



# Decision Tree – Regression

Diptangshu Banik

Twitter - @dipbanik



# Overview

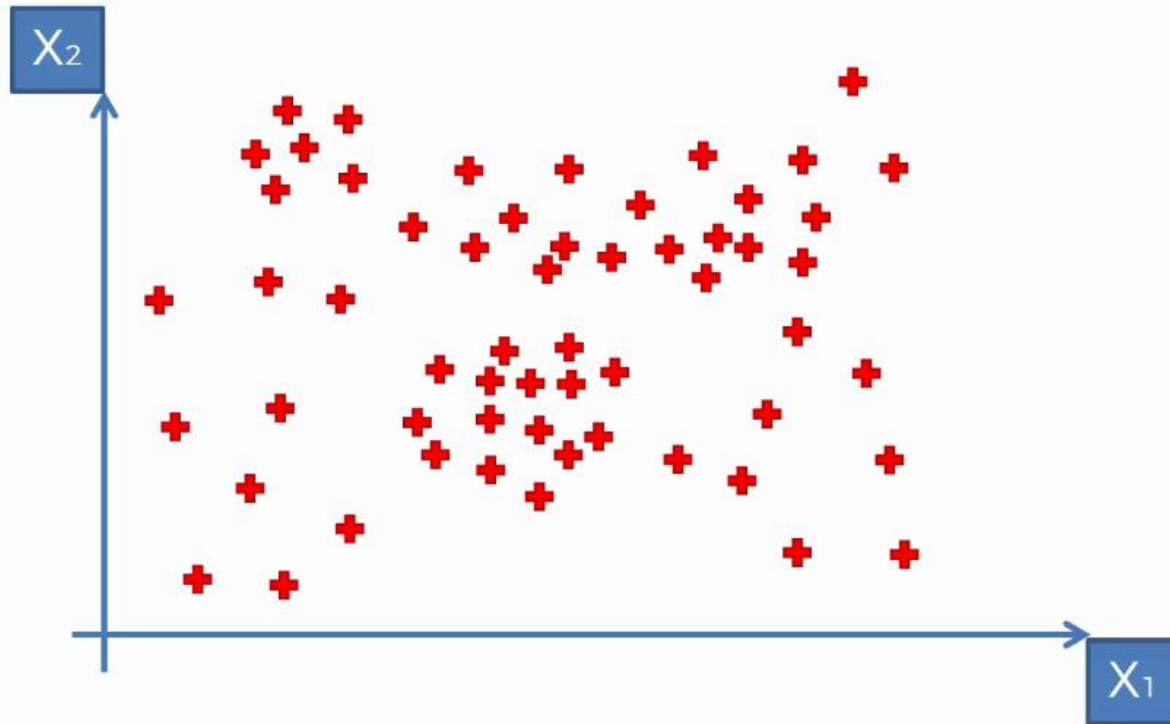
- Decision trees are of two types – Regression and classification
- It creates leaves or pockets and takes the mean of the data points in those pockets.
- If an input value falls in a leaf, the prediction for that input will be the mean of the data points in that leaf.
- How and where these splits are conducted are determined by the algorithm.



- Split occurs on the basis of information entropy (is the split increasing the amount of information that we have about our points).
- The split stops when we cannot add any more information about the leaf or the new leaf has less than say 5% of total points of the actual leaf.
- Instead of looking at all the points in our dataset, we are looking at specific cluster which localizes the prediction and can be more accurate.
- If the data is non linear, this is a good option.



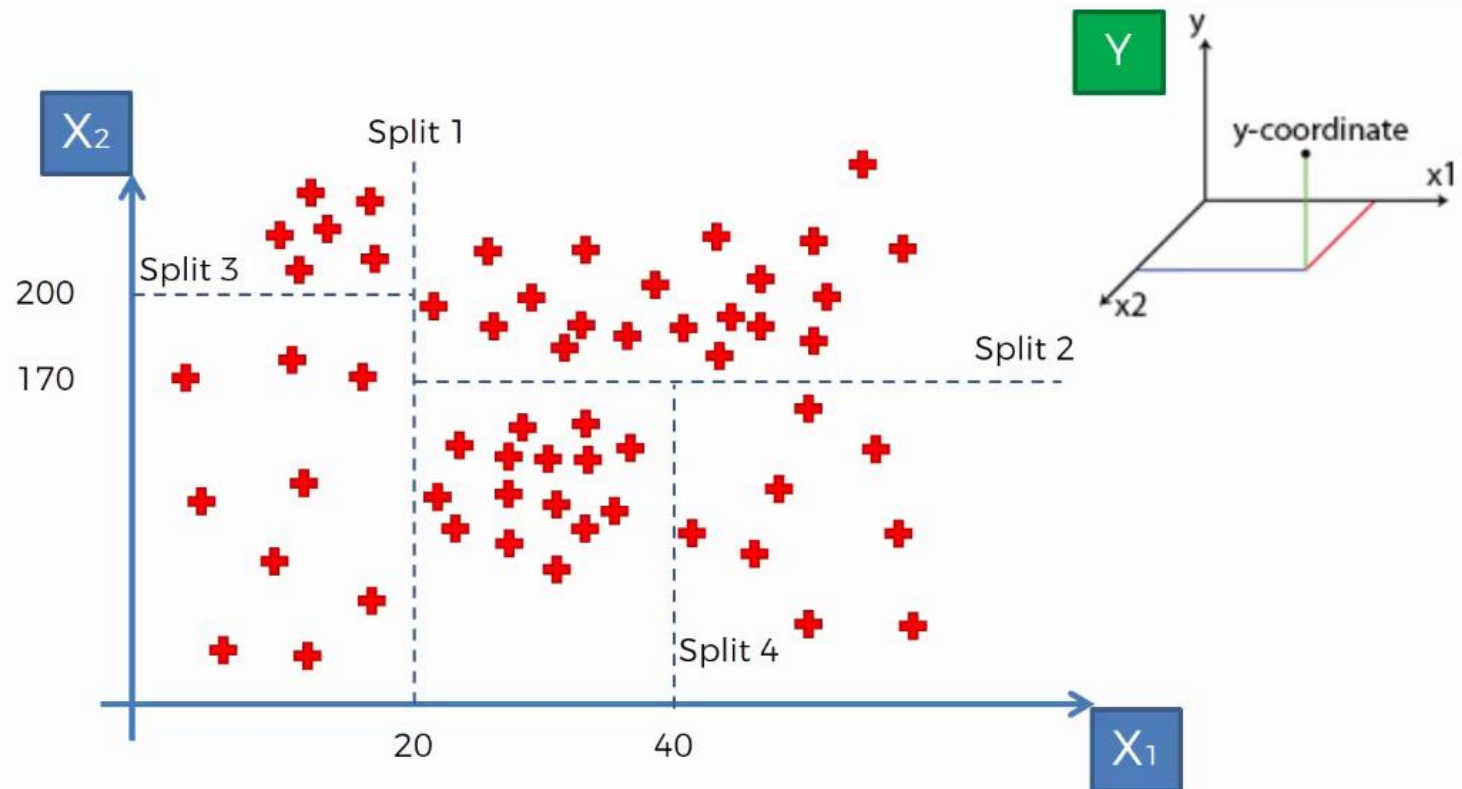
- Our data set -





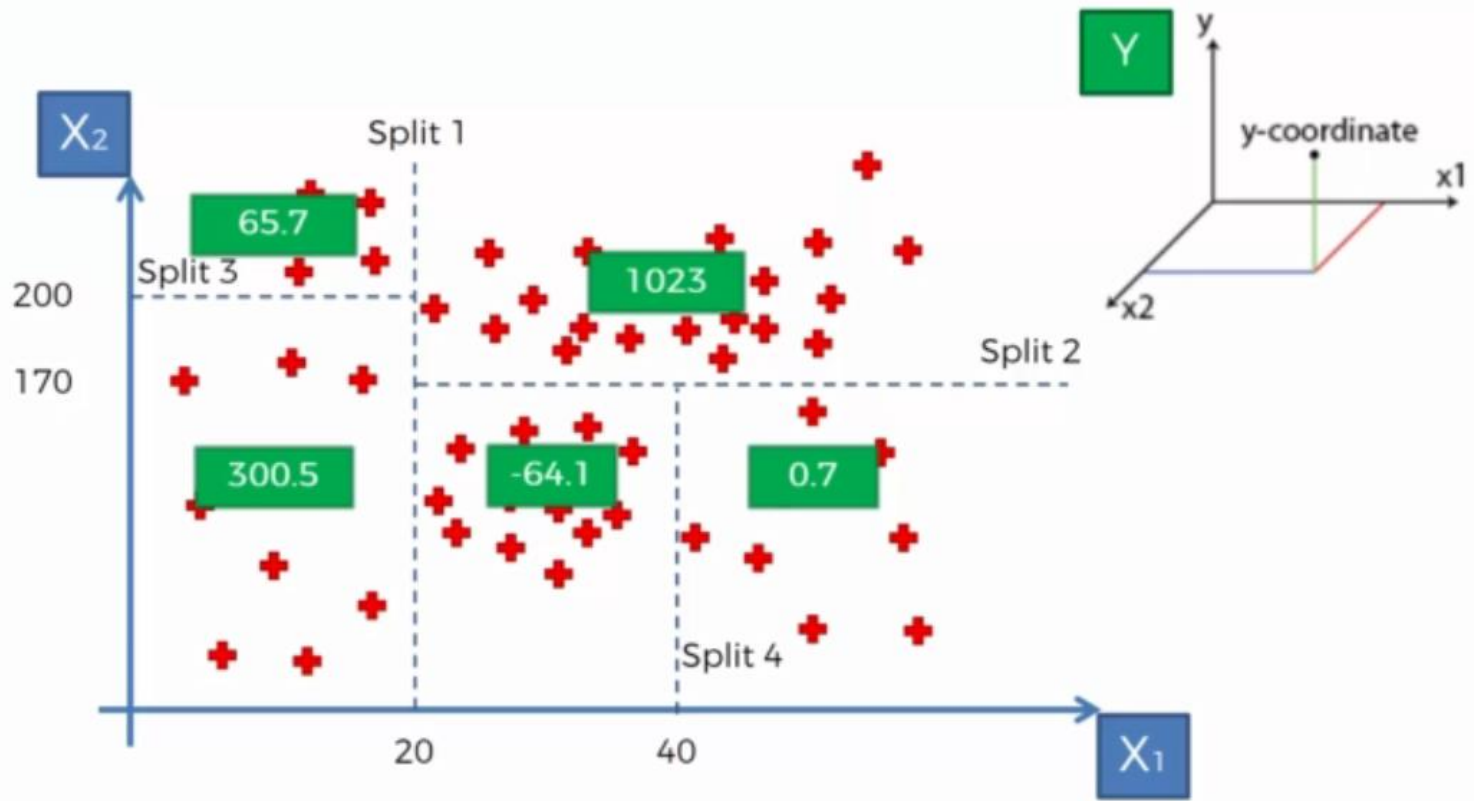
# Data representation DT

- Y is the output variable, kind of imaginary here as it is difficult to plot in a 2d setting





- Get the mean of every leaf and predict for (30,50) -





## Tree representation –

