

Decision Tree

Axis India Machine Learning
- Bhavook Chouhan

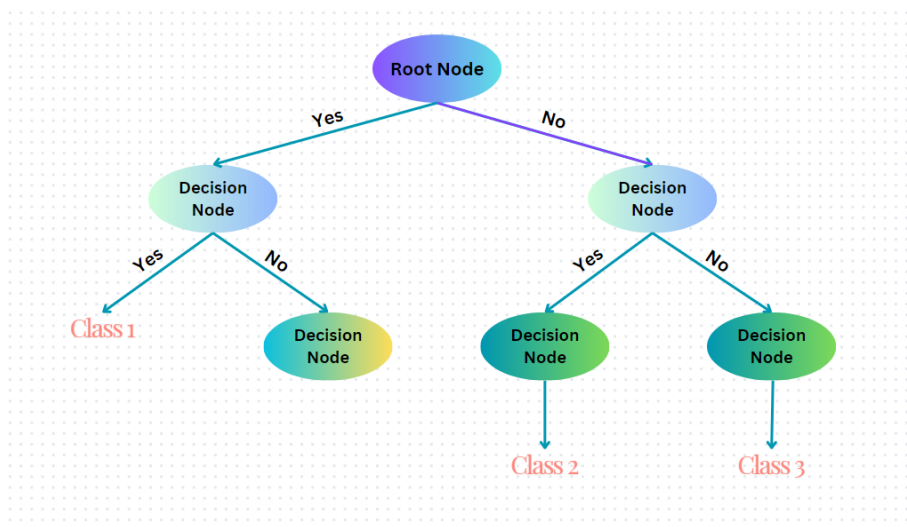
May 2024

Introduction:

A **Decision Tree** is a powerful tool used in supervised learning for both classification and regression tasks.

A Decision Tree has a tree-like structure composed of nodes. The key components of Decision Tree are:

- Root Node: The topmost node representing the entire dataset.
- Decision/Internal Nodes: Nodes that symbolize choices based on input features.
- Leaf/Terminal Nodes: Nodes without child nodes, indicating class labels or numerical values.
- Branches: Connect internal nodes to leaf nodes or other internal nodes.



Programmatically speaking, Decision Trees are nothing but a giant structure of nested if-else conditions.

Mathematically speaking, Decision Tree use hyperplanes which run parallel to any one of the axes to cut the co-ordinate system into hypercuboids.

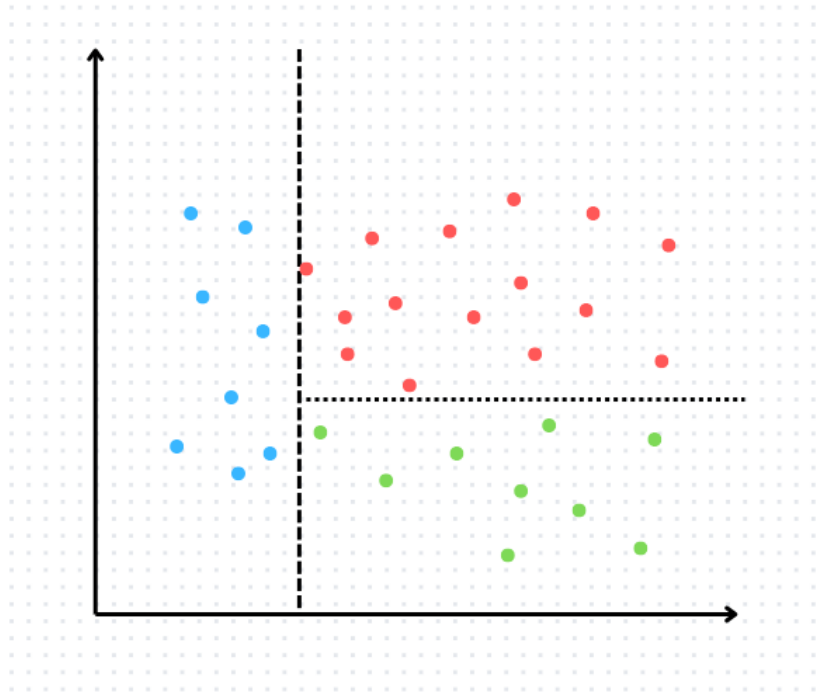


Figure 1: Geometric Intuition

How Decision Tree Works:

1. Begin with your training dataset, which should have some feature variable and classification or regression output.
2. Determine the best feature in the dataset to split the data by using the concept of **Information Gain**.
3. Split the data into subsets with respect to the best feature with highest Information Gain.
This splitting basically defines a node on the tree.
4. Recursively generate new tree node by using subset of data created from step '3'.