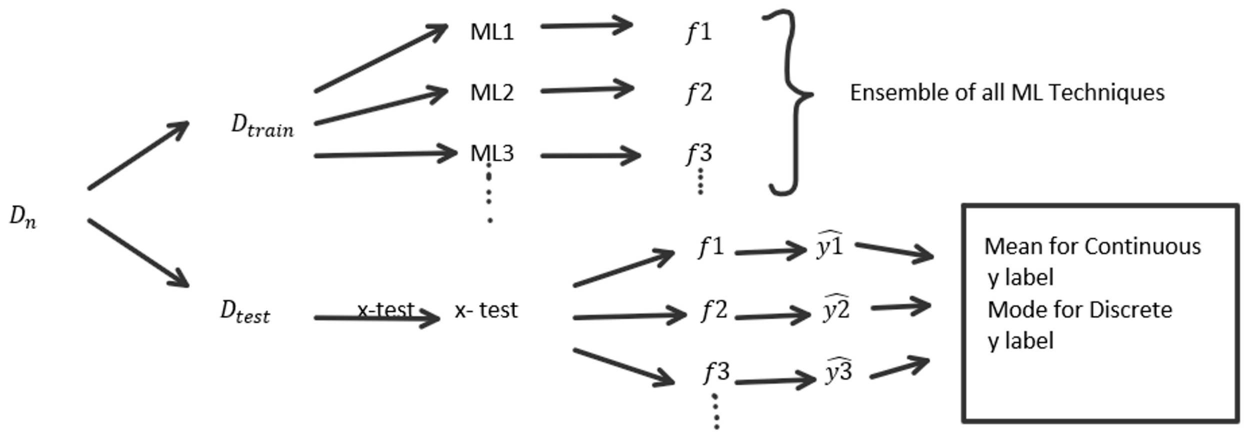
Ensemble Techniques

Friday, July 23, 2021 5:17 PM

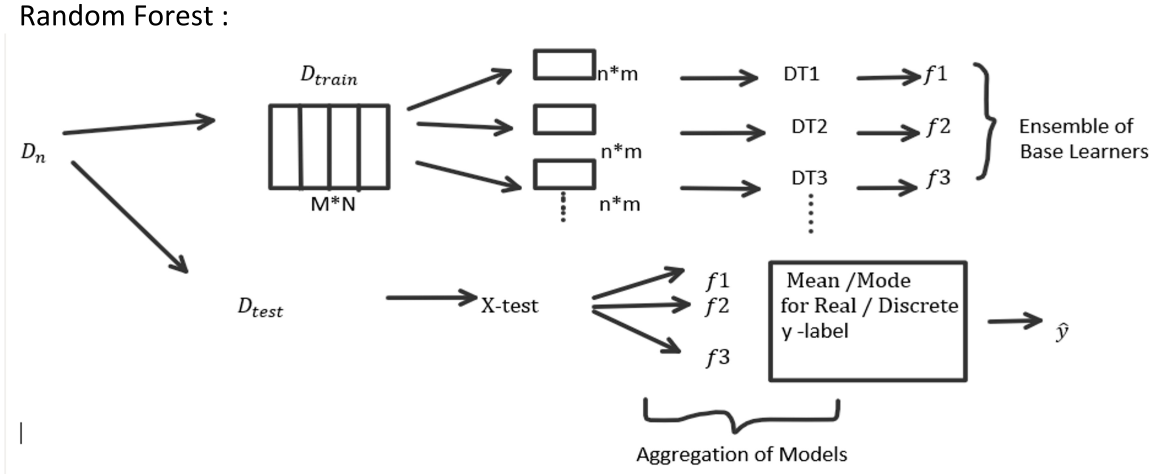
Ensemble means grouping. Instead of using only one algorithm, we use multiple algorithms in Ensemble.
In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.

- 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 is do everything similar except Picking a random threshold incase of Real Valued Feature.
It is considered theoretically better than Random Forest.

Boosting:

XG - Boost Regression:
Step 1 : No of trees
Step 2 : Most Efficient tree

Exp	Gap	Sal	Residual.
2	Y	40 k	-11 k
2.5	Y	42 k	-9 k
3	N	52 k	1 k
4	N	60 k	9 k
4.5	Y	62 k	11 k
		51 k	

[-11, -9, 1, 9, 11]

$$S.W. \text{ of Root} = \frac{-11-9+1+9}{5+1} = \frac{1}{6} = 0.16$$

$\boxed{\text{Exp}}$

<=2 : [-11]

>2 : [-9, 1, 9, 11]

$$SW = \frac{\Sigma(\text{Residual})^2}{\text{No. of Residuals} + \lambda}, \quad \lambda - \text{hyperparameter}$$

$$SW = \frac{(-11)^2}{1+1} = \frac{|-121|}{2} = 65.5$$

$$SW = \frac{(\cancel{-9}+1+\cancel{9}+11)^2}{4+1} = \frac{(12)^2}{5} = \frac{144}{5} = 28.5$$

$$\text{Gain} = \underline{65.5 + 28.5 - 0.16}$$

$$= 94 - 0.16$$

$$= 93.84$$

[-11, -9, 1, 9, 11]

Exp

<=2.5   >2.5

[-11, -9]   [1, 9, 11]

$$SW = \frac{(-11-9)^2}{2+1} = \frac{400}{3} = 133.33$$

$$SW = \frac{(1+9+11)^2}{2+1} = 110.25$$

$$gain = SW + SW - SW_{root} = 133.33 + 110.25 - 0.16 = 143.42$$

We have to select the ~~highest~~ split with highest gain.

[-11 -9 1 9 11]

Exp

<=2.5   >2.5

[-11 -9]   [1, 9, 11]

Y   N

[ ]   [11]   [1 9]

$$o/p = \frac{|-11-9|}{2} = \frac{-20}{2} = -10$$

o/p = 11   o/p = 5

$= -10$

$$o/p = 51 + (0.5)^{\alpha_1} \overset{T_1}{[-10]}$$
$$= 51 - 5$$
$$= 46$$

Exp.	Gap	salary	Res 1	o/p
2	Y	40	-11	46
2.5	Y	42	-9	46
3	N	52	1	
4	N	60	9	
4.5	Y	62	11	

We calculate again new tree using data with o/p.

so. overall $o/p = \text{Base Model} + \alpha_1(T_1) + \alpha_2(T_2) + \alpha_n(T_n)$

$\gamma = 150.0$, Hyperparameter for pruning

if $I.G < \gamma$, then we prunn the tree

$\propto$ – Learning rate to reduce Residual or error.