Decision Trees

What is?

Decision Trees are built by spitting the training set into nodes where one node contains all or most of one category of data.

Each internal node corresponds to a test (ie. A choice)

Each branch corresponds to a result of the test

Each leaf node assigns a subject to a classification

How to construct a decision tree

1. Choose an attribute
2. Calculate the significance of the attribute in splitting of data
3. Split the data based on the value of the best attribute
4. Go to step 1 and repeat

Which attribute is best?

Datasets are built using recursive partitioning to classify the data.

Basically, the best attribute is statistically the best predictor of the decision to be made (more significant) (Less impurity) (Lower Entropy)

Entropy is a measure of randomness. The lower the entropy the less uniform the distribution (more pure).

Entropy can be calculated using the formula

Ent = p(A)log2(p(A)) – p(B)log2(p(B))

Where:

P = the ratio of a category

So basically

If we have 15 total observations, 9 are Yes, 6 are no then:

Ent = (9/15)log2(9/15) – (6/15)log2(9/15)

If an attribute still isn’t completely predictive, we add more nodes until the ‘leaves’ are as ‘pure’ as possible.

We do this by testing entropy at each level for information gain.

Information gain is the increase in level of certainty after splitting ie.

Information gain = (Entropy before split) – (weighted entropy after split)

So, say we have a set of 14 split 9/5 entropy is .94 before split and after the split we have node entropies of both totaling 7 results (3,4) and (6,1) .985 and .592

Then the weight is the split of results over the total (7/14) for both meaning that:

Information gain = .94 – [(7/14)\*0.985 + (7/14)\*0.592]

Information gain = 0.151