ID3 COMPLETE EXAMPLE

Sky	Temp	Humid	Wind	Water	Forecast	EnjySp
Sunny	Warm	High	Strong	Cool	Change	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No

a) Show the decision tree that would be learnt by ID3 assuming it is given the four examples in Table 1 for *EnjoySport*

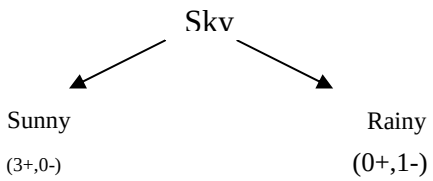
In order to make the decision tree first I have to calculate the gain of information that will give me each of the 6 'features'.
The entropy of the system is:

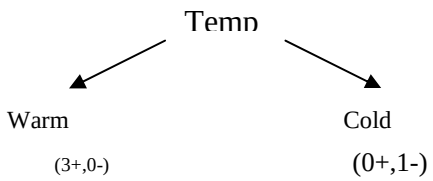
- Entropy(S) = $-p_+ \log 2p_+ - p_- \log 2p_- = -3/4*\log2(3/4) – 1/4*\log2(1/4) = 0.8113$.

The gain results of each feature are shown in the next table:

Feature	Value 1 Positive Examples	Value 1 Negative Examples	Value 2 Positive Examples	Value 2 Negative Examples	Gain Value
Sky	3	0	0	1	0.8113
Temp	3	0	0	1	0.8113
Humid	1	0	2	1	0.1226
Wind	3	1	0	0	0
Water	2	1	1	0	0.1226
Forecast	2	0	1	1	0.3113

The results show that we gain more information if our first node is Sky or Temp. Moreover, if we take one of them as our first node our training examples will be all explained.





b) Add the new training example, and compute the new decision tree. This time show the value of the information gain for each candidate attribute at each step of growing the tree.

So our new training set will be the next one:

Sky	Temp	Humid	Wind	Water	Forecast	EnjySp
Sunny	Warm	High	Strong	Cool	Change	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Warm	Same	Yes
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	Normal	Weak	Warm	Same	No

First node decision:

1. Calculate the entropy for the whole set of examples

   Entropy(S) = $-p_+\log2p_+ - p_-\log2p_- = -3/5\log2(3/5) - 2/5\log2(2/5) = 0.9710$.

2. Calculate the gain for each attribute
   a. Sky (Sunny 3+ 1-, Rainy 1-)
   Gain = 0.9710 – 4/5*(-3/4*log2(3/4) – 1/4*log2(1/4)) – 1/5*(-1/1*log2(1/1)) = 0.9710 – 4/5*0.8113 – 1/5*0 = 0.3229.

   b. Temp (Warm 3+ 1-, Cold 1-)
   Gain = 0.9710 – 4/5*(-3/4*log2(3/4) – 1/4*log2(1/4)) – 1/5*(-1/1*log2(1/1)) = 0.9710 – 4/5*0.8113 – 1/5*0 = 0.3229.

   c. Humid (Normal 1+ 1-, High 2+ 1-)
   Gain = 0.9710 – 2/5*(-1/2*log2(1/2) – 1/2*log2(1/2)) – 3/5*(-2/3*log2(2/3) - 1/3*log2(1/3)) = 0.9710 – 2/5*1 – 3/5*0.68875 = 0.0200.

   d. Wind (Strong 3+ 1-, Weak 1-)
   Gain = 0.9710 – 4/5*(-3/4*log2(3/4) – 1/4*log2(1/4)) – 1/5*(-1/1*log2(1/1)) = 0.9710 – 4/5*0.8113 – 1/5*0 = 0.3229.

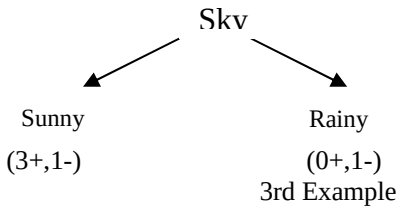
   e. Water (Warm 2+ 2-, Cool 1+)
   Gain = 0.9710 – 4/5*(-2/4*log2(2/4) – 2/4*log2(2/4)) – 1/5*(-1/1*log2(1/1)) = 0.9710 – 4/5*1 = 0.1710.

   f. Forecast (Same 2+ 1-, Change 1+ 1-)
   Gain = 0.9710 – 3/5*(-2/3*log2(2/3) – 1/3*log2(1/3)) – 2/5*(-1/2*log2(1/2) – 1/2*log2(1/2)) = 0.9710 – 3/5*0.9183 – 2/5*1 = 0.0200.

3. Decision of which attribute will be used as the first node.

Attribute	Gain Value
Sky	0.3229
Temp	0.3229
Humid	0.0200
Wind	0.3229
Water	0.1710
Forecast	0.0200

The attributes that gives us more information are Sky, Temp and Wind. I've followed the exercise using Sky as my first node. So the tree until this step is as it follows:



As it can be seen one branch is already closed so our tree will only be extended from Sunny branch. Because of that the training set that we have to consider has changed from the initial one, now we only have to consider those examples with Sunny as value for attribute Sky.

So our new training set will be the next one:

Sky	Temp	Humid	Wind	Water	Forecast	EnjySp
Sunny	Warm	High	Strong	Cool	Change	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	Normal	Weak	Warm	Same	No

Second node decision:

4. Calculate the entropy for the new set of examples

   Entropy(S) = $-p_+\log2p_+ - p_-\log2p_- = -3/4\log2(3/4) - 1/4\log2(1/4) = 0.8113$.

5. Calculate the gain for each attribute

   a. Temp (Warm 3+ 1-, Cold 0+ 0-)
   Gain = 0.8113 – 4/4*(-3/4*log2(3/4) – 1/4*log2(1/4)) = 0.8113 – 0.8113 = 0.

   b. Humid (Normal 1+ 1-, High 2+)
   Gain = 0.8113 – 2/4*(-1/2*log2(1/2) – 1/2*log2(1/2)) – 2/4*(-2/2*log2(2/2)) = 0.8113 – 0.5*1 – 0.5*0 = 0.3113.

   c. Wind (Strong 3+, Weak 1-)
   Gain = 0.9710 – 3/4*(-3/3*log2(3/3)) – 1/4*(-1/1*log2(1/1)) = 0.8113 – 0.75*0 – 0.25*0 = 0.8113.

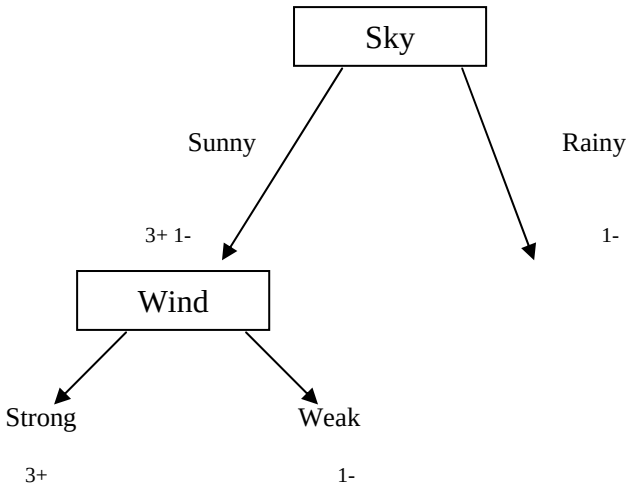
   d. Water (Warm 2+ 1-, Cool 1+)
   Gain = 0.8113 – 3/4*(-2/3*log2(2/3) – 1/3*log2(1/3)) – 1/4*(-1/1*log2(1/1)) = 0.8113 – 0.75*0.918 – 0.25*0 = 0.1226.

   e. Forecast (Same 2+ 1-, Change 1+)
   Gain = 0.8113 – 3/4*(-2/3*log2(2/3) – 1/3*log2(1/3)) – 1/4*(-1/1*log2(1/1)) = 0.8113 – 0.75*0.9183 – 1/4*0 = 0.1226.

6. Decision of which attribute will be used as the second node.

Attribute	Gain Value
Temp	0
Humid	0.3113
Wind	0.8113
Water	0.1226
Forecast	0.1226

The attribute that gives more information is Wind. If we choose Wind as our second node, the result tree is the next one:



As we can see the inferred tree is consistent with the hypotheses given in the exercise and we can stop the growth of the tree because all the branches are closed.