**Prune Regression Trees**

Pruning means to change the model by deleting the child nodes of a branch node. The pruned node is regarded as a leaf node. Leaf nodes cannot be pruned.

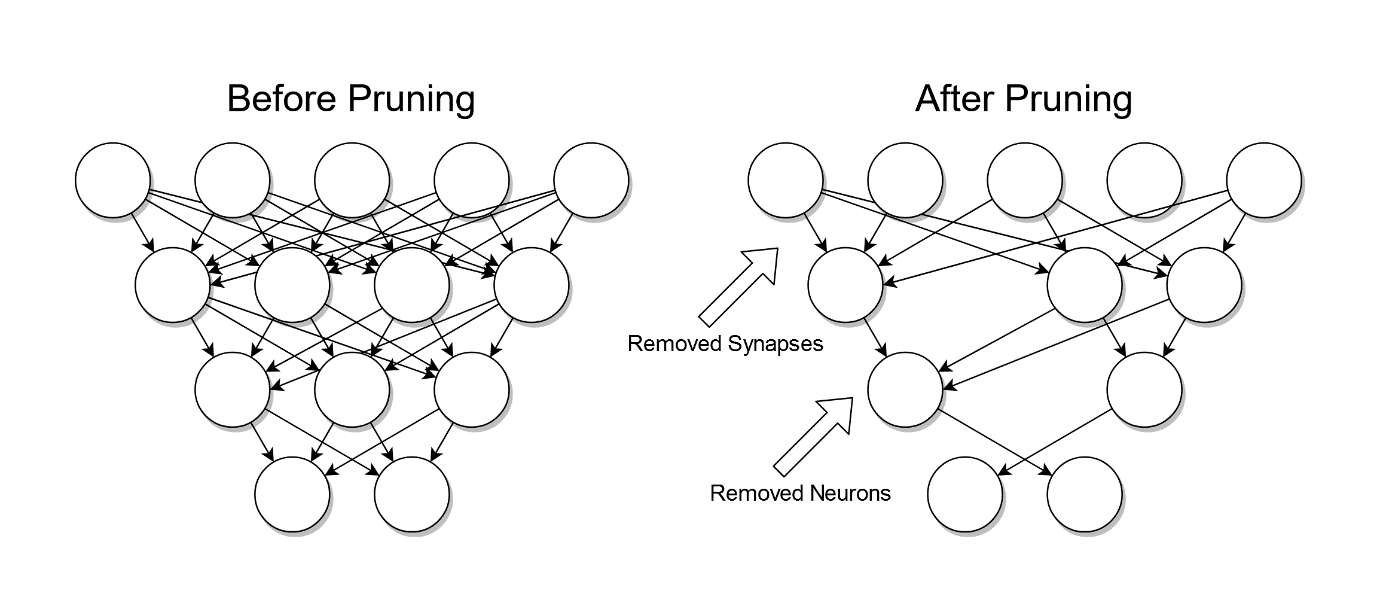
Pruning reduces the size of decision trees by removing parts of the tree that do not provide power to classify instances. Decision trees are the most susceptible out of all the machine learning algorithms to overfitting and effective pruning can reduce this likelihood.

Pruning is a data compression technique in machine learning and search algorithms that reduces the size of decision trees by removing sections of the tree

There are several methods of pruning so one of the method is **cost complexity pruning** aka weakest link pruning

**How to identify when you should prune the decision tree ?**

A common strategy is to grow the tree until each node contains a small number of instances then use pruning to remove nodes that do not provide additional information. Pruning should reduce the size of a learning tree without reducing predictive accuracy as measured by a cross-validation set.



**Types of pruning**

1)pre-pruning

2)post-pruning

pre-pruning or early stopping involves stopping the tree before it has completed classifying the training set and post-pruning refers to pruning the tree after it has finished.

Pre-pruning that stop growing the tree earlier, before it perfectly classifies the training set. Post-pruning that allows the tree to perfectly classify the training set, and then post prune the tree.

**To implement the pre-pruning**

1)Set minimum sample In order to shouldn’t go to depth

2)specify max-depth (specify how many layers should have)

**To implement the post-pruning**

Reduced error pruning

Cost complexity pruning

Reference = [ccp](https://youtu.be/D0efHEJsfHo)