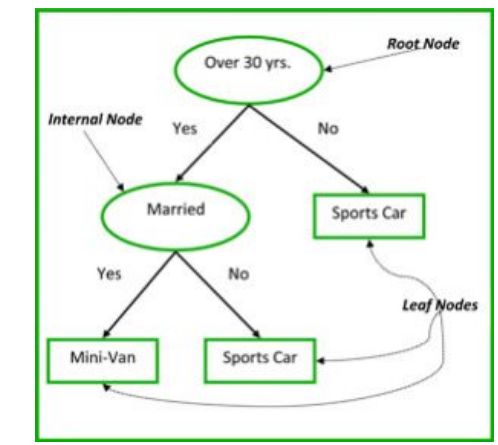
**DECISION TREES (CART)**

* CART is the modern term used for decision trees.
* CART stands for Classification and Regression Tree Algorithm
* Decision trees use a tree-like structure for representing the data in the dataset, which enables to take decisions easily since it depicts the data explicitly and visually.
* It has a root node, which represents the input variable (X), which is a non-terminal-node.
* The splitting point of the root node represents the output variable (Y), which is a leaf node.
* The above figure, makes a decision of a person will buy a sports car or mini-van based on the age and marital-status.



**Recursive Binary splitting (Greedy approach)**

* Here each attribute in the given dataset is taken for splitting and the instance with minimum splitting (less cost) is selected. This is called as greedy approach.
* In the above example, over 30 years, has the minimum split, which has least cost and hence the root node is always selected as the best predictor/classifier.
* In classification and regression trees, the cost functions will try to find the branches having the similar responses.

**Cost of regression tree:**

**Regression: Sum(Y-prediction)2** ,

Wher**e** Y is the output variable, Prediction is the mean of responses from branches within similar groups.

**Cost of classification tree:**

**Classification: G=sum (pk\*(1-pk))**

Where G is the Gini index which predicts how pure the leaf node is. That is, how mixed the training dataset assigned to each node.

G is the Gini index

Pk is the proportion of training instances with class k

For attributes or node having similar response G=0 (good purity)

For attributes or node having 50-50 split, like Binary decision tree, G=0.5 (worst purity)

**When to stop splitting:**

For large number of datasets there will be more splits which will result in a very complex structure. Hence we need to minimize the number of training instances to each leaf, like we should stop splitting when the count is less or greater than a particular threshold value.

**Pruning:**

It removes few branches from the tree that is least important to our dataset, thereby reducing the tree complexity and increasing the predicting power.

**Advantages of decision trees:**

* Simple to understand, interpret and Visualize
* Best algorithm for feature selection in Dimensionality Reduction
* Can handle multiple data( categorical and continuous)
* Less human effort in preparing the dataset

**Drawbacks of decision trees:**

* Leads to over fitting for complex data structure
* A small change to a data in a tree will have severe impact to entire tree structure
* Greedy approach to the tree structure will not give an optimal solution in all scenarios