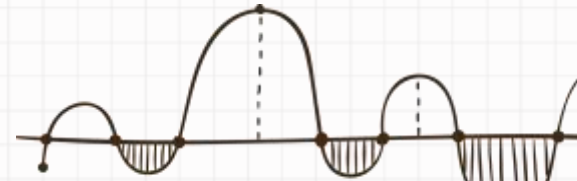
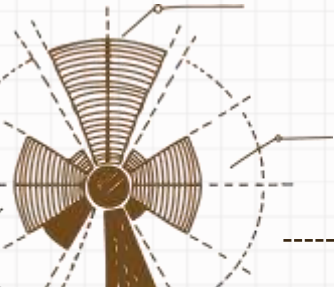
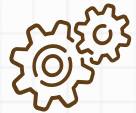




MACHINE LEARNING



DECISION TREE (METRICS FOR SPLITTING) *Variance Reduction & Mean Squared Error*



Variance Reduction (for Regression):

Used to determine the best splits in regression trees by minimizing the variance in the target variable after the split.

Variance is used for calculating the homogeneity of a node. If a node is entirely homogeneous, then the variance is zero.

Formula :

$$\text{Variance} = \frac{\sum (X - \mu)^2}{N}$$

Variance Reduction Formula :

$$\text{Variance Reduction} = \text{Variance}(\text{parent}) - \sum \frac{n_i}{N} \text{Variance}(\text{child}_i)$$

Here are the steps to split a decision tree using the reduction in variance method :

1. For each split, individually calculate the variance of each child node.
2. Calculate the variance of each split as the weighted average variance of child nodes.
3. Select the split with the lowest variance.
4. Perform steps 1-3 until completely homogeneous nodes are achieved.



Mean Squared Error(MSE) (for Regression):

Measures the average of the squared differences between the predicted and actual values in regression tasks.

Formula :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

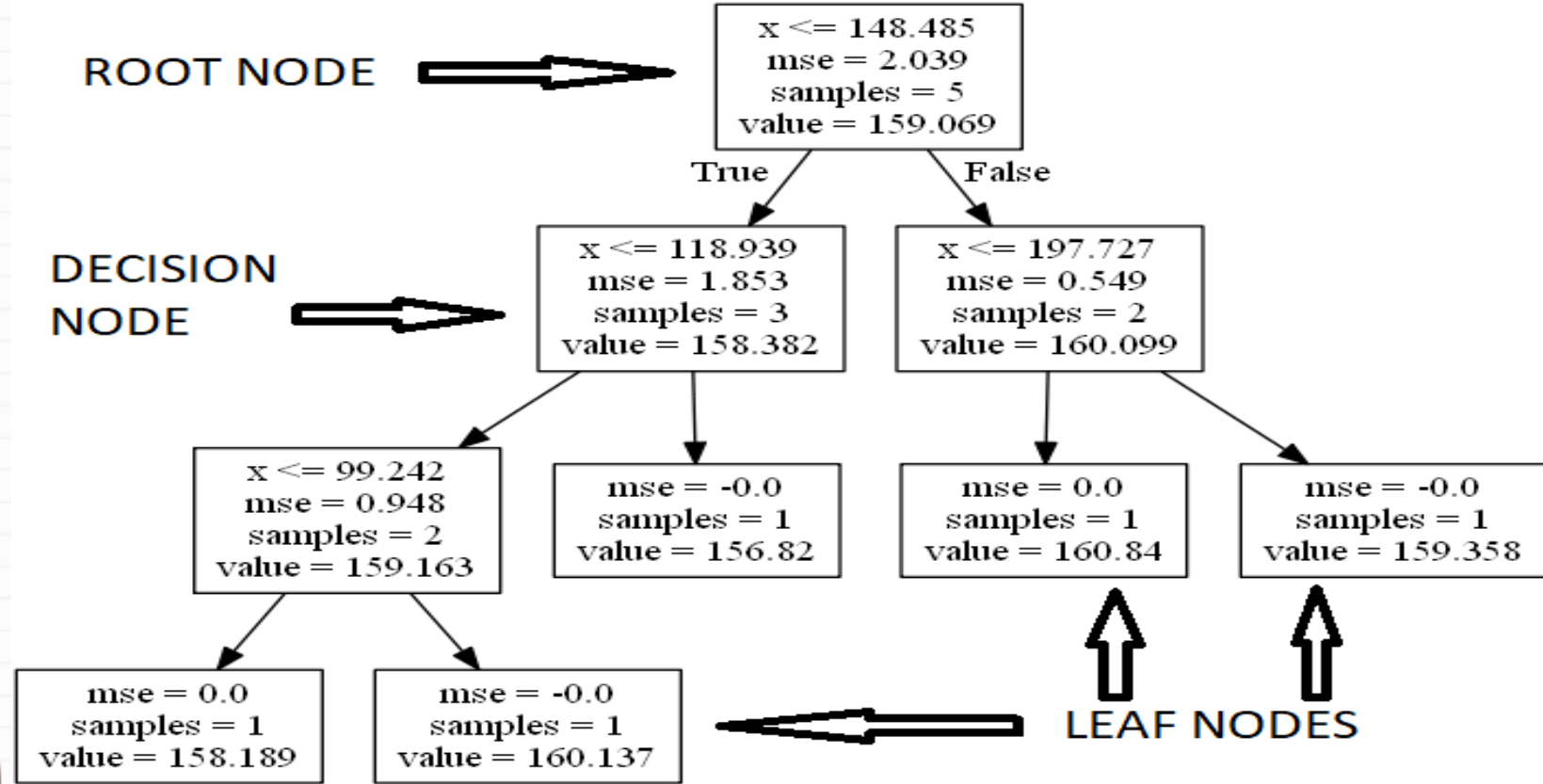
where ,n is the number of observations
 y_i is the actual or observed value for the i^{th} data point.

\hat{y}_i is the predicted value for the i^{th} data point.

MSE Reduction : Similar to variance reduction, the goal is to minimize the MSE after each split.



MSE Calculation Example:





Thank you!



Follow for more updates!!

