

Supervised Learning

Linear Algorithms



Linear Reg.

Logistic Reg.

Support vector

Machines(SVM)

K-NN



Simple algorithms



These algorithms
are highly affected
by the various
issues present in
the data



These are
parametric



They assume the
data to be normally
distributed.

issues
↓
outliers
missing value
duplicates
multicollinearity

Tree Based Algo.



Decision Trees

Random Forest

Gradient Boost

XGBoost



Complex &
Powerful
Algorithms



These algorithms
are more
robust & dynamic
and hence
are not bother
ed by many
issues in the
data.



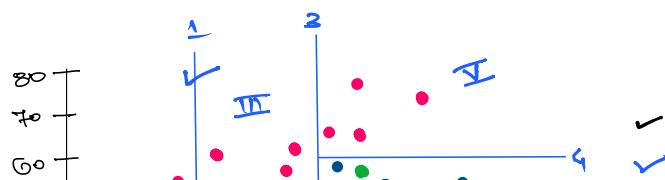
Non Parametric



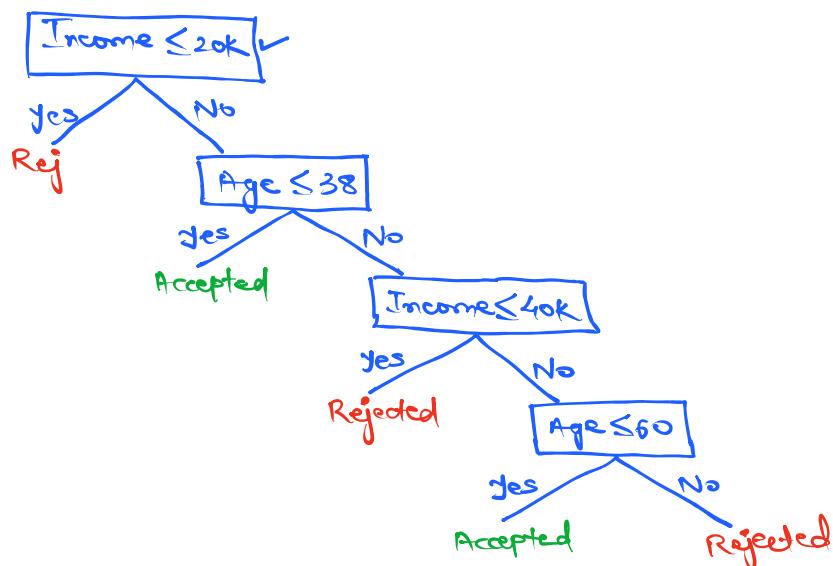
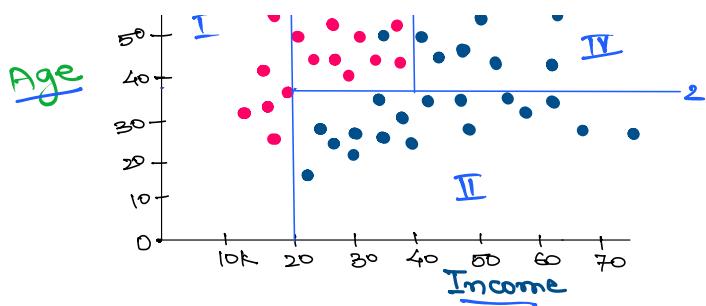
They do not care
about the
distribution of
the data.

Decision Tree Algorithm

✓✓✓
Homogeneous
↓
Maximizing the
purity of



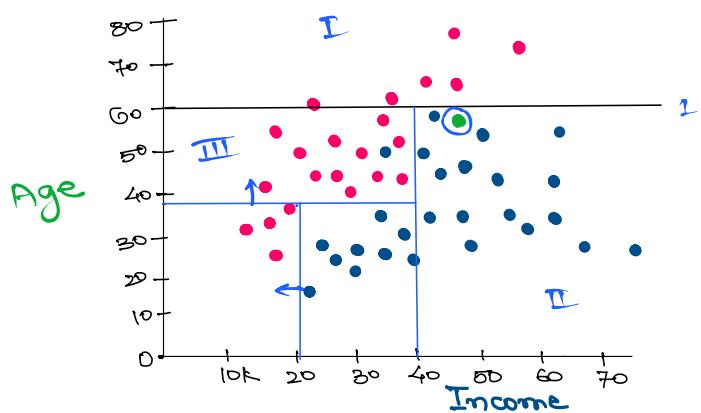
the split

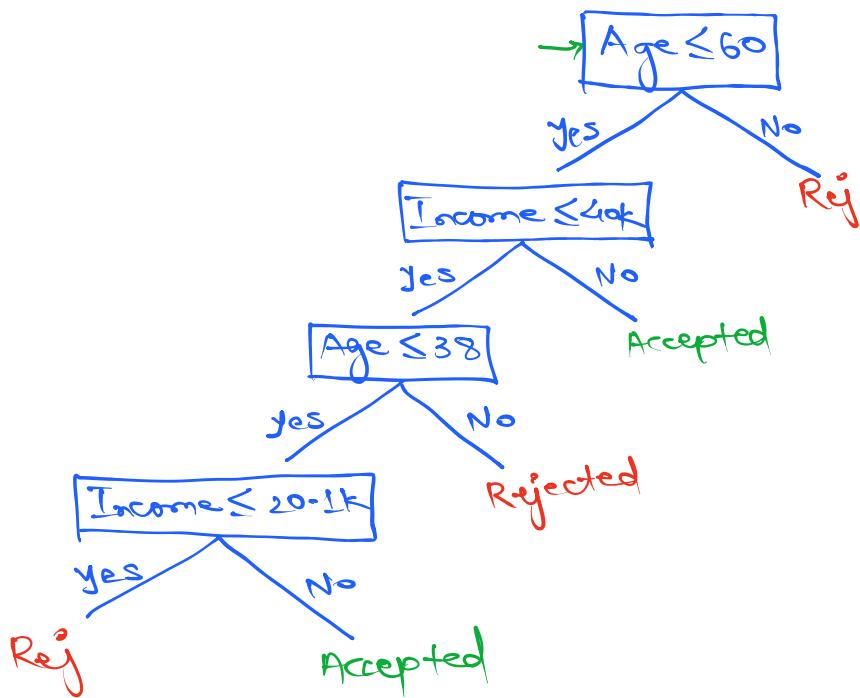


if you have more than 3 input cols



You can't visualize the graph





The most important Decision Node in a tree is the Root Node.

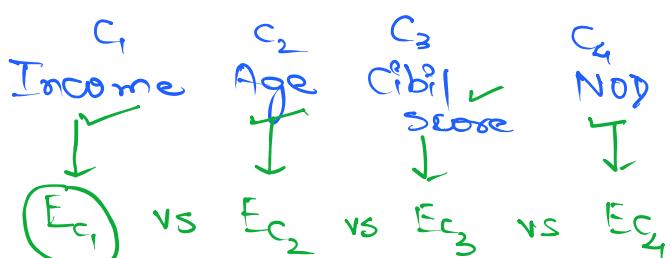
How to create the best Decision Tree on a dataset

① CART (Classification & Regression Trees) → New

↳ using Gini Index ↴

② ID3 (Iterative Dichotomizer 3) → old

↳ using Entropy & Information Gain ↴



Compare these



whichever col has the least
entropy value, that will be
our root node.



Suppose, Entropy of Income col is
least, so the root Node will
be Income column.



Income $\leq 25k$