DECISION TREE (QUESTION AND ANSWER)

- **What is the difference between supervised and unsupervised machine learning?**
- Supervised Machine learning: - Supervised machine learning requires training labeled data.
- Unsupervised Machine learning: - Unsupervised machine learning doesn't required labeled data.

- **What is a Decision Tree? How does it work?**

- Decision tree is a supervised machine learning algorithm mainly used for the **Regression and Classification**. This graphical representation which classifies the data is called as the Decision Tree. The final result is a tree with decision nodes and leaf nodes. Decision tree can handle both categorical and numerical data.
- In this technique, we split the population or sample into two or more homogeneous sets (or sub-populations) based on most important input variable.
- There are two types:-
- Classification decision tree (Categorical): - its prediction with mode value.
- Regression decision tree (Numerical): - its prediction with mean value.

- **Important Terminology related to Decision Trees**
- **Root Node:** It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
- **Splitting:** It is a process of dividing a node into two or more sub-nodes.
- **Decision Node:** When a sub-node splits into further sub-nodes, then it is called the decision node.
- **Leaf / Terminal Node:** Nodes do not split is called Leaf or Terminal node.
- **Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
- **Branch / Sub-Tree:** A subsection of the entire tree is called branch or sub-tree.

- **How do Decision Trees work?**
- some algorithms used in Decision Trees:
- **For Classification Tree: -**
- **ID3** → (extension of D3, Information gain)
- **C4.5** → (successor of ID3, Information gain)
- **CART** → (Classification And Regression Tree, Gini impurity)
- **CHAID** → (Chi-square automatic interaction detector Performs multi-level splits when computing classification trees, Statistical significance)
- **MARS** → (multivariate adaptive regression splines)

- **For Regression Tree**: -
- Reduction in variance

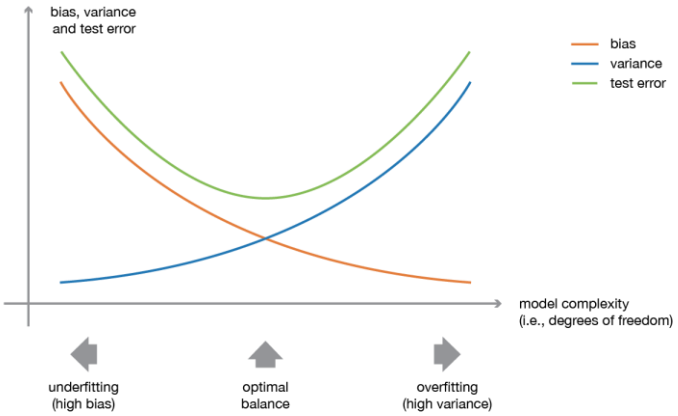
- **How to avoid/counter Overfitting in Decision Trees?**
- Here are two ways to remove overfitting:-
- Pruning Decision Trees.
- Random Forest

DECISION TREE (QUESTION AND ANSWER)

- **Setting constraints on tree size :-**
- **Minimum samples for a node split**
- Defines the minimum number of samples (or observations) which are required in a node to be considered for splitting.
- Used to control over-fitting. Higher values prevent a model from learning relations which might be highly specific to the particular sample selected for a tree.
- Too high values can lead to under-fitting hence, it should be tuned using CV(cross-validation).
- **Minimum samples for a terminal node (leaf)**
- Defines the minimum samples (or observations) required in a terminal node or leaf.
- Used to control over-fitting similar to min_samples_split.
- Generally lower values should be chosen for imbalanced class problems because the regions in which the minority class will be in majority will be very small.
- **Maximum depth of tree (vertical depth)**
- The maximum depth of a tree.
- Used to control over-fitting as higher depth will allow model to learn relations very specific to a particular sample.
- Should be tuned using CV.
- **Maximum number of terminal nodes**
- The maximum number of terminal nodes or leaves in a tree.
- Can be defined in place of max_depth. Since binary trees are created, a depth of 'n' would produce a maximum of 2^n leaves.
- **Maximum features to consider for split**
- The number of features to consider while searching for a best split. These will be randomly selected.
- As a thumb-rule, square root of the total number of features works great but we should check upto 30-40% of the total number of features.
- Higher values can lead to over-fitting but depends on case to case.

- **What is bias, variance trade off?**
- **Bias: -** Bias is error introduced in your model due to over simplification of machine learning algorithm." It can lead to underfitting. When you train your model at that time model makes simplified assumptions to make the target function easier to understand.
- **Bias means---- underfitting the data – both training and test accuracy low**
- **Low bias machine learning algorithms - Decision Trees, k-NN and SVM**
- **High bias machine learning algorithms - Linear Regression, Logistic Regression**
- **Variance: -** Variance is error introduced in your model due to complex machine learning algorithm, your model learns noise also from the training dataset and performs bad on test dataset." It can lead high sensitivity and overfitting.
- **Varience means --- overfitting the data – training accuracy is high but test accuracy,we cannot say that it high or low.**
- **Note: -** Bias measures the proximity of the average classifier created by the learning algorithm to the target function, and variance measures by how much the learning algorithm's prediction varies for different training data sets.

DECISION TREE (QUESTION AND ANSWER)



- **Bias, Variance trade off:**
- The goal of any supervised machine learning algorithm is to have low bias and low variance to achive good prediction performance.
- The k-nearest neighbors' algorithm has low bias and high variance, but the trade-off can be changed by increasing the value of k which increases the number of neighbors that contribute to the prediction and in turn increases the bias of the model.
- The support vector machine algorithm has low bias and high variance, but the trade-off can be changed by increasing the C parameter that influences the number of violations of the margin allowed in the training data which increases the bias but decreases the variance.
- There is no escaping the relationship between bias and variance in machine learning.
- Increasing the bias will decrease the variance. Increasing the variance will decrease the bias.

- **What is selection Bias?**

- Selection bias occurs when sample obtained is not representative of the population intended to be analyzed.

- **What is Ensemble Learning?**
- The art of combining diverse(various) set of learners (Individual models) together to improvise on the stability and predictive power of the model is called an Ensemble.
- A group of weak models combine to a form strong model.

- **What is Ensemble Learning?**
- Ensemble is the art of combining diverse set of learners(Individual models) together to improvise on the stability and predictive power of the model. Ensemble learning has many types but two more popular ensemble learning techniques are mentioned below.
- **Bagging**: - let considers homogeneous weak learners, learns them independently from each other in parallel and combines them following some kind of deterministic averaging process.
- **Boosting**: - considers homogeneous weak learners, learns them sequentially in a very adaptative way (a base model depends on the previous ones) and combines them following a deterministic strategy
- **Stacking**:- Let us considers heterogeneous weak learners, learns them in parallel and combines them by training a meta-model to output a prediction based on the different weak models predictions.

DECISION TREE (QUESTION AND ANSWER)

- **Differences Between Bagging and Boosting –**

S.NO	BAGGING	BOOSTING
1.	Simplest way of combining predictions that belong to the same type.	A way of combining predictions that belong to the different types.
2.	Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
3.	Each model receives equal weight.	Models are weighted according to their performance.
4.	Each model is built independently.	New models are influenced by performance of previously built models.
5.	Different training data subsets are randomly drawn with replacement from the entire training dataset.	Every new subsets contains the elements that were misclassified by previous models.
6.	Bagging tries to solve over-fitting problem.	Boosting tries to reduce bias.
7.	If the classifier is unstable (high variance), then apply bagging.	If the classifier is stable and simple (high bias) the apply boosting.
8.	Random forest.	Gradient boosting.