Sample data：



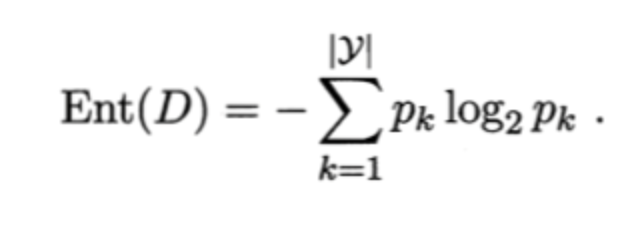
1. Observe the data and analyses it.

The example shows that the result “Conversion” could be predicted by the three independent variables: Demand, Campaign and Strategic. So, we know we could build a decision tree model for predictions.

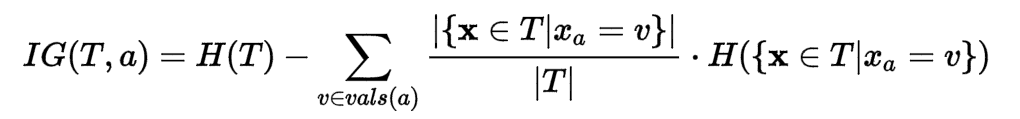
The algorithm is that we want to get the best predictability performance of this model. After we split from every node, we want the higher purity. In this example, I’d like to use the ID3 algorithm to make its decision tree.

1. Know about the concept of Entropy and Information Gain

Entropy is a way to determine the uncertainty of a subset.



1. Information gain can be used to define a preferred sequence of attributes to investigate to most rapidly narrow down the state of X.



1. Now we could know about how to split the leaf in the decision tree. We should choose a variable to be the leaf which create the most information gain.

Now we can change the sequence of the sample to look clearer:



The ratio of high conversion: 5/16

The ratio of low conversion: 6/16 = 3/8

The ratio of medium conversion: 5/16

So, the entropy of the root is:

Ent(S) = = -( =

-

-(-0.5394-0.5306-0.5394) = 1.5794

1. Alternative Splits from Root Node:

Now we need to calculate the information gain of every attribute :

5.1 There are three

When the Demand = low



Ent (Demand = low) =- ( 0.6\*-0.737+0.4\*-1.32 = -0.529 -0.4422 = 0.97

Similarly,



Ent (Demand = moderate) = =

-(0.5\*-1 +0.25\*-2\*2) = 1.5



Ent (Demand = heavy) = =

-(-0.524-0.516-0.516) = 1.55

Gain (S, Demand) = Ent(S) -

= 1.5794– (

= 1.5794 – 1.3562 = 0.2232

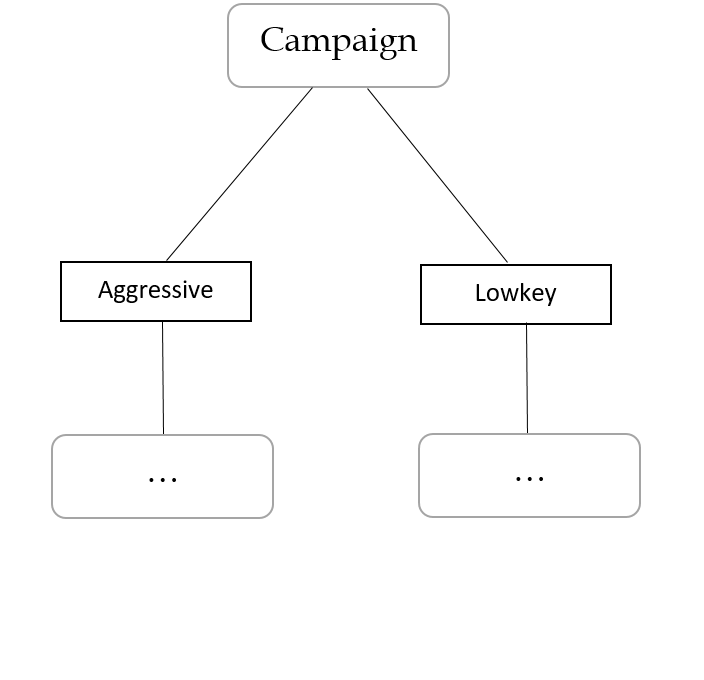
As the same with the attribute ‘demand’, the Gain of other two attributes can be calculated like this.

So, we get:

Gain (S, Strategic) = Ent(S) - = 0.006

Gain (S, Campaign) = Ent(S) - = 0.293

Because the gain of attribute Campaign is the largest one, so the first brunch should be Campaign. So, the first step will be:



* 1. And then we need to calculate the new information gain and choose the best two. 

We could see that at the ‘Aggressive’ part:

Ent(C) = = -( =1.3921

Ent (Demand = low’) = 0

Ent (Demand = heavy’) = -( 0.2724

Ent (Demand = moderate’) = -( 1

Gain (C, Demand) = Ent(C) -

= 1.3921- (

= 1.1580

Ent (Strategic = yes’) = -( 1.4591

Ent (Strategic = no’) = = -( 1

Gain (C, Strategic) = Ent(C) -

= 1.3921- (

= 1.2303

So, Gain (C, Demand) > Gain (C, Strategic):

Campaign

lowkey

aggressive

demand

“Lowkey” part:

Ent(C) = = -( =1.1488

Ent (Demand = low’) = -( 0.8113

Ent (Demand = heavy’) = 0

Ent (Demand = moderate’) = -( 1

Gain (C, Demand) = Ent(C) -

= 1.1488- (

= 0.3995

Ent (Strategic = yes’) = 0

Ent (Strategic = no’) = = -( 0.2265

Gain (C, Strategic) = Ent(C) -

= 1.1488- (

= 0.9870

Campaign

lowkey

aggressive

strategic

demand