

Question What is Decision Tree learning?!

- Decision tree learning is the construction of a decision tree from class labeled training tuples.
 - A decision tree is a flow-chart-like structure, where each internal (non-leaf) node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label.
 - The population or sample is split into two or more homogeneous sets (or sub populations) based on most significant splitter/differentiator in input variables.
 - Classification and Regression trees follow a top-down greedy approach known as recursive binary splitting. Greedy only looks for best variable available on the split (current), and does not care about future splits which will lead to a better tree.
 - The splitting process continues until a user defined stopping criteria is reached. If there is no stop criteria the tree is prone to overfit! Pruning is one of the techniques used to tackle overfitting.
 - Decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes.
- ① Easy to understand
 - ② Useful for Data exploration
 - ③ Less data cleaning required
 - ④ Data type is not a constraint (it can handle both)