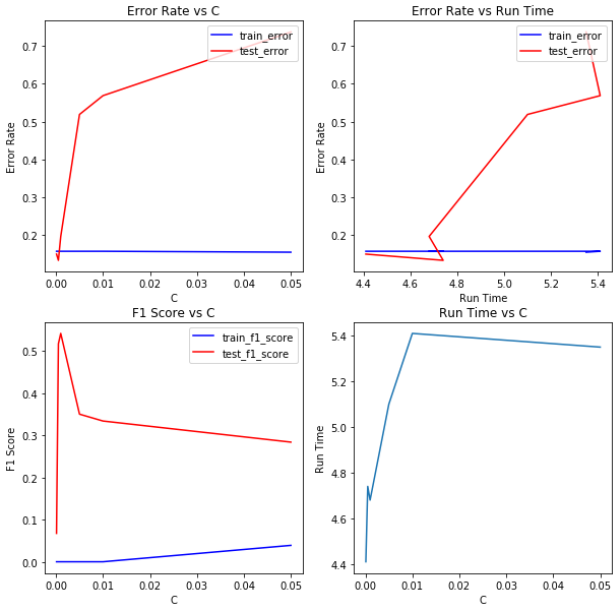
The least Error Rates are achieved when Estimators = 10 and Learning Rate = 0.05, but the highest F1 Score is achieved when Estimators = 500 or when Estimators = 1000 and when the Learning Rate = 0.2.



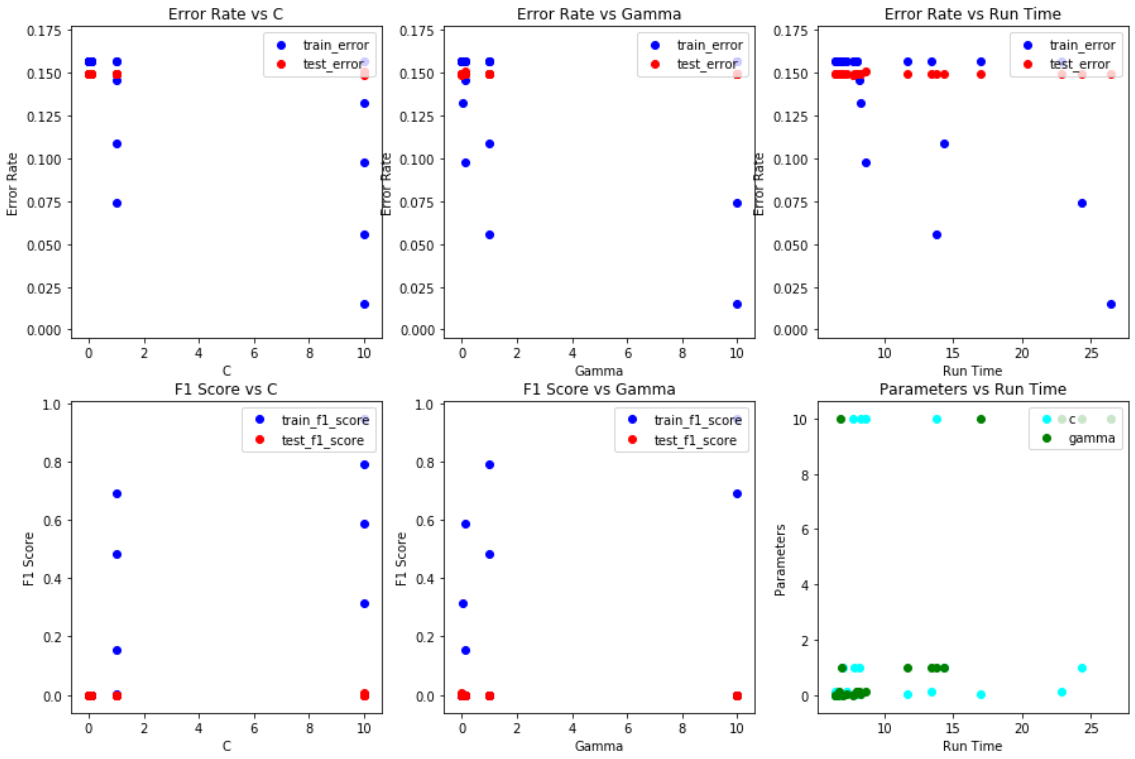
Support Vector Machines (Dataset 2 - Online Shopping Intention Prediction)

a. Linear Kernel: When C = 0.0005, Linear kernel is giving the least error and when C = 0.001, best F1 Score is achieved. As C increases, Run Time increases with slightly non-linear relation between them.

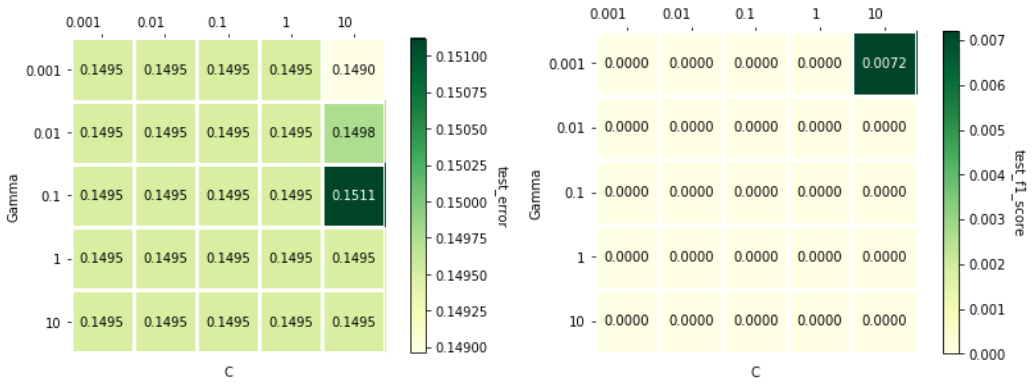
c	train_error	test_error	train_f1_score	test_f1_score
0.0010	0.156992	0.196269	0.000000	0.542245
0.0005	0.156992	0.133009	0.000000	0.517647
0.0050	0.156992	0.519059	0.000000	0.350474
0.0100	0.156992	0.568802	0.000000	0.334177
0.0500	0.154559	0.737226	0.038905	0.284064
0.0001	0.156992	0.149500	0.000000	0.067454



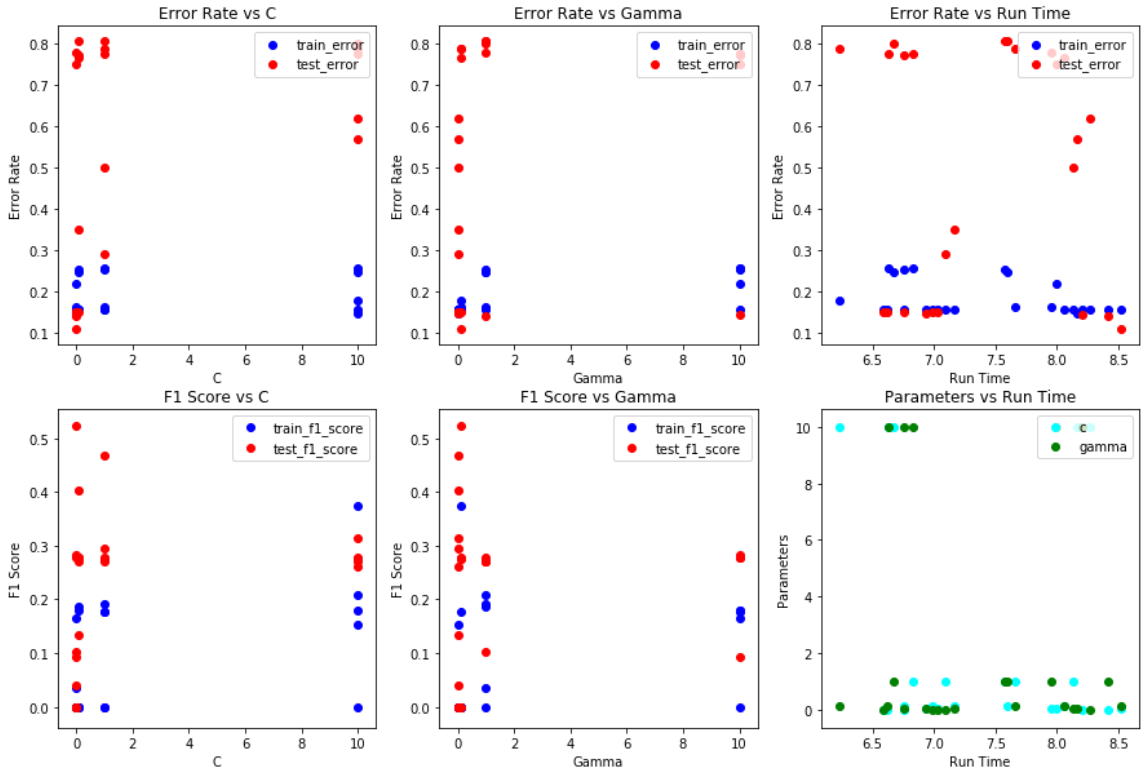
b. RBF Kernel: Following are various plots between the parameters, Error Rate, F1 Score and Run Time.



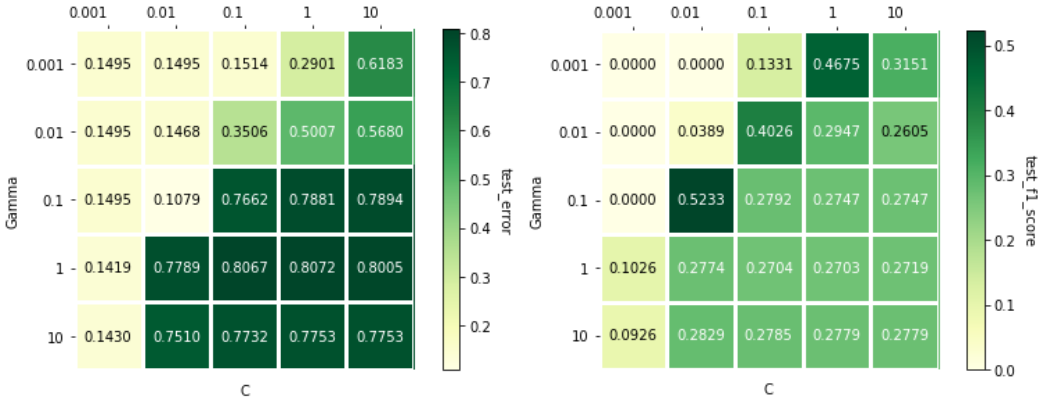
For all combinations of C and Gamma, the Error Rate and F1 Score are shown below. The combinations with higher F1 Score and lease Error Rate is C = 10 and Gamma = 0.001. But all the results are very poor.



c. Sigmoid Kernel: Following are some plots between parameters, Error Rate, F1 Score and Run Time.

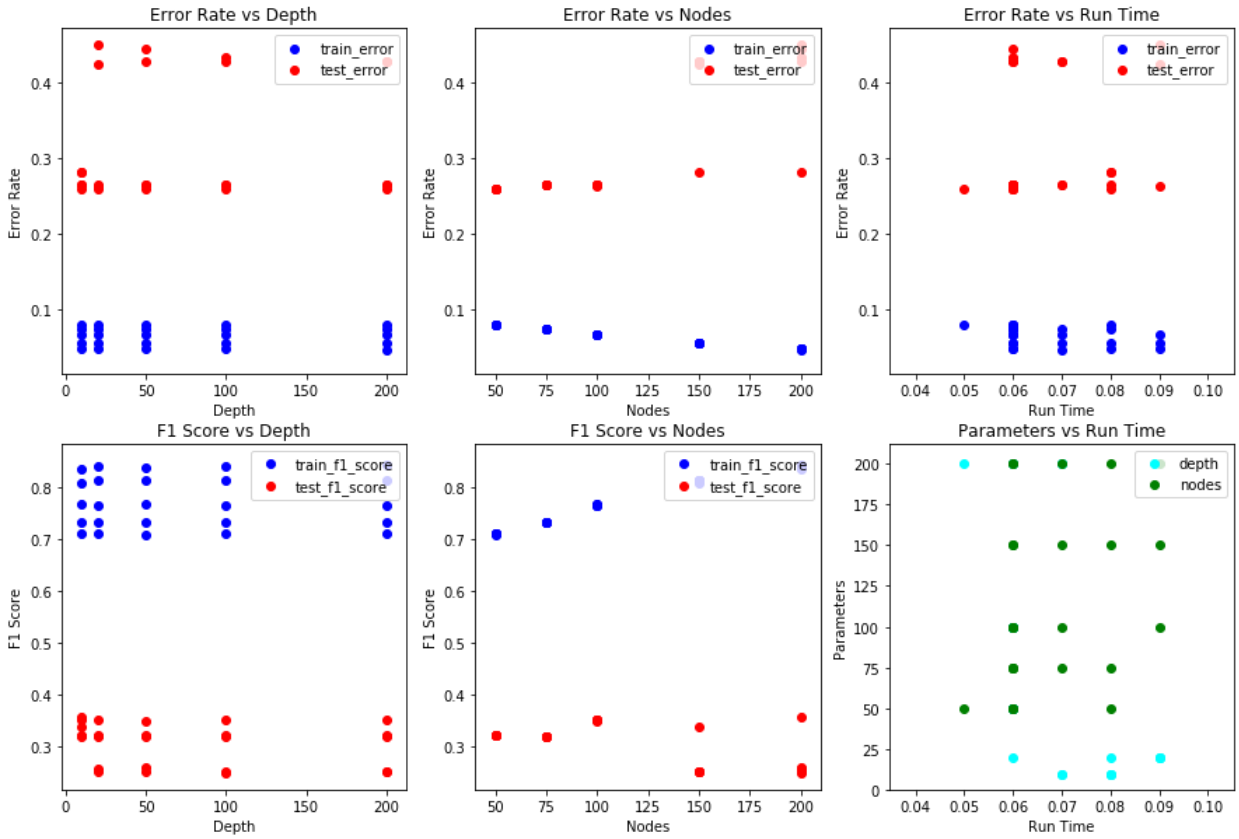


The highest F1 Score and the least Error Rate are obtained when parameters C = 0.01 and Gamma = 0.1.

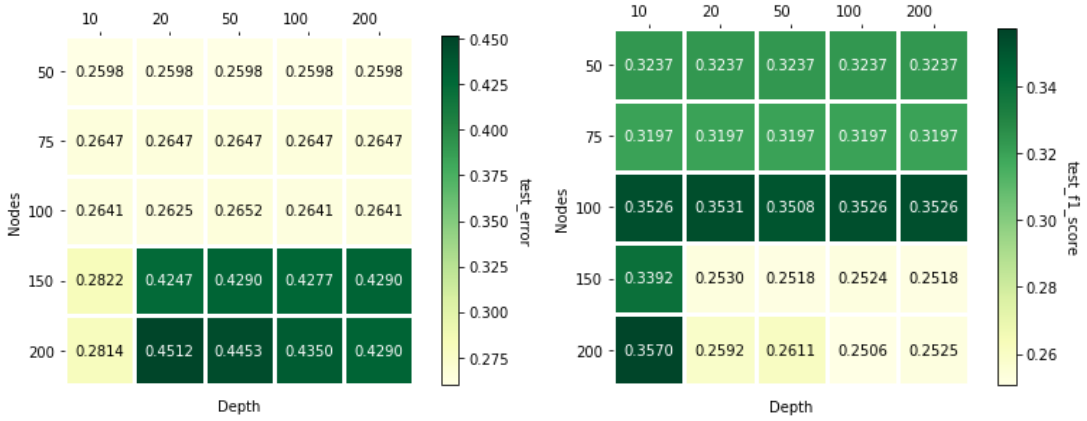


Decision Trees (Dataset 2 - Online Shopping Intention Prediction)

Following are few plots between parameters, Error Rate, F1 Score and Run Time. As nodes number gets increasing, the train error is decreasing, test error is increasing and Run Time is increasing slightly. From below, Error Rate and F1 Score seems independent of Depth and have positive linearity with Leaf Nodes.

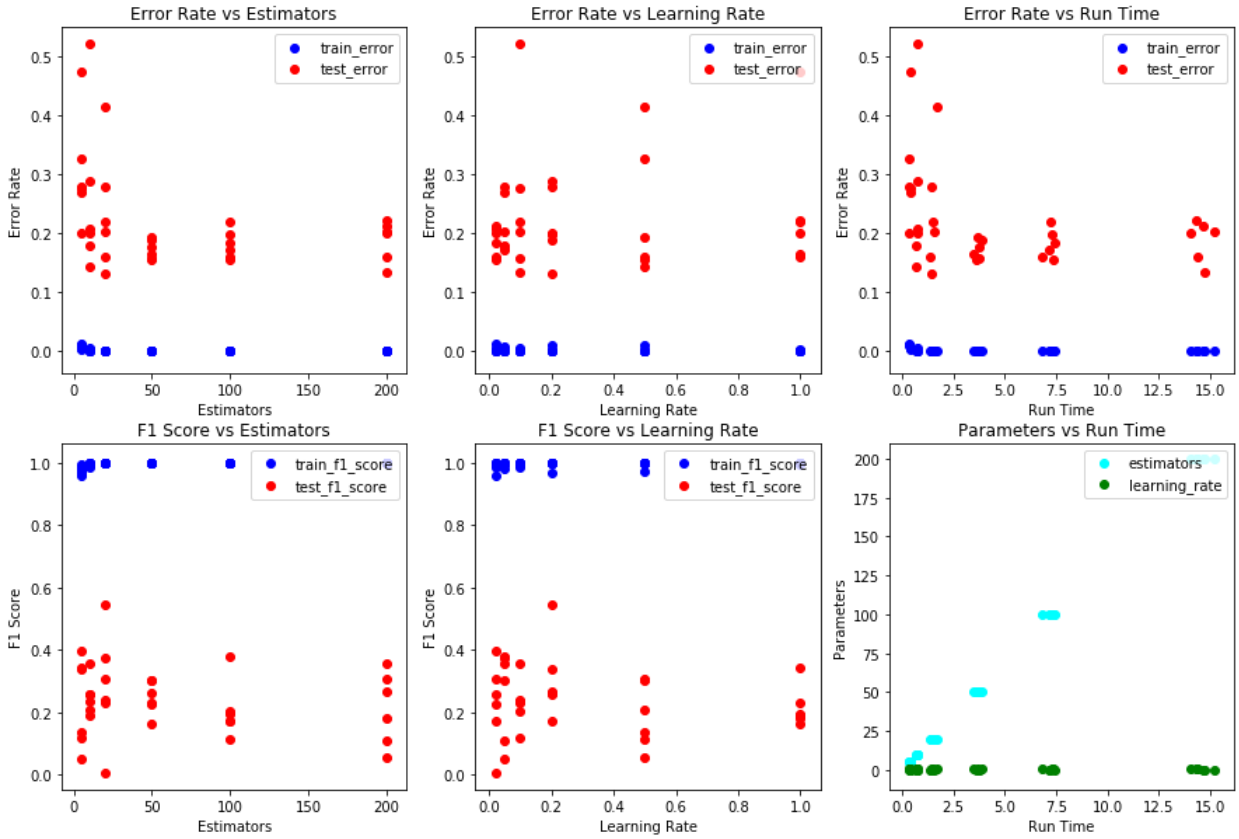


The least Error Rates are achieved when Nodes = 50, but highest F1 Score is achieved when Nodes = 100 irrespective of the Depth. For less complexity we can consider Depth = 10 as best parameter in this case.

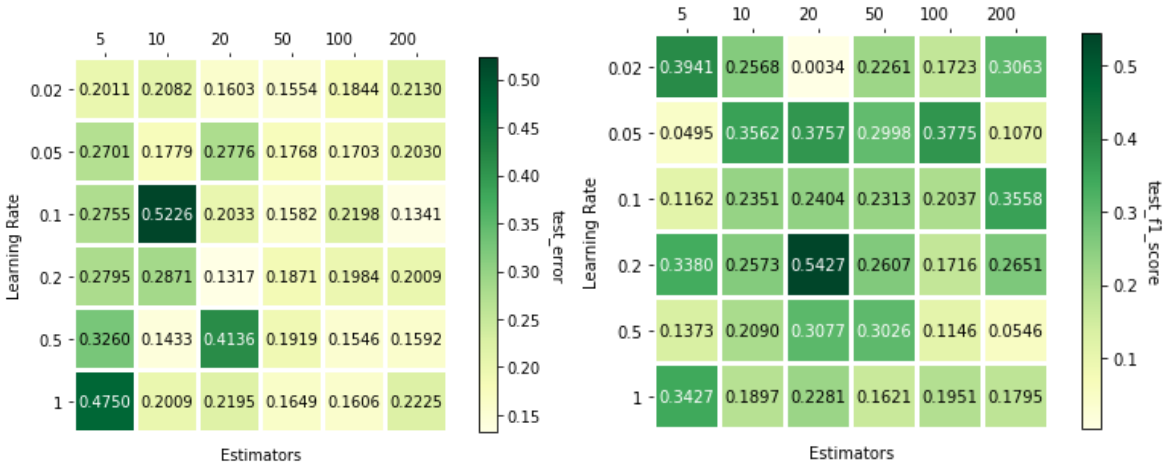


Adaptive Boosting on Decision Trees (Dataset 2 - Online Shopping Intention Prediction)

Following are plots between parameters, Error Rate, F1 Score and Run Time. Estimators and Run Time have positive linear relation. Error Rate has high variance for Learning Rate = 0.5 and Estimators < 50.



Both least Error Rate and highest F1 Score are achieved when Estimators = 20 and Learning Rate = 0.2.



Algorithm Comparison (Dataset 1)

Comparing all algorithms F1 Scores, AdaBoost on Decision Trees achieved the highest F1 Score. Doing cross validation with 5 folds on AdaBoost, best parameters are Learning Rate = 0.5, Estimators = 1000.

SVM (Linear) - Highest F1 Score and Least Error Rate

c	train_error	test_error	train_f1_score	test_f1_score	run_time
20.00	0.234327	0.393008	0.737235	0.284220	13.72

SVM (RBF) - Highest F1 Score but not Least Error Rate

c	gamma	train_error	test_error	train_f1_score	test_f1_score	run_time
0.1	0.100	0.268641	0.5418	0.687600	0.628446	20.37

SVM (Sigmoid) - Highest F1 Score but not Least Error Rate

c	gamma	train_error	test_error	train_f1_score	test_f1_score	run_time
1.0	10.00	0.472781	0.476102	0.479892	0.634986	16.88

Decision Tree - Highest F1 Score but not Least Error Rate

depth	nodes	train_error	test_error	train_f1_score	test_f1_score	run_time
25	10	0.278775	0.5418	0.697510	0.628446	0.12

AdaBoost - Highest F1 Score but not Least Error Rate

estimators	learning_rate	train_error	test_error	train_f1_score	test_f1_score	run_time
1000	0.20	0.000000	0.512245	1.000000	0.638714	223.73

Algorithm Comparison (Dataset 2)

Comparing all algorithms F1 Scores, AdaBoost on Decision Trees achieved the highest F1 Score. Doing cross validation with 5 folds on AdaBoost, best parameters are Learning Rate = 0.1, Estimators = 50.

SVM (Linear) - Highest F1 Score but not Least Error Rate

c	train_error	test_error	train_f1_score	test_f1_score	run_time
0.0010	0.156992	0.196269	0.000000	0.542245	4.68

SVM (RBF) - Highest F1 Score and Least Error Rate

c	gamma	train_error	test_error	train_f1_score	test_f1_score	run_time
10.0	0.001	0.156645	0.148959	0.004418	0.007207	7.74

SVM (Sigmoid) - Highest F1 Score and Least Error Rate

c	gamma	train_error	test_error	train_f1_score	test_f1_score	run_time
0.01	0.100	0.156992	0.107867	0.0	0.523297	8.52

Decision Tree - Highest F1 Score but not Least Error Rate

depth	nodes	train_error	test_error	train_f1_score	test_f1_score	run_time
10	200	0.047851	0.281427	0.834998	0.357011	0.08

AdaBoost - Highest F1 Score but and Least Error Rate

estimators	learning_rate	train_error	test_error	train_f1_score	test_f1_score	run_time
20	0.20	0.000000	0.131657	1.000000	0.542723	1.46