**DECISION TREE**

CART

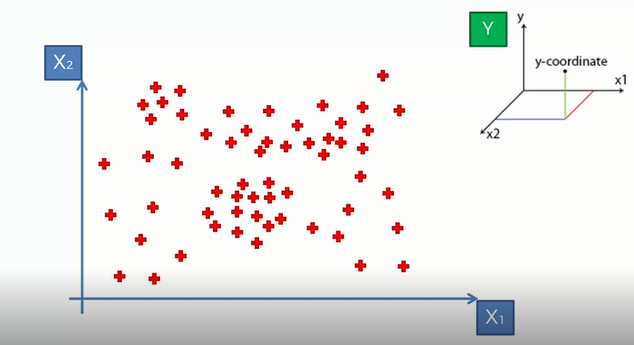
Classification trees Regression tress

Decision Tree regression:-

We have a scattered plot of a given dataset. We have got 2 independent variables x1 and x2 and we have to predict a third variable y which is a dependent variable.

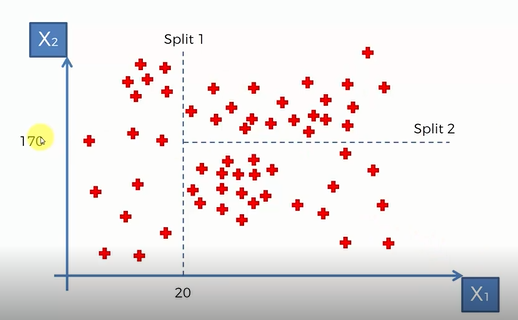
(y is a third dimension).

What we have here is projection of all the points on x1x2 plane.

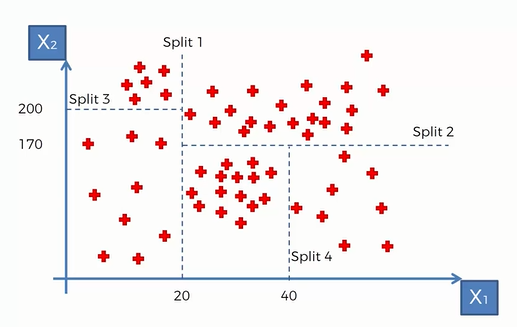


When we run the regression tree or decision tree algorithm in regression, scatter plot splits up into segments.

So, when we execute the algorithm will split the digram into 2 parts.



Any number of splits are possible. How and where these splits are conducted is determined by the algorithm.

This is actually called information entropy. It basically means when we perform a split it is increasing the information. And once it cannot add any more information to our setup by splitting our leaves then it stops.

(Every part obtained by splitting is known as a leaf.)

The algorithm is finding the optimal splits of our dataset and the final leaves are called terminal leaves.

Now, we have to predict the value of y for a new observation that gets added to our scattered plot.

So, let’s say w have a observation which has x1=30 and x2=50. By this we know the value lies somewhere between which terminal leaf. Now, we just have to take the averages of each terminal leaf.

So, we take the average of y for these points (for every terminal leaf)and that will be the value assigned to any pont that fall in that terminal leaf.

