**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](https://archive.ics.uci.edu/ml/datasets/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](https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset))

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.

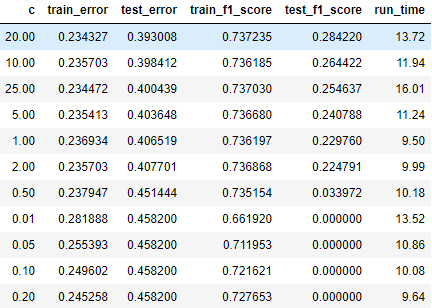
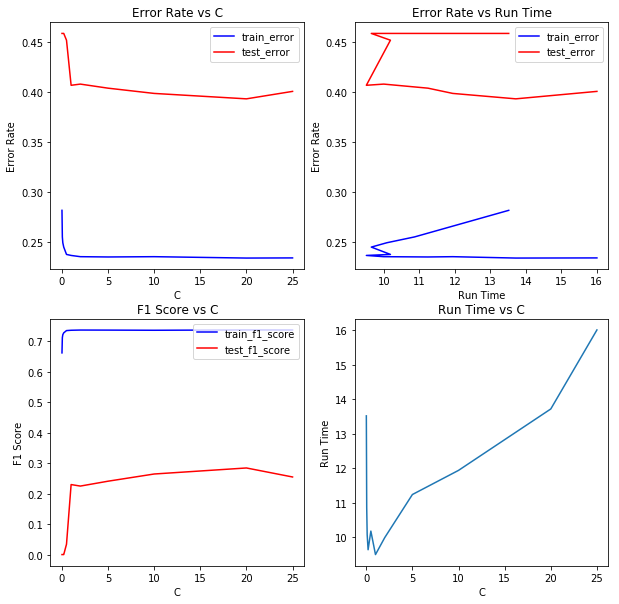
|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Kernel** | **C** | **Gamma** |
| Appliances Energy | Linear | 0.01,0.05,0.1,0.2,0.5,1,2,5,10,20,25 | - |
| Appliances Energy | RBF | 0.001,0.01,0.1,1,10 | 0.001,0.01,0.1,1,10 |
| Appliances Energy | Sigmoid | 0.001,0.01,0.1,1,10 | 0.001,0.01,0.1,1,10 |
| 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.

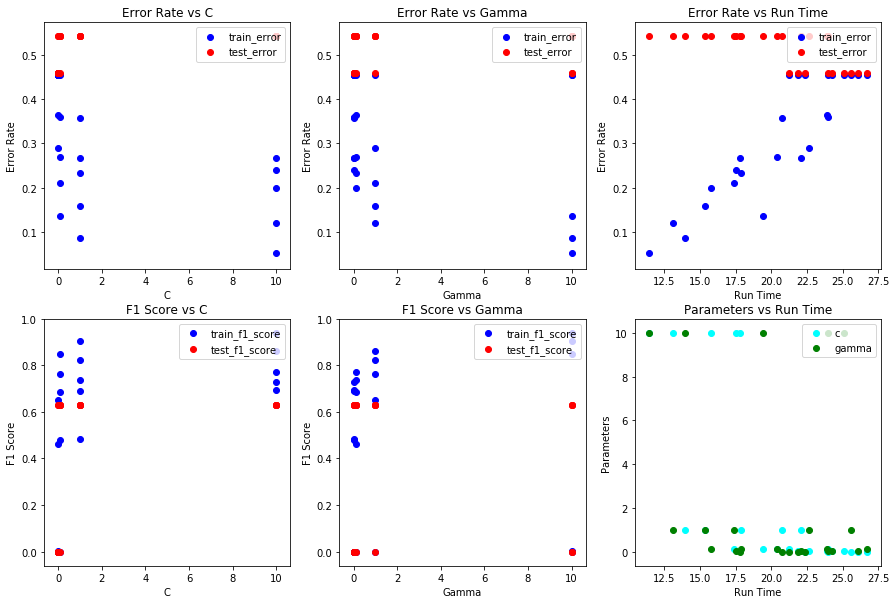
|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Algorithm** | **Depth / Number of Estimators** | **Nodes / Learning Rate** |
| Appliances Energy | Decision Trees | 5,10,15,20,25 | 5,10,15,20,25 |
| Appliances Energy | AdaBoost on Trees | 10,50,100,200,500,1000 | 0.005,0.01,0.05,0.1,0.2,0.5 |
| 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)**

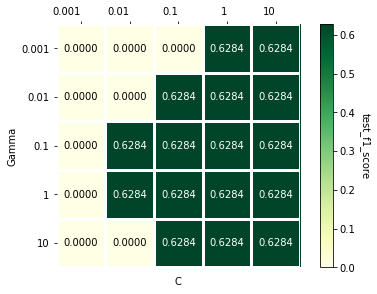
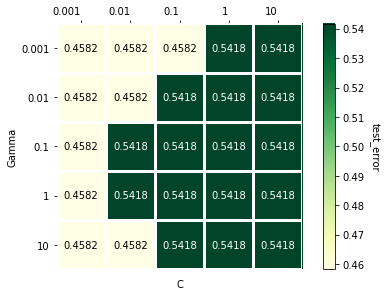
1. **Linear Kernel:** When C=20, Linear kernel is giving the least error and best F1 Score for test set. As C increases, Run Time increases and there is non-linear relationship between Error Rate and Run Time.

****

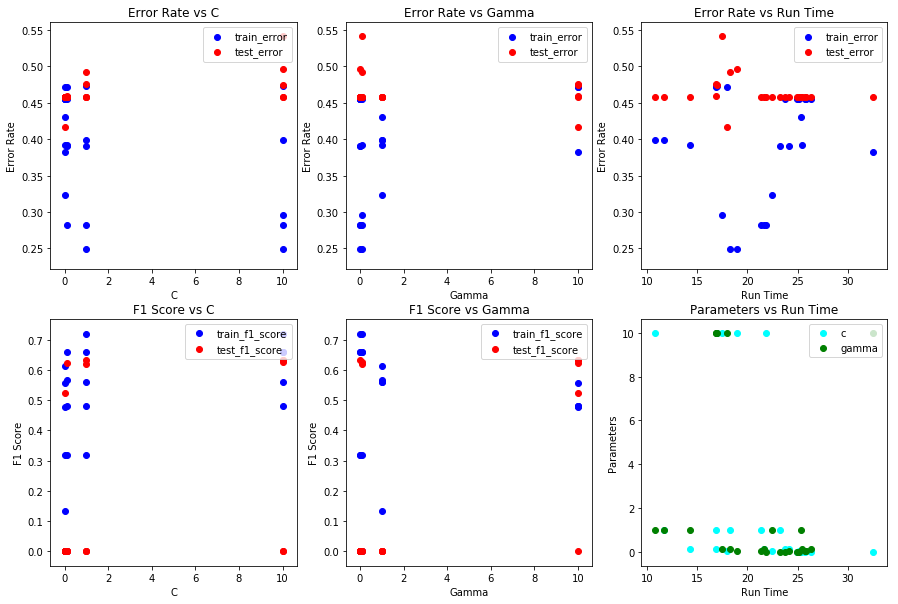
1. **RBF Kernel:** Following are various plots between the parameters, Error Rate, F1 Score and Run Time.

****

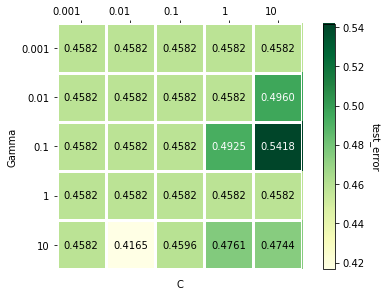
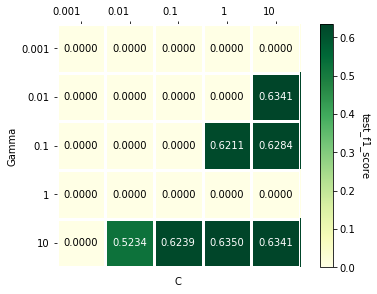
For all combinations of C and Gamma, the Error Rate and F1 Score are shown below. The combinations with higher F1 Score have slightly higher Error Rate. We can look at this as trade-off for better F1 Score.

****

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

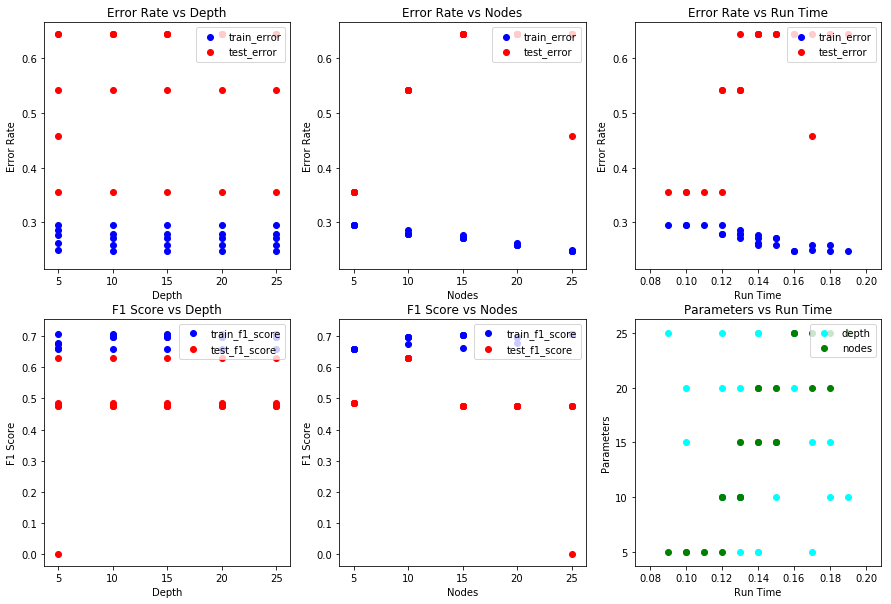
****

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

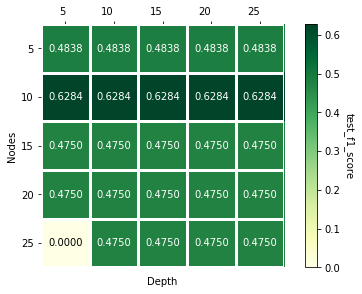
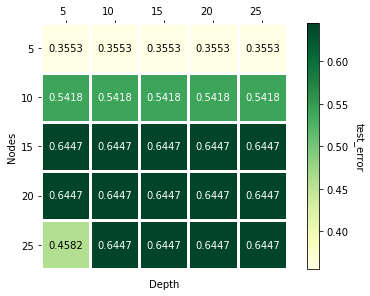
** **

**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.

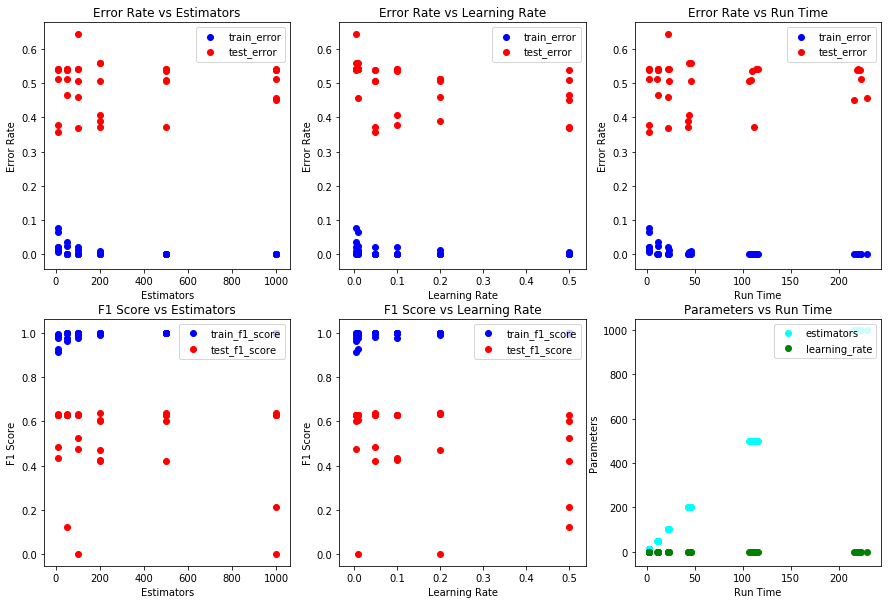
****

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.

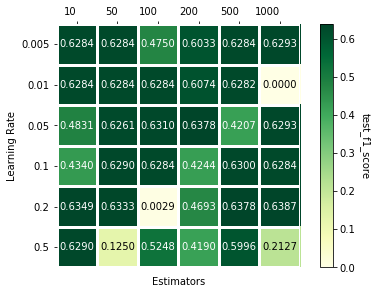
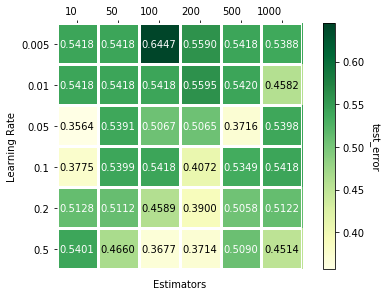
****

**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.

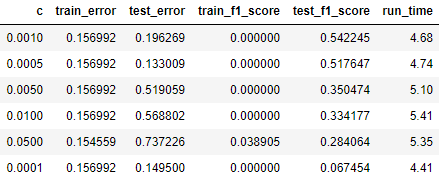
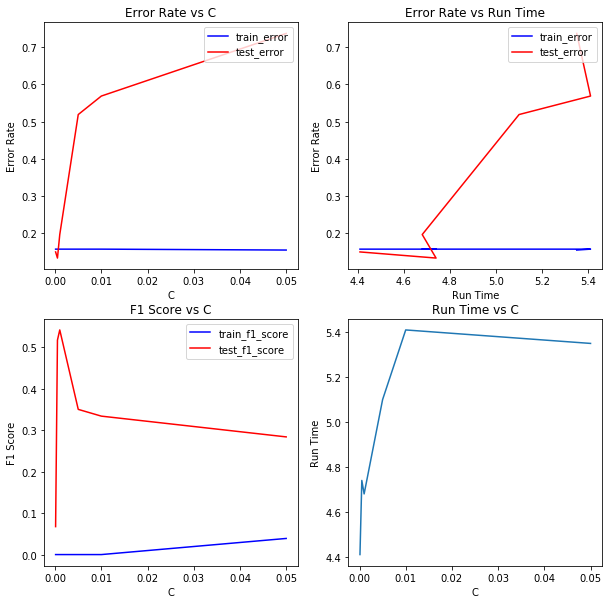
****

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.

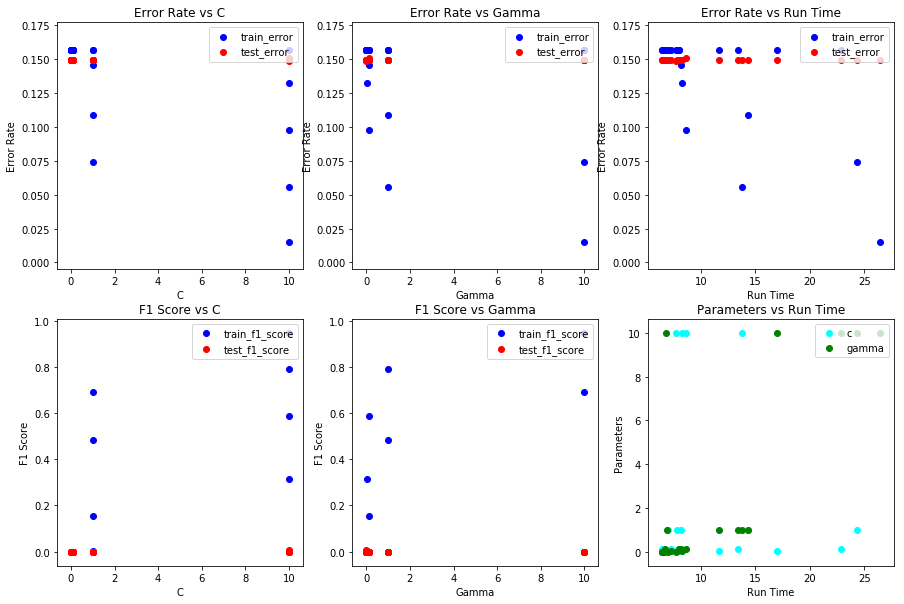
****

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

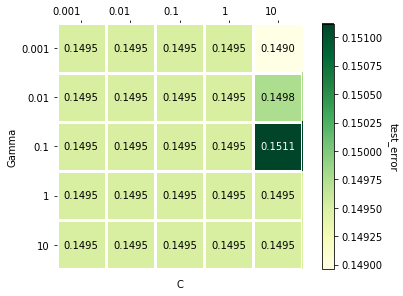
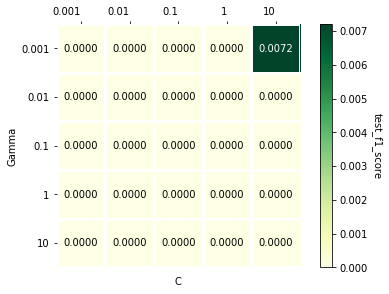
1. **Linear Kernel:** When C = 0.0005, Linear kernel is giving the least error and when C = 0.001, best F1 Score is achieved. As C increases, Run Time increases with slightly non-linear relation between them.

 ****

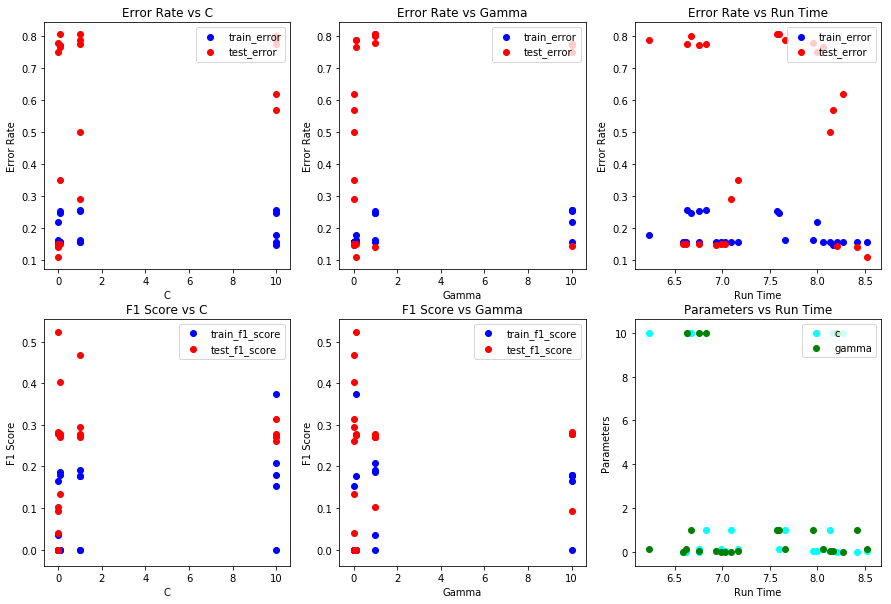
1. **RBF Kernel:** Following are various plots between the parameters, Error Rate, F1 Score and Run Time.

****

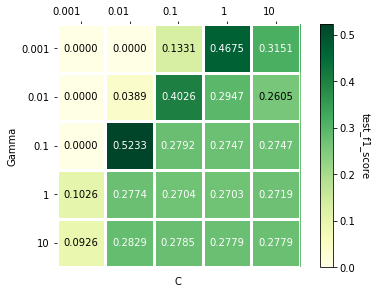
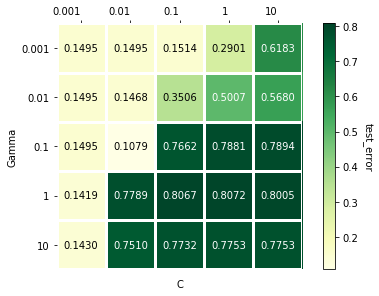
For all combinations of C and Gamma, the Error Rate and F1 Score are shown below. The combinations with higher F1 Score and lease Error Rate is C = 10 and Gamma = 0.001. But all the results are very poor.

** **

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

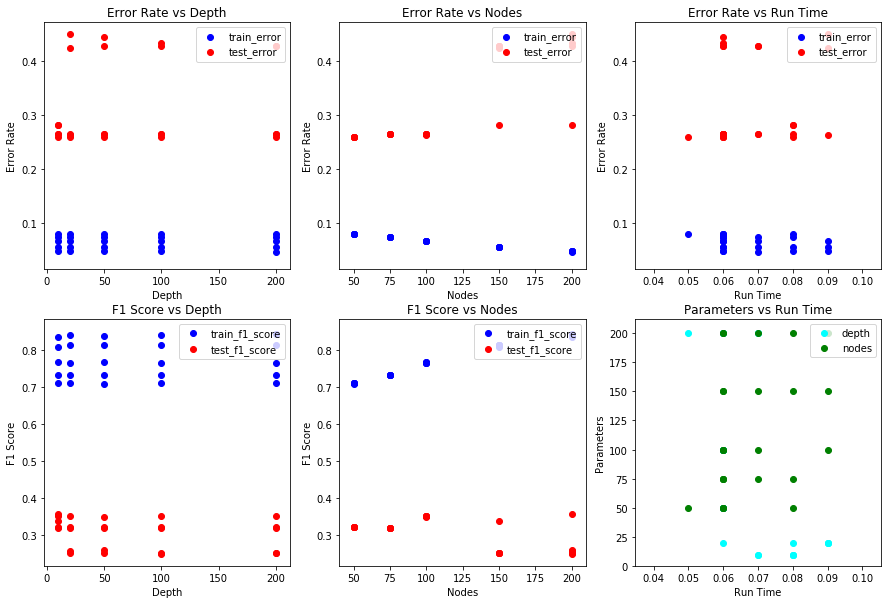
****

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

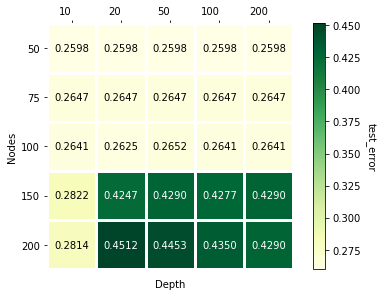
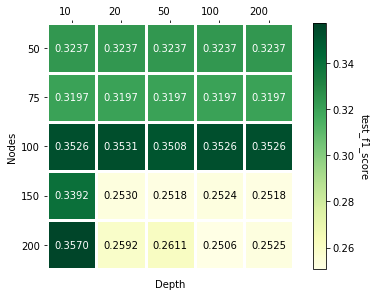


**Decision Trees (Dataset 2 - Online Shopping Intention Prediction)**

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.

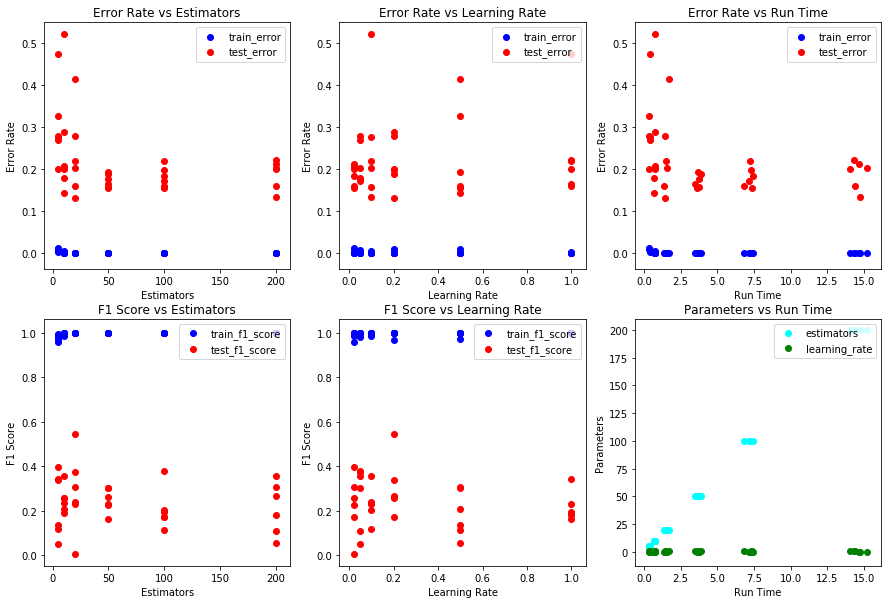
****

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.

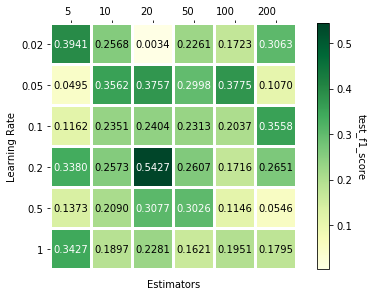
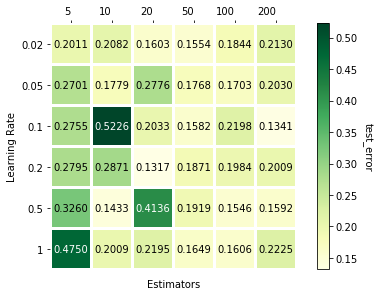
** **

**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**



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



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



**Decision Tree - Highest F1 Score but not Least Error Rate**



**AdaBoost - Highest F1 Score but not Least Error Rate**



**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**



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



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



**Decision Tree - Highest F1 Score but not Least Error Rate**



**AdaBoost - Highest F1 Score but and Least Error Rate**

