Random Forests

Recall from Bagging we said that using Decision Trees leverages the bagging technique well since Decision Trees are high variance low bias weak learners so averaging multiple over-fitted tress will result in a reduce of variance (check bagging notes for more details).

Why Random Forests

In Bagging Tress we consider the full set of **Predictors**(Features) and as discussed in Decision Trees they select the best split(Feature) which increase the information gain so when we **bootstrap** our data set and fit multiple tress we will end up similar hence the predictions of all the tress will be highly correlated and averaging highly correlated results wont reduce the variance much.

Random Forests

Random Forests tweak the bagging process by limiting the amount of predictors p each tree can use which adds randomness and get rid of **dominate predictor** typically selecting :

$$m \approx \sqrt{p} \ \text{ For classification}$$

$$m \approx p/3 \ \text{ For Regression}$$

- With m being a subset of the full predictors set p

This process can be understood by de-correlating trees since $\frac{p-m}{p}$ of splits didn't even include the dominate predictors diversity

$$\text{Subset } m \implies \text{more diversity} \implies \text{less correlation} \implies \text{Reduced Varaince}$$

In random forests at each split it consider a **random** subset of predictors m , but this comes at a small cost following the Bias-Variance Trade-Off it increase the **bias** a bit more compared to Bagging due to sub-optimal splits when not considering all predictors

Recalling the **variance formula** derived in the Bagging notes:

$$\text{Var(RF)} = \frac{1}{B}(1-\rho)\sigma^2 + \rho\sigma^2$$

- While Bagging reduce the first term with B being number of **weak learners**
- Random forests reduce the second term $\rho\sigma^2$ by decreasing the correlation between fitted trees