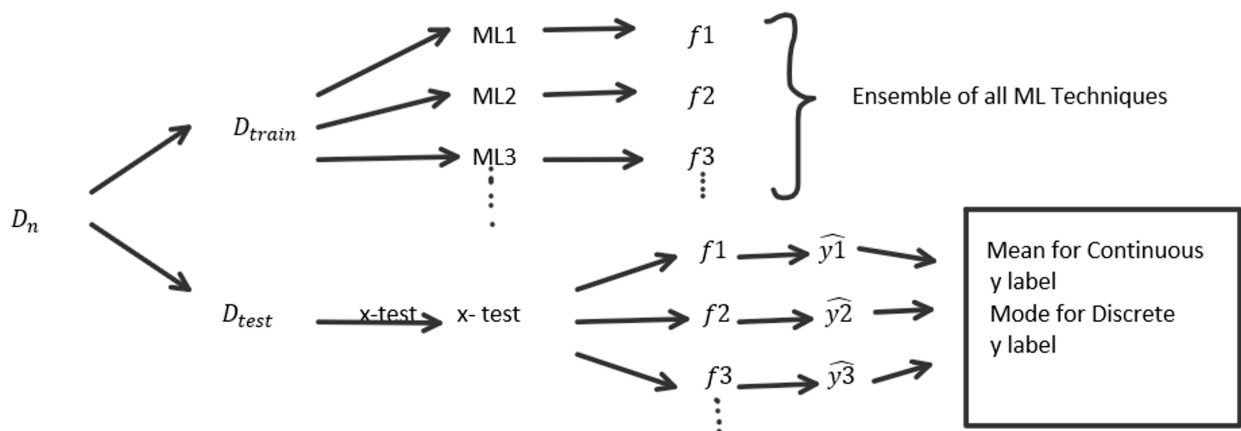


Ensemble Techniques

Friday, July 23, 2021 5:17 PM

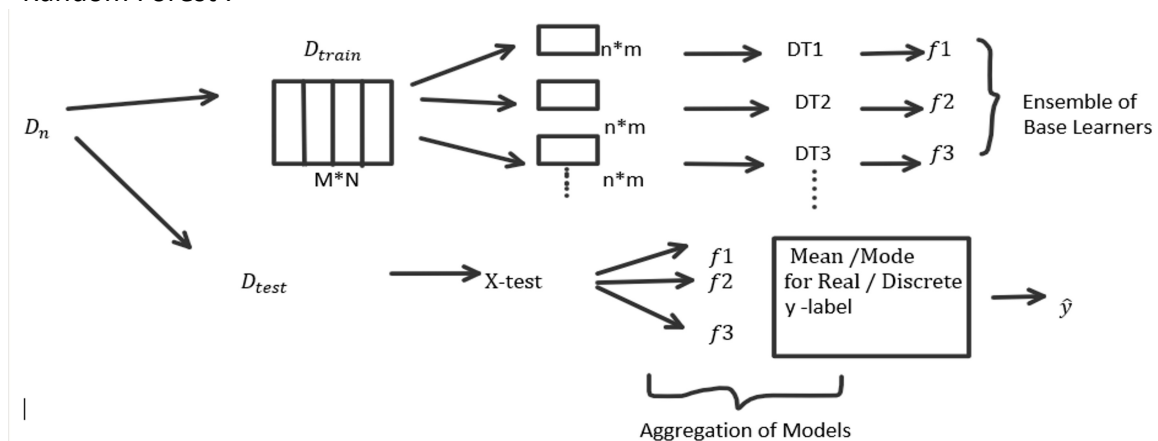
Ensemble means grouping. Instead of using only one algorithm, we use multiple algorithms in Ensemble. In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.

- Voting Ensemble Technique :



Even ML1 and ML3 can also be same but with different hyperparameters. Models are called Base Learners.

- Bagging (Bootstrap and Aggregation):
Random Forest :



Bootstrap means doing Row sampling and Column sampling (as in CLT). Base Learners in Bagging i.e. Decision Tree should have High Depth and large no of base learners because the decision tree should be overfitted (Low Bias, High Variance). We get the Overall Model which has Low Bias and Low Variance. .

Extra Randomized Tree : It is similar to Random Forest. It does everything similar except Picking a random threshold in case of Real Valued Feature. It is considered theoretically better than Random Forest.

Boosting (Adaboost):

Steps for Adaboost Algorithm ^{1/n.}

Step 1 : initialize the weights as $\frac{1}{n}$ to every n observations

Step 2 : Select the 1 features according to Lowest Gini/Highest information Gain and calculate the Total Error

Step 3 : calculate the Performance of the stump.

Step 4: Calculate the new weights for each misclassification (increase) and right classification (decrease)

Step 5 : Normalize the new weights so that the sum of weights is 1

Step 6: Now Repeat from Step 2 and so on till the configured number of estimators reached or the accuracy achieved.

