#100DAYSOFMLCODE
DAY 23
©AVIK JAIN

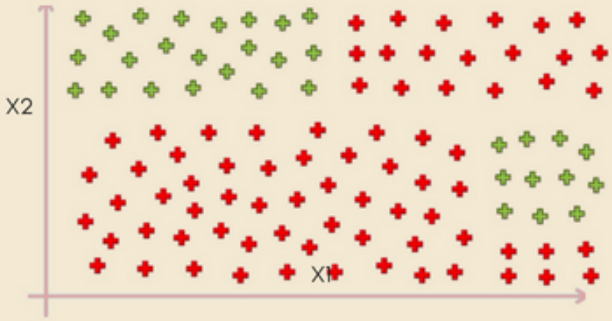
DECISION TREE

AN INTUITION ON DECISION TREE FOR CLASSIFICATION

1 WHAT IS A DECISION TREE?

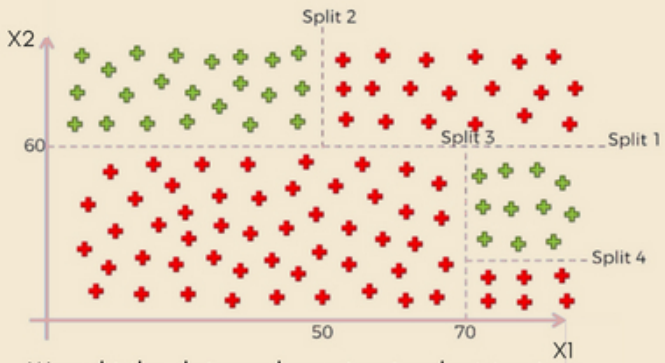
It is a type of supervised learning algorithm that is mostly used in classification problems and works for both categorical and continuous input and output variables.

A decision tree is a tree in which each branch node represents a choice between a number of alternatives and each leaf node represents a decision.



Here we've got an example with lots of points on our two dimensional scatter plot.
Now how does a decision tree work.

So what it is going to do is cut it up into slices in several iterations.



We split the data and construct a decision tree side by side which we will use later. This very task

The resulting Tree (obtained by applying algorithms like CART, ID3) which will be later used

is achieved by using various algorithms. It builds a decision tree from a fixed set of examples and the resulting tree is used to classify future samples.

algorithms like CART, ID3) which will be later used to predict the outcomes

3 DECISION TREE ALGORITHM: ID3

ID3 stands for Iterative Dichotomizer 3. The basic idea is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node.
Sounds simple — but which node should we select to build the correct and most precise decision tree? How would we decide that? Well, we have some measures that can help us in selecting the best choice!

Loop:
A -> Best Attribute
Assign A as decision attribute for node.
For each value of A, create a descendant of node.
Sort training examples to leaves.
If examples perfectly classified:
STOP
Else:
Iterate over leaves

INFORMATION GAIN

The best attribute is the one which gives us maximum Information Gain. Broadly speaking, it is a mathematical way to capture the amount of information we want by picking a particular attribute. But what it really speaks about us the reduction in the randomness, over the tables that we have with the set of data, based upon knowing the value of a particular attribute. Information gain is defined by:

$$Gain(S, A) = Entropy(S) - \sum_{v} \frac{|S_v|}{|S|} Entropy(S_v)$$

S = Collection of training examples
A = Particular attribute
|Sv| = Number of elements in Sv
|S| = Number of elements in S
v = All the possible values of the attribute

ENTROPY

Entropy in machine learning also carries almost the same meaning as it does in Thermodynamics. It is a measure of randomness.

$$Entropy = -\sum_{v} p(v) \log_2 p(v)$$

Where v = possible values for the attribute.

Steps :
1.compute the entropy for data-set
2.for every attribute/feature:
1. calculate entropy for all categorical values
2. take average information entropy for the current attribute
3. calculate gain for the current attribute
3. pick the highest gain attribute.
4. Repeat until we get the tree we desired

Check out the Repository at: github.com/Avik-Jain/100-Days-Of-ML-Code
Follow Me For More Updates