Assignment 1

Adriaan Louw (53031377)

April 21, 2018

Contents

1 Question 1 **1**

2 Question 2 **1**

3 Question 3 **2**

4 Question 4 **2**

5 Question 5 **2**
5.1 Question 5.1 2

6 Question 6 **3**
6.1 Question 6a 3
6.2 Question 6b 9

1 Question 1

The book *Introduction to Machine Learning* by Nils J, Nilsson can be found at http://robotics.stanford.edu/people/nilsson/MLBOOK.pdf and is 1.855 MB.

The book *A first encounter with Machine Learning* by Max Welling can be found at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.441.6238&rep=rep1&type=pdf and is 416 KB.

2 Question 2

We can represent some learning functions in machine learning as Boolean function. We can define a Boolean function as a function of the form

$$f(x_1, x_2, x_3, ...x_n) \tag{1}$$

Which maps a n-tuple of (0,1) values to 0,1. 0,1 can also be expressed as false,true.

There are 3 basic types of operations can be performed in Boolean functions. Firstly we have the "and" operation which uses the connective "." as in

$$x_1.z_2 \tag{2}$$

Which only returns true if x_1 and x_2 is true. Then there is the "or" operation represented as a "+"

$$x_1 + x_2 \tag{3}$$

Where it returns true if x_1 is true or if x_2 or both are true. Thirdly we have the negation operation indicated by a $^-$ as in

$$\bar{x_1} \tag{4}$$

This operation returns true if x_1 is false and false if x_1 is true.

The . and + operations are commutative

$$\begin{aligned} x_1.x_2 &= x_2.x_1 \\ x_1 + x_2 &= x_2 + x_1 \end{aligned} \tag{5}$$

and associative

$$\begin{aligned} x_1.x_2(x_3) &= x_1(x_2.x_3) \\ x_1 + x_2 + (x_3) &= x_1 + (x_2 + x_3) \end{aligned} \tag{6}$$

To commute between . and + we use DeMorgan's laws

$$\begin{aligned} \overline{x_1.x_2} &= \overline{x_1} + \overline{x_2} \\ \overline{x_1 + x_2} &= \overline{x_1}.\overline{x_2} \end{aligned} \tag{7}$$

Boolean functions can be broken down into various sub-classes. The first subclass is called *terms*. We can write these as $k_1k_2k_3...k_n$ where k_i are literals. An example would be the following term of size 4, $x_2.x3.\overline{x_6}.x_{12}$. This is also called a conjunctive literal as in all the terms are separated by the and operation.

Secondly we have *clauses* or . A clause is a function where the literals are separated by the or function. As in $k_1 + k_2 + ... + k_n$. An example would be $x_1 + \overline{x_5} + x_8$.

Disjunctive normal functions are functions that can be written as a disjunction of terms.

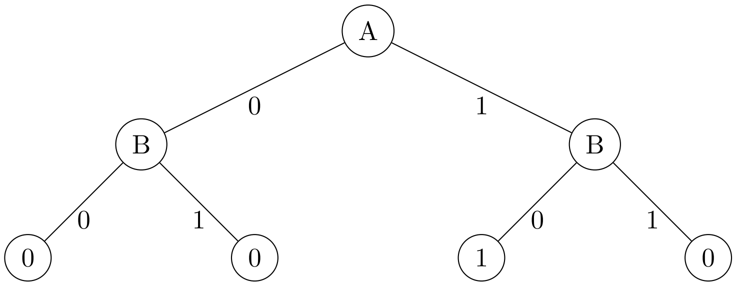
3 Question 3

4 Question 4

5 Question 5

5.1 Question 5.1

$A \wedge B$



6 Question 6

6.1 Question 6a

Entropy is used to calculate a decision tree in the ID3 algorithm. For a collection of Samples (S):

$$Entropy(S) = \sum_{i=1}^{c} -p_i log_2 p_i \tag{8}$$

If we assume we only have 2 outcomes, then we let the number of positive outcomes by p and the number of negative outcomes be n. Then we can write Equation 8 as

$$Entropy(p, n) = -\frac{p}{p+n} log_2 \frac{p}{p+n} - \frac{n}{p+n} log_2 \frac{n}{p+n} \tag{9}$$

More specifically we compare the gain of different attributes of the sample set. Then create the tree based on the attributes with the greatest gain

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{10}$$

We need to determine $Entropy(S)$

$$\begin{aligned} Entropy(p, n) &= -\frac{p}{p+n} log_2 \frac{p}{p+n} - \frac{n}{p+n} log_2 \frac{n}{p+n} \\ &= -\frac{3}{3+1} log_2 \frac{3}{3+1} - \frac{1}{3+1} log_2 \frac{1}{3+1} \\ &= 0.811 \end{aligned} \tag{11}$$

Now we need to determine the gain for the 6 attributes to determine which one will be the root node.

For Sky we have sunny $p_1 = 3, n_1 = 0)$ and rainy $p_2 = 0, n_2 = 1)$

$$\begin{aligned}
Gain(S, Sky) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{3}{4} Entropy(S_{Sunny}) - \frac{1}{4} Entropy(S_{Rainy}) \\
&= 0.811 - \frac{3}{4} Entropy(3, 0) - \frac{1}{4} Entropy(0, 1) \\
&= 0.811 - \frac{3}{4}(-\frac{3}{3} log_2 \frac{3}{3} - \frac{0}{3} log_2 \frac{0}{3}) - \frac{1}{4}(-\frac{0}{1} log_2 \frac{0}{1} - \frac{1}{1} log_2 \frac{1}{1}) \\
&= 0.811 - 0 \\
&= 0.811
\end{aligned} \tag{12}$$

For attribute AirTemp we have warm $p_1 = 3, n_1 = 0)$ and cold $p_2 = 0, n_2 = 1)$

$$\begin{aligned}
Gain(S, AirTemp) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{3}{4} Entropy(S_{Warm}) - \frac{1}{4} Entropy(S_{Cold}) \\
&= 0.811 - \frac{3}{4} Entropy(3, 0) - \frac{1}{4} Entropy(0, 1) \\
&= 0.811 - \frac{3}{4}(-\frac{3}{3} log_2 \frac{3}{3} - \frac{0}{3} log_2 \frac{0}{3}) - \frac{1}{4}(-\frac{0}{1} log_2 \frac{0}{1} - \frac{1}{1} log_2 \frac{1}{1}) \\
&= 0.811 - 0 \\
&= 0.811
\end{aligned} \tag{13}$$

For attribute Humidity we have normal $p_1 = 1, n_1 = 0)$ and high $p_1 = 2, n_2 = 1)$

$$\begin{aligned}
Gain(S, Humidity) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{1}{4} Entropy(S_{normal}) - \frac{3}{4} Entropy(S_{high}) \\
&= 0.811 - \frac{1}{4} Entropy(1, 0) - \frac{3}{4} Entropy(2, 1) \\
&= 0.811 - \frac{1}{4}(-\frac{1}{1} log_2 \frac{1}{1} - \frac{0}{1} log_2 \frac{0}{1}) - \frac{3}{4}(-\frac{2}{3} log_2 \frac{2}{3} - \frac{1}{3} log_2 \frac{1}{3}) \\
&= 0.811 - 0.689 \\
&= 0.123
\end{aligned} \tag{14}$$

For attribute Wind we only have strong $p_1 = 3, n_1 = 1$

$$\begin{aligned}
Gain(S, Wind) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{4}{4} Entropy(S_{strong}) \\
&= 0.811 - \frac{4}{4} Entropy(3, 1) \\
&= 0.811 - \frac{4}{4}(-\frac{4}{4} log_2 \frac{4}{4} - \frac{0}{4} log_2 \frac{0}{4}) \\
&= 0.811 - 0.811 \\
&= 0.0
\end{aligned} \tag{15}$$

For attribute water we have warm $p_1 = 2, n_1 = 1)$ and cold $p_1 = 1, n_2 = 0)$

$$\begin{aligned}
Gain(S, Water) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{3}{4} Entropy(S_{warm}) - \frac{1}{4} Entropy(S_{high}) \\
&= 0.811 - \frac{3}{4} Entropy(2, 1) - \frac{1}{4} Entropy(1, 0) \\
&= 0.811 - \frac{3}{4}(-\frac{2}{3} log_2 \frac{2}{3} - \frac{1}{3} log_2 \frac{1}{3}) - \frac{1}{4}(-\frac{1}{1} log_2 \frac{1}{1} - \frac{0}{1} log_2 \frac{0}{1}) \\
&= 0.811 - 0.689 \\
&= 0.123
\end{aligned} \tag{16}$$

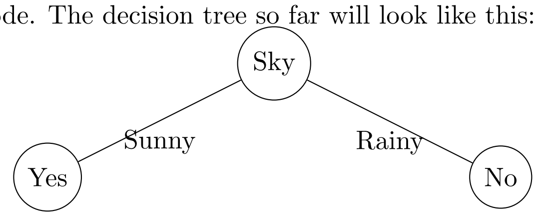
For attribute forecast we have same $p_1 = 2, n_1 = 0)$ and change $p_1 = 1, n_2 = 1)$

$$\begin{aligned}
Gain(S, Forecast) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{2}{4} Entropy(S_{same}) - \frac{2}{4} Entropy(S_{change}) \\
&= 0.811 - \frac{2}{4} Entropy(2, 0) - \frac{1}{1} Entropy(1, 0) \\
&= 0.811 - \frac{2}{4}(-\frac{2}{2} log_2 \frac{2}{2} - \frac{0}{2} log_2 \frac{0}{2}) - \frac{2}{4}(-\frac{1}{2} log_2 \frac{1}{2} - \frac{1}{2} log_2 \frac{1}{2}) \\
&= 0.811 - 0.5 \\
&= 0.311
\end{aligned} \tag{17}$$

To recap:
Gain(Sky) = 0.811 from Equation 12
Gain(AirTemp) = 0.811 from Equation 13
Gain(Humidity) 0.123 from Equation 14
Gain(Wind) = 0 from Equation 15

Gain(Water) = 0.123 from Equation 16
Gain(Forecast) = 0.311 from Equation 17

We find that Sky and Airtemp have the greatest information gain. We could choose either to be the root node. We arbitrarily will choose Sky as the root node. The decision tree so far will look like this:



For the leaf node Sky = Sunny, we need to determine the sub-tree.
For Sky = Sunny:

$$\begin{aligned} Entropy(S_{Sky=Sunny}) &= Entropy(p,n) \\ &= -\frac{p}{p+n}log_2\frac{p}{p+n} - \frac{n}{p+n}log_2\frac{n}{p+n} \\ &= -\frac{3}{3}log_2\frac{3}{3} - \frac{0}{3}log_2\frac{0}{3} \\ &= 0 \end{aligned} \tag{18}$$

We need to calculate the information gain for the other attributes when Sky = sunny
For AirTemp we only have warm ($p_1 = 3, n_1 = 0$)

$$\begin{aligned} Gain(S_{Sky=Sunny}, AirTemp) &= 0.0 - \frac{3}{3}Entropy(S_{Sky=Sunny;AirTemp=Warm}) \\ &= 0.0 - \frac{3}{3}Entropy(3,0) \\ &= 0.0 - 1(-\frac{3}{3}log_2\frac{3}{3} - \frac{0}{3}log_2\frac{0}{3}) \\ &= 0.0 \end{aligned} \tag{19}$$

For Humidity normal ($p_1 = 1, n_1 = 0$) and high ($p_2 = 2, n_2 = 0$)

$$\begin{aligned} Gain(S_{Sky=Sunny}, Humidity) &= 0.0 - \frac{1}{3}Entropy(S_{Sky=Sunny;Humidity=Normal}) \\ &- \frac{2}{3}Entropy(S_{Sky=Sunny;Humidity=High}) \\ &= 0.0 - \frac{1}{3}Entropy(1,0) - \frac{2}{3}Entropy(2,0) \\ &= 0.0 - \frac{1}{3}(-\frac{1}{1}log_2\frac{1}{1} - \frac{0}{1}log_2\frac{0}{1}) - \frac{2}{3}(-\frac{2}{2}log_2\frac{2}{2} - \frac{0}{2}log_2\frac{0}{2}) \\ &= 0.0 \end{aligned} \tag{20}$$

For Wind we only have strong ($p_1 = 3, n_1 = 0$)

$$
\begin{aligned}
Gain(S_{Sky=Sunny}, Wind) &= 0.0 - \frac{3}{3}Entropy(S_{Sky=Sunny;Wind=Strong}) \\
&= 0.0 - \frac{3}{3}Entropy(3,0) \\
&= 0.0 - 1(-\frac{3}{3}log_2\frac{3}{3} - \frac{0}{3}log_2\frac{0}{3}) \\
&= 0.0
\end{aligned} \tag{21}
$$

For Water we have warm($p_1 = 2, n_1 = 0$) and cool($p_2 = 1, n_2 = 0$)

$$
\begin{aligned}
Gain(S_{Sky=Sunny}, Water) &= 0.0 - \frac{2}{3}Entropy(S_{Sky=Sunny;Water=Warm}) \\
&- \frac{1}{3}Entropy(S_{Sky=Sunny;Water=Cool}) \\
&= 0.0 - \frac{2}{3}Entropy(2,0) - \frac{1}{3}Entropy(1,0) \\
&= 0.0 - \frac{2}{3}(-\frac{2}{2}log_2\frac{2}{2} - \frac{0}{2}log_2\frac{0}{2}) - \frac{1}{3}(-\frac{1}{1}log_2\frac{1}{1} - \frac{0}{1}log_2\frac{0}{1}) \\
&= 0.0
\end{aligned} \tag{22}
$$

For Forecast we have same($p_1 = 2, n_1 = 0$) and change($p_2 = 1, n_2 = 0$)

$$
\begin{aligned}
Gain(S_{Sky=Sunny}, Forecast) &= 0.0 - \frac{2}{3}Entropy(S_{Sky=Sunny;Forecast=Same}) \\
&- \frac{1}{3}Entropy(S_{Sky=Sunny;Forecast=Change}) \\
&= 0.0 - \frac{2}{3}Entropy(2,0) - \frac{1}{3}Entropy(1,0) \\
&= 0.0 - \frac{2}{3}(-\frac{2}{2}log_2\frac{2}{2} - \frac{0}{2}log_2\frac{0}{2}) - \frac{1}{3}(-\frac{1}{1}log_2\frac{1}{1} - \frac{0}{1}log_2\frac{0}{1}) \\
&= 0.0
\end{aligned} \tag{23}
$$

This means that the node Sky = Sunny in our tree will be a leaf node, sice at this node the gain gor all the other attributes are 0.

Next we need to calculate the subtree under the node Sky = Rainy

Starting with

$$
\begin{aligned}
Entropy(S_{Sky=Rainy}) &= Entropy(p, n) \\
&= -\frac{p}{p+n}log_2\frac{p}{p+n} - \frac{n}{p+n}log_2\frac{n}{p+n} \\
&= -\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1} \\
&= 0
\end{aligned} \tag{24}
$$

For AirTemp we only have cold ($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, AirTemp) &= 0.0 - \frac{1}{1}Entropy(S_{Sky=Rainy;AirTemp=Cold}) \\
&= 0.0 - \frac{1}{1}Entropy(0,1) \\
&= 0.0 - 1(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{25}$$

For Humidity we only have High ($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Humidity) &= 0.0 - \frac{1}{1}Entropy(S_{Sky=Rainy;Humidity=High}) \\
&= 0.0 - \frac{1}{1}Entropy(0,1) \\
&= 0.0 - 1(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{26}$$

For Wind we only have Strong ($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Wind) &= 0.0 - \frac{1}{1}Entropy(S_{Sky=Rainy;Wind=Strong}) \\
&= 0.0 - \frac{1}{1}Entropy(0,1) \\
&= 0.0 - 1(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{27}$$

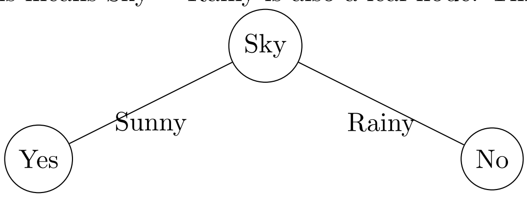
For Water we only have Warm ($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Water) &= 0.0 - \frac{1}{1}Entropy(S_{Sky=Rainy;Water=Warm}) \\
&= 0.0 - \frac{1}{1}Entropy(0,1) \\
&= 0.0 - 1(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{28}$$

For Forecast we only have Change ($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Forecast) &= 0.0 - \frac{1}{1}Entropy(S_{Sky=Rainy;ForeCast=Change}) \\
&= 0.0 - \frac{1}{1}Entropy(0,1) \\
&= 0.0 - 1(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{29}$$

This means when Sky = Rainy, the gain from all the other attributes are 0. This means Sky = Rainy is also a leaf node. This means the final tree will be:



6.2 Question 6b

Now we add a new row of training data. We will be recalculating the decision tree. Starting by determining the root node of the tree by comparing the information gain each attribute gives us.

First we need to determine the entropy of S

$$\begin{aligned} Entropy(p,n) &= -\frac{p}{p+n}log_2\frac{p}{p+n} - \frac{n}{p+n}log_2\frac{n}{p+n} \\ &= -\frac{3}{3+2}log_2\frac{3}{3+2} - \frac{2}{3+2}log_2\frac{2}{3+2} \\ &= 0.971 \end{aligned} \tag{30}$$

For the attribute Sky we have Sunny ($p_1 = 3, n_1 = 1$) and Rainy ($p_2 = 0, n_2 = 1$)

$$\begin{aligned} Gain(S,Sky) &= 0.971 - \frac{4}{5}Entropy(S_{Sunny}) - \frac{1}{5}Entropy(S_{Rainy}) \\ &= 0.971 - \frac{4}{5}Entropy(3,1) - \frac{1}{5}Entropy(0,1) \\ &= 0.971 - \frac{4}{5}(-\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4}) - \frac{1}{5}(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\ &= 0.971 - 0.649 \\ &= 0.322 \end{aligned} \tag{31}$$

For the attribute AirTemp we have Warm ($p_1 = 3, n_1 = 1$) and Cold ($p_2 = 0, n_2 = 1$)

$$\begin{aligned} Gain(S,AirTemp) &= 0.971 - \frac{4}{5}Entropy(S_{Warm}) - \frac{1}{5}Entropy(S_{Cold}) \\ &= 0.971 - \frac{4}{5}Entropy(3,1) - \frac{1}{5}Entropy(0,1) \\ &= 0.971 - \frac{4}{5}(-\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4}) - \frac{1}{5}(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\ &= 0.971 - 0.649 \\ &= 0.322 \end{aligned} \tag{32}$$

For the attribute Humidity we have Normal ($p_1 = 1, n_1 = 1$) and High ($p_2 = 2, n_2 = 1$)

$$\begin{aligned}
Gain(S, Humidity) &= 0.971 - \frac{2}{5}Entropy(S_{Normal}) - \frac{3}{5}Entropy(S_{High}) \\
&= 0.971 - \frac{2}{5}Entropy(1,1) - \frac{3}{5}Entropy(2,1) \\
&= 0.971 - \frac{2}{5}(-\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4}) - \frac{3}{5}(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\
&= 0.971 - 0.951 \\
&= 0.020
\end{aligned} \tag{33}$$

For the attribute Wind we have Strong($p_1 = 3, n_1 = 1$) and Weak ($p_2 = 0, n_2 = 1$)

$$\begin{aligned}
Gain(S, Wind) &= 0.971 - \frac{4}{5}Entropy(S_{Strong}) - \frac{1}{5}Entropy(S_{Weak}) \\
&= 0.971 - \frac{4}{5}Entropy(3,1) - \frac{1}{5}Entropy(0,1) \\
&= 0.971 - \frac{4}{5}(-\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4}) - \frac{1}{5}(-\frac{0}{1}log_2\frac{0}{1} - \frac{1}{1}log_2\frac{1}{1}) \\
&= 0.971 - 0.649 \\
&= 0.322
\end{aligned} \tag{34}$$

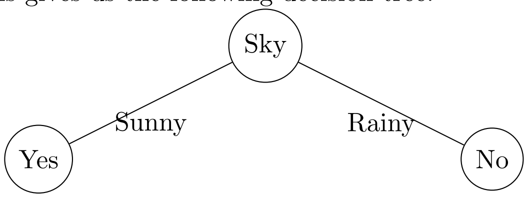
For the attribute Water we have Warm ($p_1 = 3, n_1 = 1$) and Cool ($p_2 = 1, n_2 = 0$)

$$\begin{aligned}
Gain(S, Water) &= 0.971 - \frac{4}{5}Entropy(S_{Warm}) - \frac{1}{5}Entropy(S_{Cool}) \\
&= 0.971 - \frac{4}{5}Entropy(3,1) - \frac{1}{5}Entropy(0,1) \\
&= 0.971 - \frac{4}{5}(-\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4}) - \frac{1}{5}(-\frac{1}{1}log_2\frac{1}{1} - \frac{0}{1}log_2\frac{0}{1}) \\
&= 0.971 - 0.649 \\
&= 0.322
\end{aligned} \tag{35}$$

For the attribute Forecast we have Same ($p_1 = 2, n_1 = 1$) and Change ($p_2 = 1, n_2 = 1$)

$$
\begin{aligned}
Gain(S, Forecast) &= 0.971 - \frac{3}{5}Entropy(S_{Same}) - \frac{2}{5}Entropy(S_{Change}) \\
&= 0.971 - \frac{3}{5}Entropy(2,1) - \frac{2}{5}Entropy(1,1) \\
&= 0.971 - \frac{3}{5}(-\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3}) - \frac{2}{5}(-\frac{1}{2}log_2\frac{1}{2} - \frac{1}{2}log_2\frac{1}{2}) \\
&= 0.971 - 0.951 \\
&= 0.020
\end{aligned}
\tag{36}
$$

We find that the attributes Sky,AirTemp,Wind and Water all have the highest gain of 0.322. This is already very different from the previous question. For consistency we now choose Sky from these 4 attributes to be our root node. This gives us the following decision tree:



References