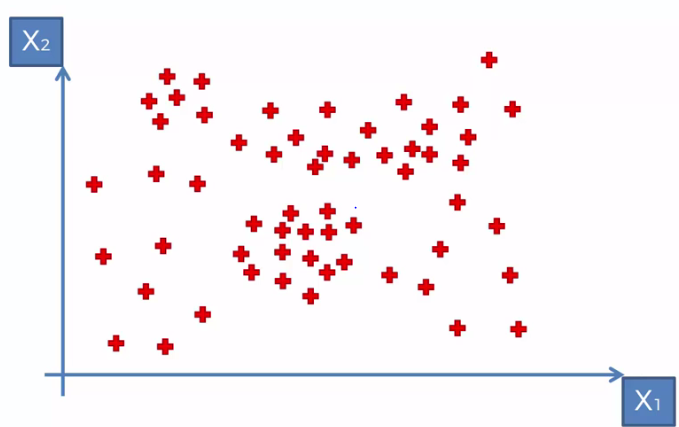
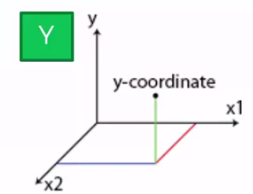
***Decision Tree Regression***

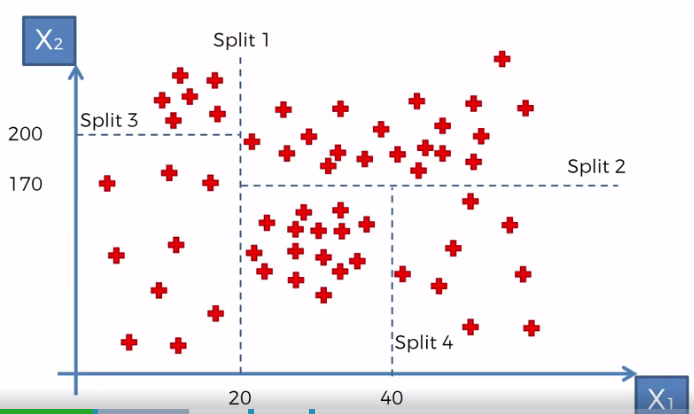
* **CART =** classification and regression trees 🡪 term that encapsulates 2 types of decision trees
* Regression are a bit more complicated than classification trees
* Have 2 predictors and 1 variable to predict:



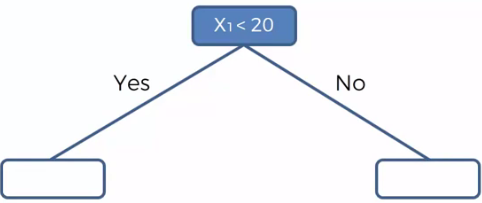
* Can’t see y yet, because we need to play around with our x1 and x2 values
* y will be a 3rd dimensions projection over all values of the x1 and x2 plane



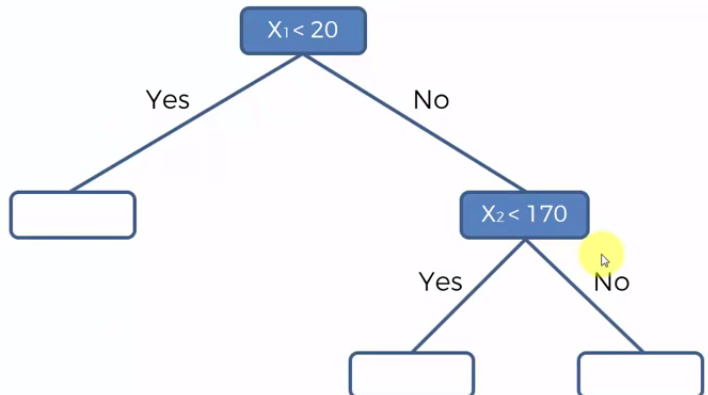
* Once we run the algorithm, the scatterplot will be split into segments, say when x1 = 20, then when x2 = 170, then when x2 = 200, and finally when x1 = 40



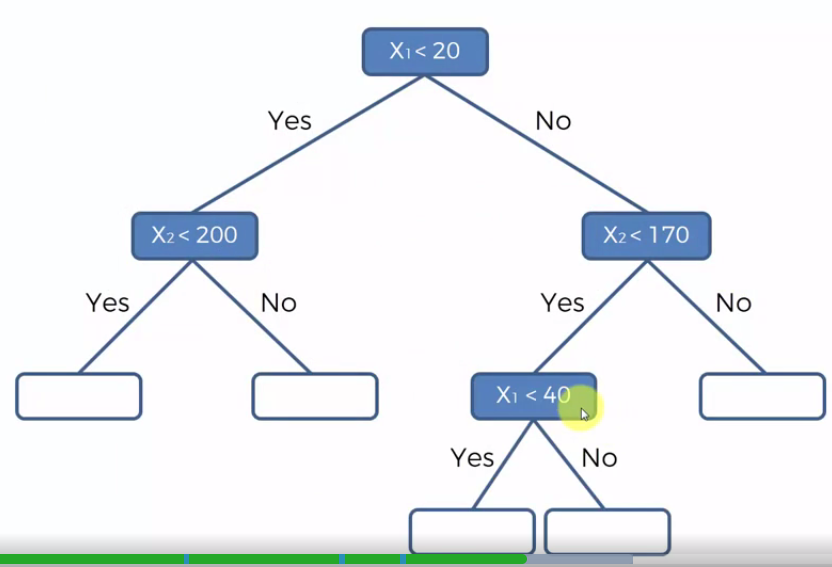
* How and when splits occur is decided upon by the algorithm by looking at **information entropy**
* In simple words, when we perform a split, do we have more info about our points?
* Is the split adding value to how we want to group our points?
* Algorithm stops at an optimal # of splits due to a threshold of a minimum amount of info to be added
* Once we cannot add any more info above this threshold by splitting the **leaves** (adding more splits), it stops
* Or it can stop when we have less than a certain amount of points (say 5%) in a leaf
* The final leaves = **terminal leaves** (the groups)
* The first split is at x1 = 20



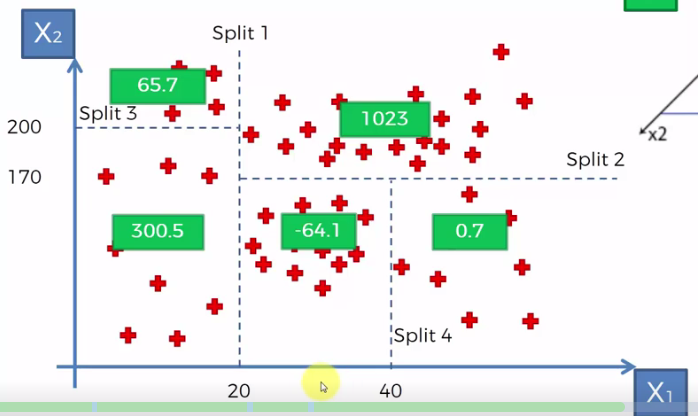
* Split 2 happens at x1 = 70, *and only happens for points w/ x1 > 20*

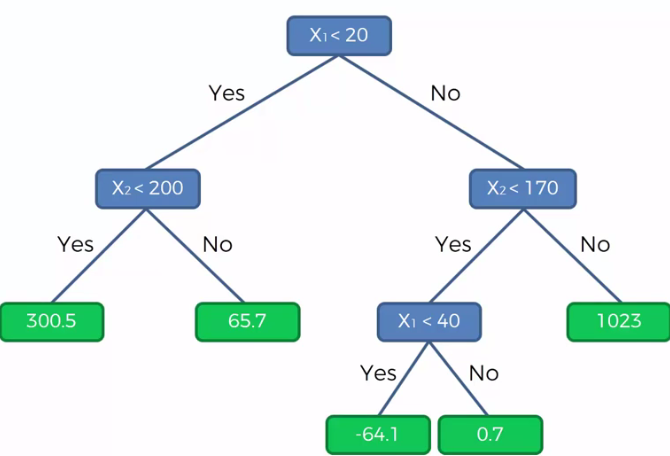


* Then split 3 happens when x1 < 20, and asks if x2 < 200
* Finally the 4th split occurs when x1 > 20, x2 < 170, and says “yes” if x < 40 (i.e. between 20-40)



* Now we need to figure out what to predict for y for a new observation based on these leaves
* By adding leaves, we added *info into our system*, and the system creates these terminal leave groups that help us predict values of y
* *Take the averages of y w/in each of the terminal leaves and that will be assigned to any new observation added to the terminal leaf*





* Now if a new point comes in w/ x1 = 30 and x = 50, we’d predict y’s value to be -64.1
* We want to add as much info as possible so that we have accurate leaves to give good average value to more accurately predict values of y for new observations