

Assignment 1

Adriaan Louw (53031377)

May 6, 2018

Contents

1	Question 1	1
2	Question 2	1
3	Question 3: Report on k-Nearest Neighbour Algorithm	3
4	Question 4	4
4.1	Question 4a	5
4.2	Question 4b	6
4.3	Question 4c	7
4.4	Question 4d	7
5	Question 5	8
5.1	Question 5.1	8
5.2	Question 5.2	9
5.3	Question 5.3	10
5.4	5.4	10
6	Question 6	13
6.1	Question 6a	13
6.2	Question 6b	19

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

$$f(x_1, x_2, x_3, \dots x_n) \quad (1)$$

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

$$x_1.z_2 \quad (2)$$

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

$$x_1 + x_2 \quad (3)$$

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 $\bar{}$ as in

$$\bar{x}_1 \quad (4)$$

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

The . and + operations are commutative

$$\begin{aligned} x_1.x_2 &= x_2.x_1 \\ x_1 + x_2 &= x_2 + x_1 \end{aligned} \quad (5)$$

and associative

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

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

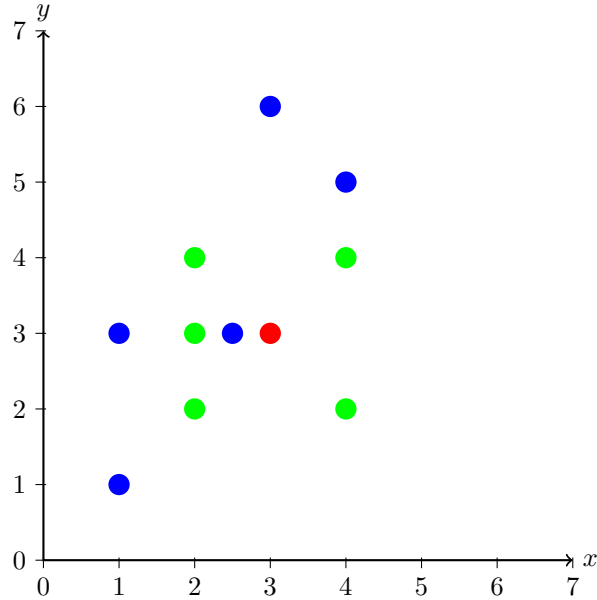
$$\begin{aligned} \overline{x_1.x_2} &= \bar{x}_1 + \bar{x}_2 \\ \overline{x_1 + x_2} &= \bar{x}_1.\bar{x}_2 \end{aligned} \quad (7)$$

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, $x_2.x_3.\bar{x}_6.x_{12}$. 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 $x_1 + \bar{x}_5 + x_8$.

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

3 Question 3: Report on k-Nearest Neighbour Algorithm



Consider the above diagram. Suppose we need to classify the red dot as either a blue dot or a green dot for an arbitrary machine learning task. We could use the k-nearest neighbour algorithm. Where k refers to how many of the red dots neighbours we use to deduce its value.

Let the Euclidean distance be defined as

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (8)$$

(University of Wasthington, 2012)

We use the euclidean distance to determine the nearest neighbour. If we only use the 1-nearest neighbour we find the nearest point to be (2.5, 3). We can deduce that the dot should be blue. On the other hand, if we use the 6-nearest neighbours we can see that the value should be green, since most of the 6 nearest neighbours are green. (Specifically referring to the points $\{(2.5, 3), (2, 3), (2, 2), (4, 2), (4, 4), (2, 4)\}$)

An odd number of nearest neighbours should be chosen in order to eliminate the possibility of a tie. As in the case with 2 nearest neighbours $\{(2.5, 3), (2, 3)\}$.

This case be generalised to a problem space (as in Mitchell (1997)) where an instance x is described by the feature vector

$$\langle a_1(x), a_2(x), \dots, a_n(x) \rangle \quad (9)$$

where a denotes the attributes of x . Then the distance will be defined as

$$d(x_i, x_j) \equiv \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (10)$$

(Mitchell, 1997)

This algorithm is susceptible to the *curse of dimensionality*. As the number of dimensions (or attributes) increases the contribution of an individual attribute decreases. If the data can only be categorised through a few or even 1 attribute, then the contribution of the important attributes will be "drowned out". (Murphy, 2012) To reduce the *curse of dimensionality*, the attributes could be weighted and additionally weighting attributes, we want to ignore, with 0. (Mitchell, 1997)

One issue with the k-nearest neighbour algorithm is that all computation generally done is done during the querying phase. Unlike decision trees where most of the computation can be done initially and the querying process being quite simple. (Mitchell, 1997)

This algorithm can also be extended to continuous functions as in

$$\hat{f}(x_q) \leftarrow \frac{\sum_{i=1}^k f(x_i)}{k} \quad (11)$$

4 Question 4

Firstly we need to convert the positions (x, y) into our hypothesis space $\langle a, b \rangle$. The quantities are related by the equation:

$$h \leftarrow \langle a < \sqrt{x^2 + y^2} < b \rangle \quad (12)$$

Given dataset

index	x	y
1	5	5
2	-6	4
3	-3	-4
4	2	-4
5	-1	2
6	-2	0
7	6	7
8	8	-8

For datapoint 1 we have

$$\begin{aligned} a &< \sqrt{x^2 + y^2} < b \\ a &< \sqrt{5^2 + 5^2} < b \\ a &< \sqrt{50} < b \\ a &< 7.071 < b \end{aligned} \quad (13)$$

Given that a and b are integers it is safe to assume they are the next integer i.e. $\langle a, b \rangle = \langle 7, 8 \rangle$.

A similar calculation gives us the values of the other data points in the hypothesis space. Note that in cases like $(x, y) = (-3, -4)$

$$\begin{aligned}
a &< \sqrt{x^2 + y^2} < b \\
a &< \sqrt{(-3)^2 + (-4)^2} < b \\
a &< \sqrt{25} < b \\
a &< 5 < b
\end{aligned} \tag{14}$$

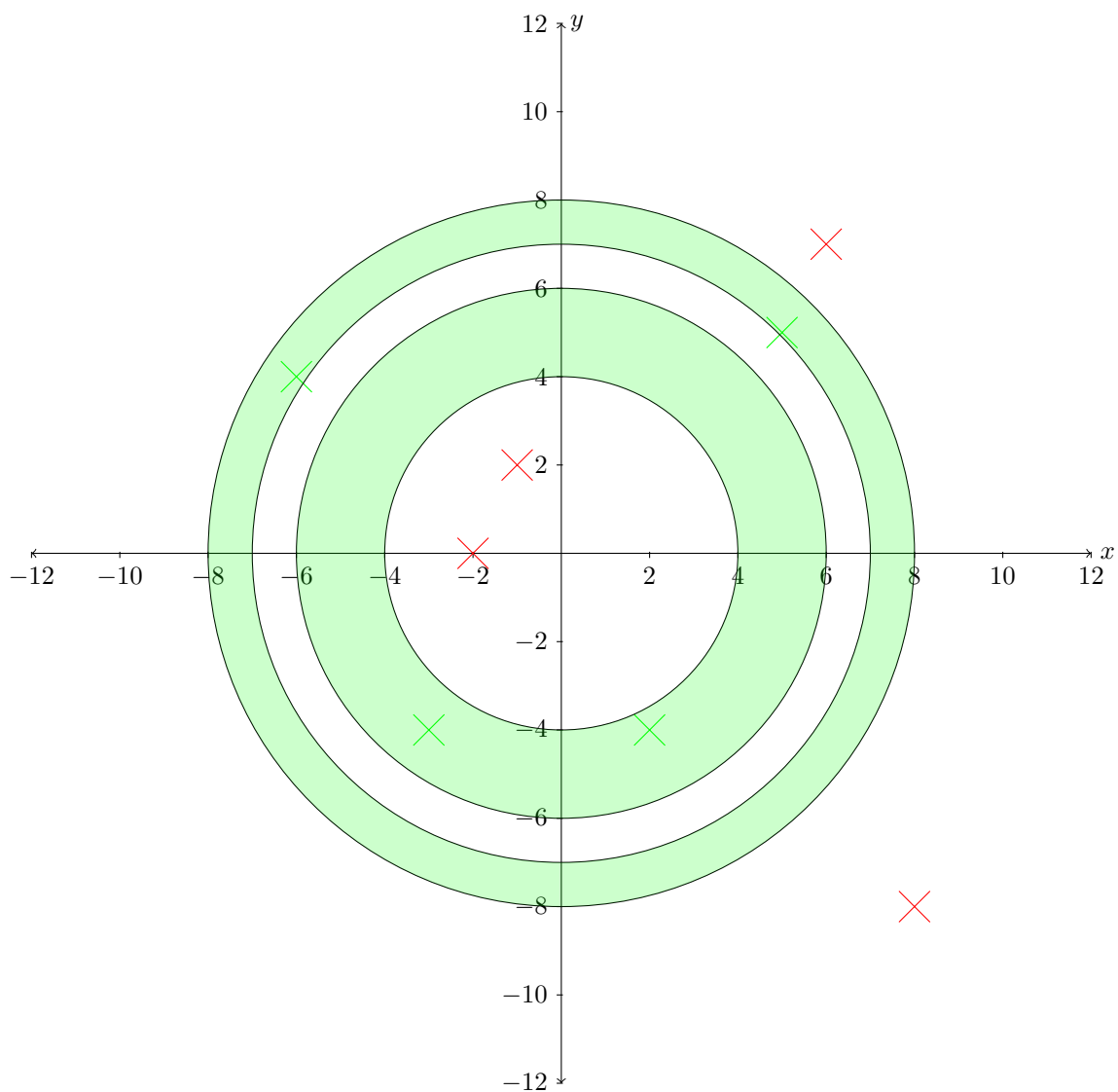
the middle term is a whole integer. Meaning $\langle a, b \rangle = \langle 4, 6 \rangle$

Thus our dataset becomes

index	x	y	a	b
1	5	5	7	8
2	-6	4	7	8
3	-3	-4	4	6
4	2	-4	4	5
5	-1	2	2	3
6	-2	0	1	3
7	6	7	9	10
8	8	-8	11	12

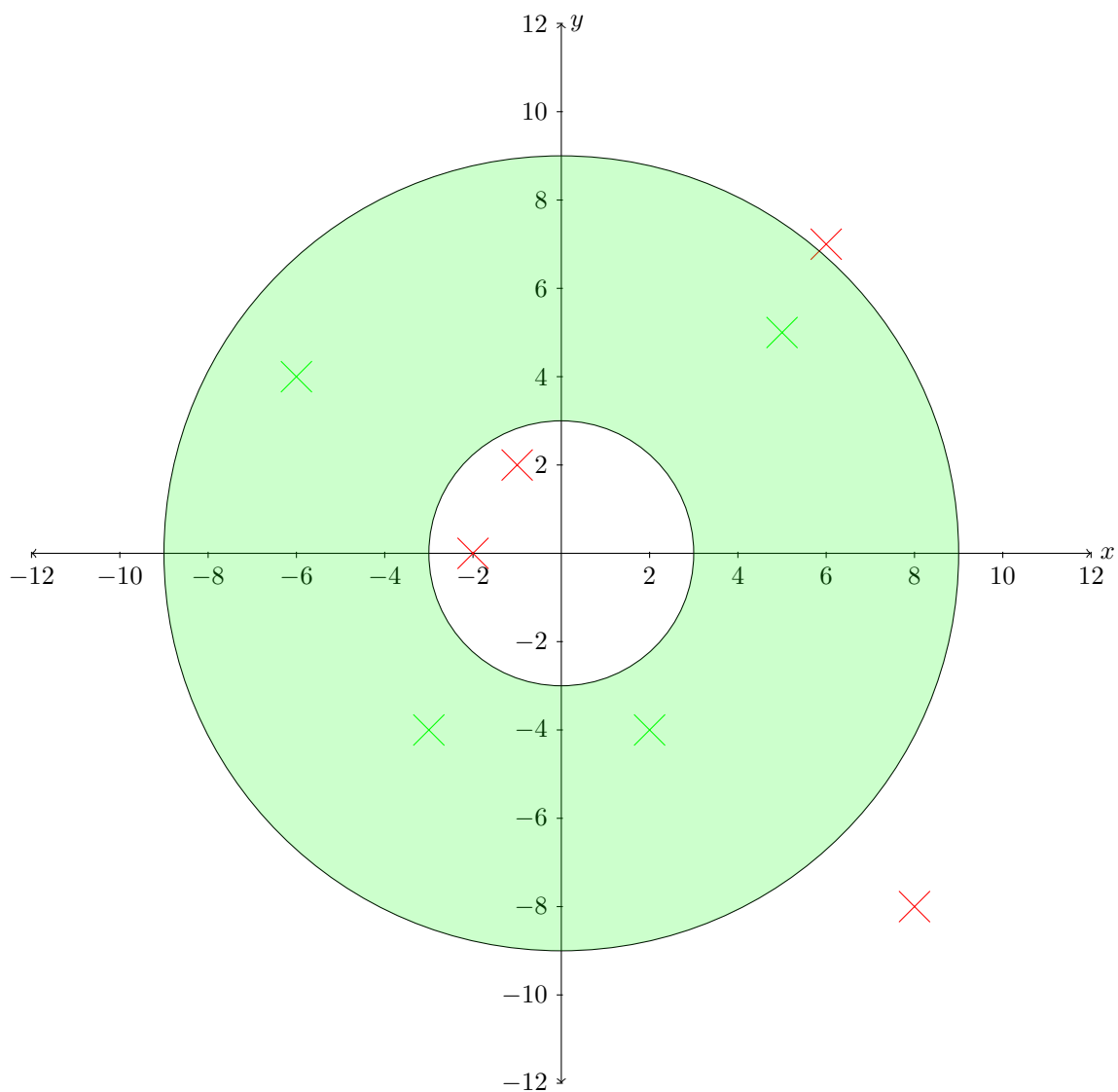
4.1 Question 4a

$$S = \{\langle 4, 6 \rangle, \langle 7, 8 \rangle\} \tag{15}$$



4.2 Question 4b

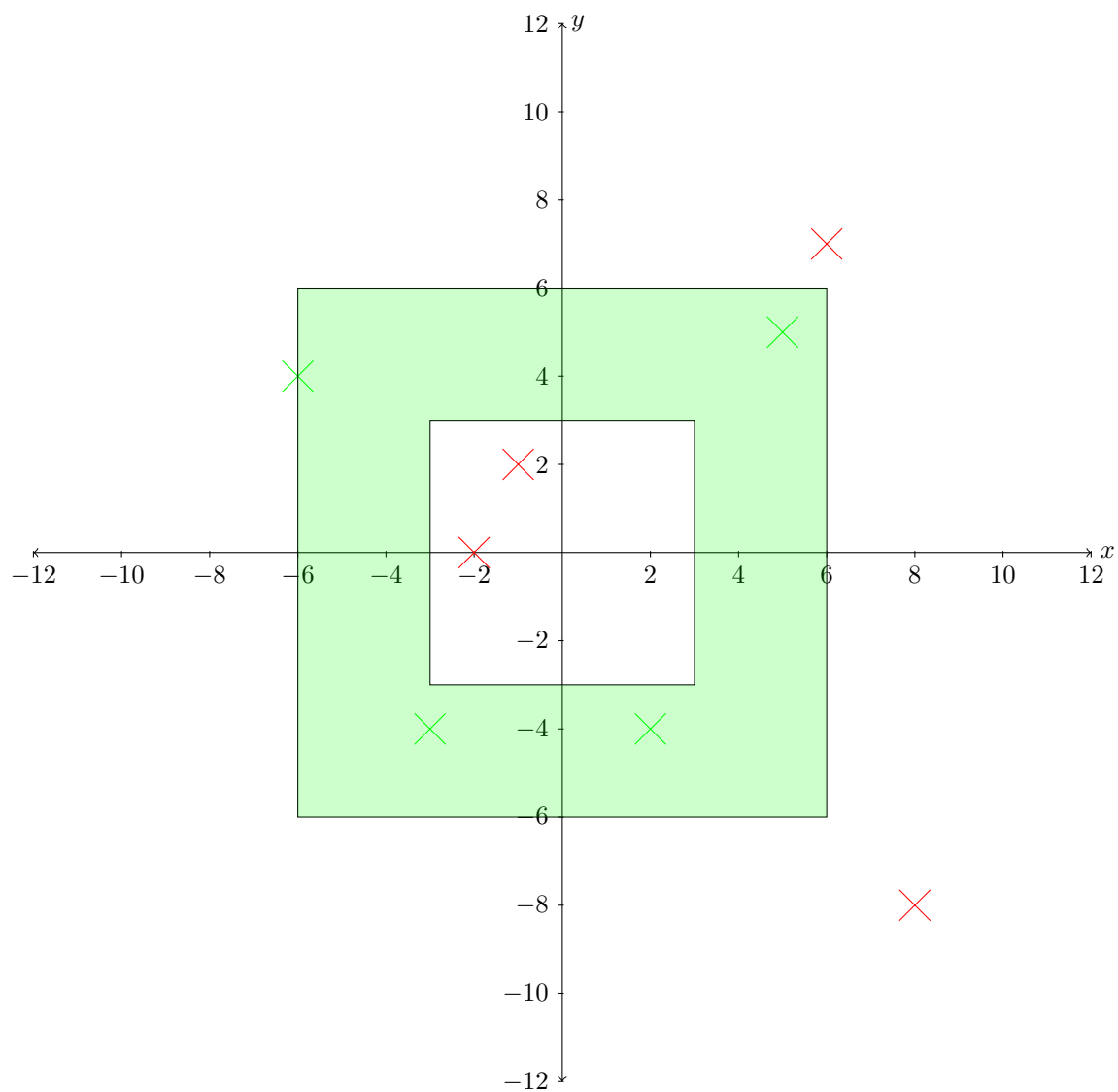
$$S = \{(3, 9)\} \quad (16)$$



4.3 Question 4c

4.4 Question 4d

We could use a rectangle of the form $\langle (a, b), (c, d) \rangle$ where $a, b, c, d \in \mathbb{Z}$. (a, b) is the position of the lower left hand corner of the outer square and (c, d) is the lower left hand corner of the inner square. $h \leftarrow \langle (-6, -6), (-3, -3) \rangle$

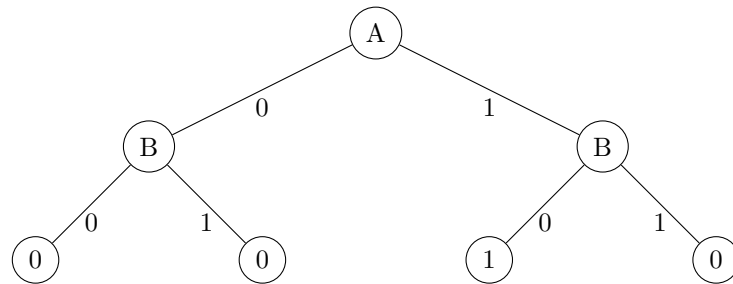


5 Question 5

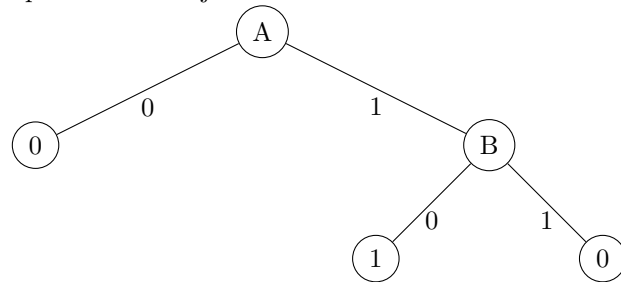
5.1 Question 5.1

For the equation $A \wedge \neg B$:

A	B	$\neg B$	$A \wedge \neg B$
0	0	1	0
0	1	1	0
1	0	1	1
1	1	0	0



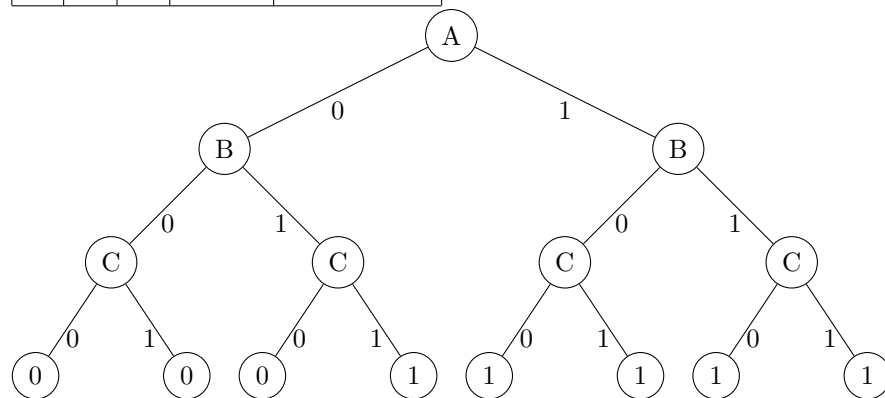
When $A=0$, B returns 0 whether we have the value 0 or 1 for B . We can collapse this B and just return 0.



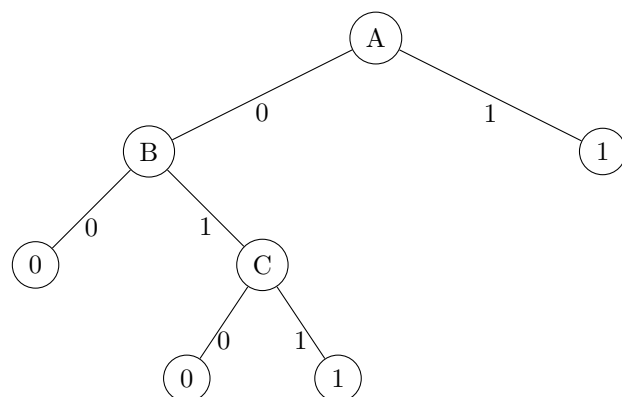
5.2 Question 5.2

For the equation $A \vee [B \wedge C]$:

A	B	C	$B \wedge C$	$A \vee [B \wedge C]$
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	1	1
1	0	0	0	1
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1



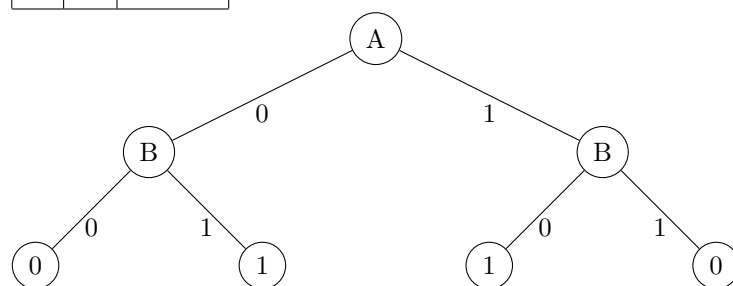
When $A = 0$ and $B = 0$, the result is 0 independent form the value of C . This branch will be reduced. Additionally When $A = 1$ the result will be 1 no matter the value of B or C . This will also be reduced. Our new decision tree will be:



5.3 Question 5.3

For the equation $A \oplus B$:

A	B	$A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

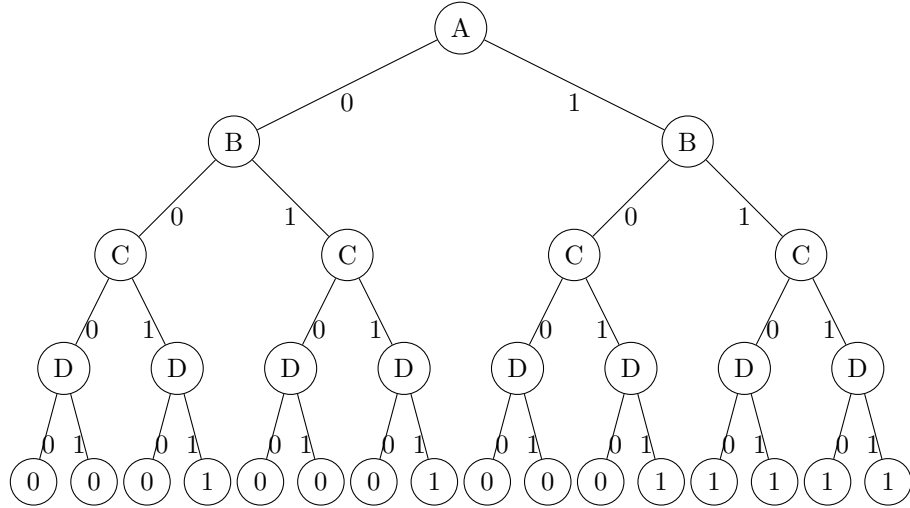


This tree cannot be simplified further

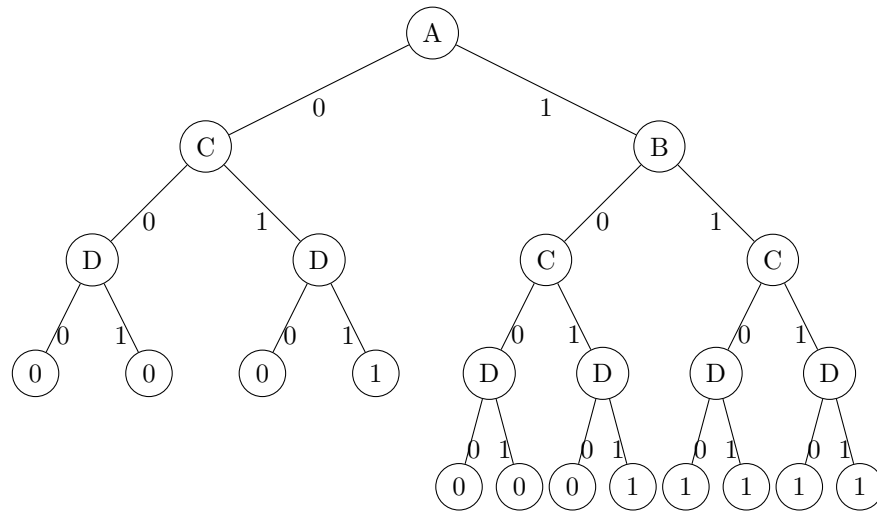
5.4 5.4

For the equation $[A \wedge B] \vee [C \wedge D]$:

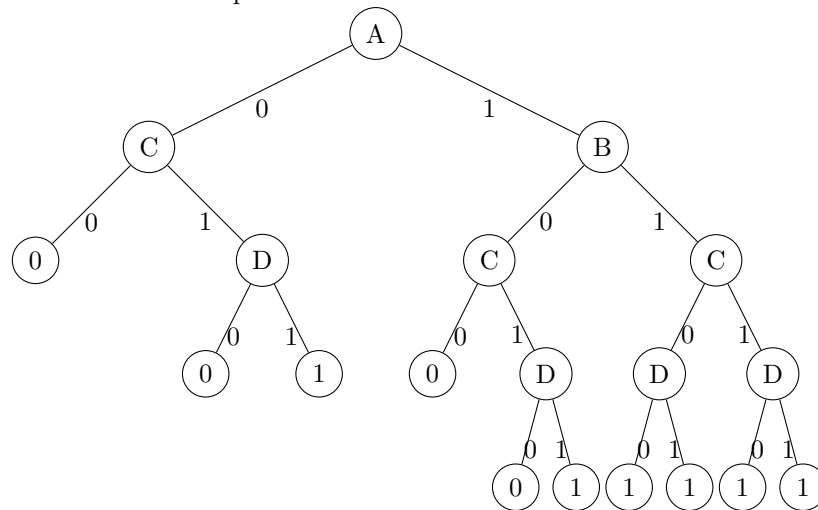
A	B	C	D	$A \wedge B$	$C \wedge D$	$[A \wedge B] \vee [C \wedge D]$
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	0	1	0	0	0	0
0	0	1	1	0	1	1
0	1	0	0	0	0	0
0	1	0	1	0	0	0
0	1	1	0	0	0	0
0	1	1	1	0	1	1
1	0	0	0	0	0	0
1	0	0	1	0	0	0
1	0	1	0	0	0	0
1	0	1	1	1	1	1
1	1	0	0	1	0	1
1	1	0	1	1	0	1
1	1	1	0	1	0	1
1	1	1	1	1	1	1



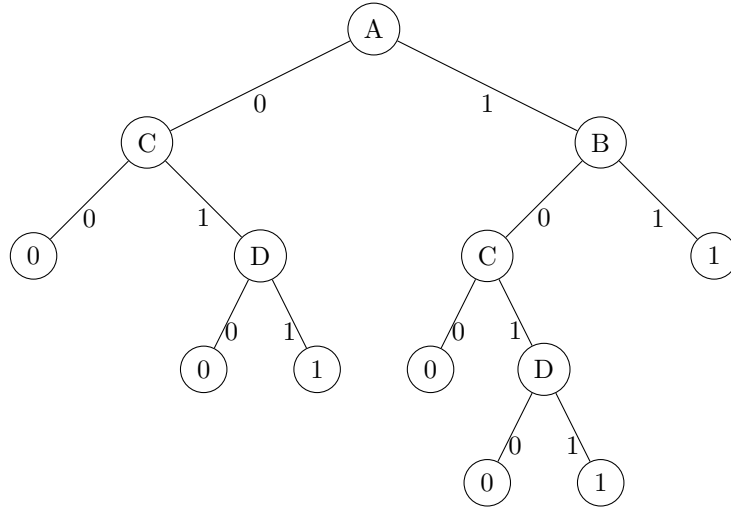
Firstly we can see that when $A = 0$, the trees under $B = 0$ and $B = 1$ are identical so they can be combined.



When $A = 0$ and $C = 0$, the result is not dependent on D . D collapses to 0. Additionally when $A = 1$, $B = 0$ and $C = 0$ the answer is always 0. That subbranch is also collapsed.



Finally when $A = 1$ and $B = 1$ the answer will always be 0 and not dependent on C or D . This branch is collapsed.



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):

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (17)$$

If we assume we only have 2 outcomes, then we let the number of positive outcomes be p and the number of negative outcomes be n . Then we can write Equation 17 as

$$Entropy(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \quad (18)$$

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

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (19)$$

We need to determine $Entropy(S)$

$$\begin{aligned} Entropy(p, n) &= -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \\ &= -\frac{3}{3+1} \log_2 \frac{3}{3+1} - \frac{1}{3+1} \log_2 \frac{1}{3+1} \\ &= 0.811 \end{aligned} \quad (20)$$

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 $p_2 = 0, n_2 = 1$)

$$\begin{aligned}
Gain(S, Sky) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{3}{4} Entropy(S_{Sunny}) - \frac{1}{4} Entropy(S_{Rainy}) \\
&= 0.811 - \frac{3}{4} Entropy(3, 0) - \frac{1}{4} Entropy(0, 1) \\
&= 0.811 - \frac{3}{4} \left(-\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} \right) - \frac{1}{4} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.811 - 0 \\
&= 0.811
\end{aligned} \tag{21}$$

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

$$\begin{aligned}
Gain(S, AirTemp) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{3}{4} Entropy(S_{Warm}) - \frac{1}{4} Entropy(S_{Cold}) \\
&= 0.811 - \frac{3}{4} Entropy(3, 0) - \frac{1}{4} Entropy(0, 1) \\
&= 0.811 - \frac{3}{4} \left(-\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} \right) - \frac{1}{4} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.811 - 0 \\
&= 0.811
\end{aligned} \tag{22}$$

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

$$\begin{aligned}
Gain(S, Humidity) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{1}{4} Entropy(S_{Normal}) - \frac{3}{4} Entropy(S_{High}) \\
&= 0.811 - \frac{1}{4} Entropy(1, 0) - \frac{3}{4} Entropy(2, 1) \\
&= 0.811 - \frac{1}{4} \left(-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} \right) - \frac{3}{4} \left(-\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \right) \\
&= 0.811 - 0.689 \\
&= 0.123
\end{aligned} \tag{23}$$

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

$$\begin{aligned}
Gain(S, Wind) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{4}{4} Entropy(S_{strong}) \\
&= 0.811 - \frac{4}{4} Entropy(3, 1) \\
&= 0.811 - \frac{4}{4} \left(-\frac{4}{4} \log_2 \frac{4}{4} - \frac{0}{4} \log_2 \frac{0}{4} \right) \\
&= 0.811 - 0.811 \\
&= 0.0
\end{aligned} \tag{24}$$

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

$$\begin{aligned}
Gain(S, Water) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{3}{4} Entropy(S_{warm}) - \frac{1}{4} Entropy(S_{high}) \\
&= 0.811 - \frac{3}{4} Entropy(2, 1) - \frac{1}{4} Entropy(1, 0) \\
&= 0.811 - \frac{3}{4} \left(-\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \right) - \frac{1}{4} \left(-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} \right) \\
&= 0.811 - 0.689 \\
&= 0.123
\end{aligned} \tag{25}$$

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

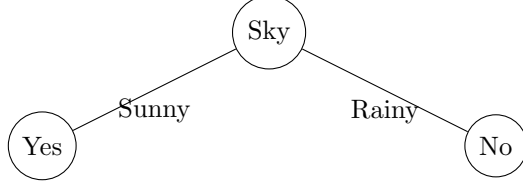
$$\begin{aligned}
Gain(S, Forecast) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= 0.811 - \frac{2}{4} Entropy(S_{same}) - \frac{2}{4} Entropy(S_{change}) \\
&= 0.811 - \frac{2}{4} Entropy(2, 0) - \frac{1}{1} Entropy(1, 0) \\
&= 0.811 - \frac{2}{4} \left(-\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2} \right) - \frac{2}{4} \left(-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \right) \\
&= 0.811 - 0.5 \\
&= 0.311
\end{aligned} \tag{26}$$

To recap:

Gain(Sky) = 0.811 from Equation 21
Gain(AirTemp) = 0.811 from Equation 22
Gain(Humidity) 0.123 from Equation 23
Gain(Wind) = 0 from Equation 24

Gain(Water) = 0.123 from Equation 25
Gain(Forecast) = 0.311 from Equation 26

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:

$$\begin{aligned}
Entropy(S_{Sky=Sunny}) &= Entropy(p, n) \\
&= -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \\
&= -\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} \\
&= 0
\end{aligned} \tag{27}$$

We need to calculate the information gain for the other attributes when Sky = sunny

For AirTemp we only have warm ($p_1 = 3, n_1 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny}, AirTemp) &= 0.0 - \frac{3}{3} Entropy(S_{Sky=Sunny; AirTemp=Warm}) \\
&= 0.0 - \frac{3}{3} Entropy(3, 0) \\
&= 0.0 - 1(-\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3}) \\
&= 0.0
\end{aligned} \tag{28}$$

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

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Humidity) &= 0.0 - \frac{1}{3} Entropy(S_{Sky=Sunny; Humidity=Normal}) \\
&\quad - \frac{2}{3} Entropy(S_{Sky=Sunny; Humidity=High}) \\
&= 0.0 - \frac{1}{3} Entropy(1, 0) - \frac{2}{3} Entropy(2, 0) \\
&= 0.0 - \frac{1}{3}(-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1}) - \frac{2}{3}(-\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2}) \\
&= 0.0
\end{aligned} \tag{29}$$

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

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Wind) &= 0.0 - \frac{3}{3} Entropy(S_{Sky=Sunny}; Wind=Strong) \\
&= 0.0 - \frac{3}{3} Entropy(3, 0) \\
&= 0.0 - 1(-\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3}) \\
&= 0.0
\end{aligned} \tag{30}$$

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

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Water) &= 0.0 - \frac{2}{3} Entropy(S_{Sky=Sunny}; Water=Warm) \\
&\quad - \frac{1}{3} Entropy(S_{Sky=Sunny}; Water=Cool) \\
&= 0.0 - \frac{2}{3} Entropy(2, 0) - \frac{1}{3} Entropy(1, 0) \\
&= 0.0 - \frac{2}{3}(-\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2}) - \frac{1}{3}(-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1}) \\
&= 0.0
\end{aligned} \tag{31}$$

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

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Forecast) &= 0.0 - \frac{2}{3} Entropy(S_{Sky=Sunny}; Forecast=Same) \\
&\quad - \frac{1}{3} Entropy(S_{Sky=Sunny}; Forecast=Change) \\
&= 0.0 - \frac{2}{3} Entropy(2, 0) - \frac{1}{3} Entropy(1, 0) \\
&= 0.0 - \frac{2}{3}(-\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2}) - \frac{1}{3}(-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1}) \\
&= 0.0
\end{aligned} \tag{32}$$

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

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

$$\begin{aligned}
Entropy(S_{Sky=Rainy}) &= Entropy(p, n) \\
&= -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \\
&= -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \\
&= 0
\end{aligned} \tag{33}$$

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

$$\begin{aligned}
Gain(S_{Sky=Rainy}, AirTemp) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Rainy}; AirTemp=Cold) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - 1(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{34}$$

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

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Humidity) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Rainy}; Humidity=High) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - 1(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{35}$$

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

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Wind) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Rainy}; Wind=Strong) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - 1(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{36}$$

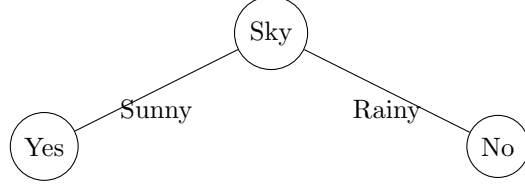
For Water we only have Warm ($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Water) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Rainy}; Water=Warm) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - 1(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{37}$$

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

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Forecast) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Rainy}; ForeCast=Change) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - 1(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1}) \\
&= 0.0
\end{aligned} \tag{38}$$

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

$$\begin{aligned}
 Entropy(p, n) &= -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \\
 &= -\frac{3}{3+2} \log_2 \frac{3}{3+2} - \frac{2}{3+2} \log_2 \frac{2}{3+2} \\
 &= 0.971
 \end{aligned} \tag{39}$$

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

$$\begin{aligned}
 Gain(S, Sky) &= 0.971 - \frac{4}{5} Entropy(S_{Sunny}) - \frac{1}{5} Entropy(S_{Rainy}) \\
 &= 0.971 - \frac{4}{5} Entropy(3, 1) - \frac{1}{5} Entropy(0, 1) \\
 &= 0.971 - \frac{4}{5} \left(-\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \right) - \frac{1}{5} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
 &= 0.971 - 0.649 \\
 &= 0.322
 \end{aligned} \tag{40}$$

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

$$\begin{aligned}
 Gain(S, AirTemp) &= 0.971 - \frac{4}{5} Entropy(S_{Warm}) - \frac{1}{5} Entropy(S_{Cold}) \\
 &= 0.971 - \frac{4}{5} Entropy(3, 1) - \frac{1}{5} Entropy(0, 1) \\
 &= 0.971 - \frac{4}{5} \left(-\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \right) - \frac{1}{5} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
 &= 0.971 - 0.649 \\
 &= 0.322
 \end{aligned} \tag{41}$$

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

$$\begin{aligned}
Gain(S, Humidity) &= 0.971 - \frac{2}{5}Entropy(S_{Normal}) - \frac{3}{5}Entropy(S_{High}) \\
&= 0.971 - \frac{2}{5}Entropy(1, 1) - \frac{3}{5}Entropy(2, 1) \\
&= 0.971 - \frac{2}{5}(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}) - \frac{3}{5}(-\frac{0}{1}\log_2\frac{0}{1} - \frac{1}{1}\log_2\frac{1}{1}) \\
&= 0.971 - 0.951 \\
&= 0.020
\end{aligned} \tag{42}$$

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

$$\begin{aligned}
Gain(S, Wind) &= 0.971 - \frac{4}{5}Entropy(S_{Strong}) - \frac{1}{5}Entropy(S_{Weak}) \\
&= 0.971 - \frac{4}{5}Entropy(3, 1) - \frac{1}{5}Entropy(0, 1) \\
&= 0.971 - \frac{4}{5}(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}) - \frac{1}{5}(-\frac{0}{1}\log_2\frac{0}{1} - \frac{1}{1}\log_2\frac{1}{1}) \\
&= 0.971 - 0.649 \\
&= 0.322
\end{aligned} \tag{43}$$

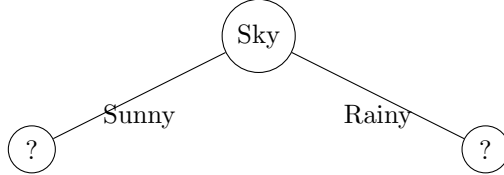
For the attribute Water we have Warm ($p_1 = 3, n_1 = 1$) and Cool ($p_2 = 1, n_2 = 0$)

$$\begin{aligned}
Gain(S, Water) &= 0.971 - \frac{4}{5}Entropy(S_{Warm}) - \frac{1}{5}Entropy(S_{Cool}) \\
&= 0.971 - \frac{4}{5}Entropy(3, 1) - \frac{1}{5}Entropy(0, 1) \\
&= 0.971 - \frac{4}{5}(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}) - \frac{1}{5}(-\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1}) \\
&= 0.971 - 0.649 \\
&= 0.322
\end{aligned} \tag{44}$$

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

$$\begin{aligned}
Gain(S, Forecast) &= 0.971 - \frac{3}{5}Entropy(S_{Same}) - \frac{2}{5}Entropy(S_{Change}) \\
&= 0.971 - \frac{3}{5}Entropy(2, 1) - \frac{2}{5}Entropy(1, 1) \\
&= 0.971 - \frac{3}{5}(-\frac{2}{3}\log_2 \frac{2}{3} - \frac{1}{3}\log_2 \frac{1}{3}) - \frac{2}{5}(-\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2}) \\
&= 0.971 - 0.951 \\
&= 0.020
\end{aligned} \tag{45}$$

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:



Continuing with Sky = Sunny, we need to determine the gain of the other attributes so that we can determine the sub-tree. We know that there are 3 positive data rows and one negative for Sky = Sunny:

$$\begin{aligned}
Entropy(S_{Sky=Sunny}) &= Entropy(p, n) \\
&= -\frac{p}{p+n}\log_2 \frac{p}{p+n} - \frac{n}{p+n}\log_2 \frac{n}{p+n} \\
&= -\frac{3}{4}\log_2 \frac{3}{4} - \frac{1}{4}\log_2 \frac{1}{4} \\
&= 0.811
\end{aligned} \tag{46}$$

For attribute AirTemp we have only the value Warm ($p_1 = 3, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Sunny}, AirTemp) &= 0.811 - \frac{3}{4}Entropy(S_{Sky=Sunny; AirTemp=Warm}) \\
&= 0.811 - \frac{4}{4}Entropy(3, 1) \\
&= 0.811 - 1(-\frac{3}{4}\log_2 \frac{3}{4} - \frac{1}{4}\log_2 \frac{1}{4}) \\
&= 0.811 - 0.811 \\
&= 0
\end{aligned} \tag{47}$$

For attribute Humidity we have Normal($p_1 = 1, n_1 = 1$) and High ($p_2 = 2, n_2 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Humidity) &= 0.811 - \frac{2}{4} Entropy(S_{Sky=Sunny}; Humidity=Normal) \\
&\quad - \frac{2}{4} Entropy(S_{Sky=Sunny}; Humidity=High) \\
&= 0.811 - \frac{2}{4} Entropy(1, 1) - \frac{2}{4} Entropy(2, 0) \\
&= 0.811 - \frac{2}{4} \left(-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} \right) - \frac{2}{4} \left(-\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2} \right) \\
&= 0.811 - 0.5 \\
&= 0.311
\end{aligned} \tag{48}$$

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

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Wind) &= 0.811 - \frac{3}{4} Entropy(S_{Sky=Sunny}; Wind=Strong) \\
&\quad - \frac{1}{4} Entropy(S_{Sky=Sunny}; Wind=Weak) \\
&= 0.811 - \frac{3}{4} Entropy(3, 0) - \frac{1}{4} Entropy(0, 1) \\
&= 0.811 - \frac{3}{4} \left(-\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} \right) - \frac{1}{4} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.811 - 0.0 \\
&= 0.811
\end{aligned} \tag{49}$$

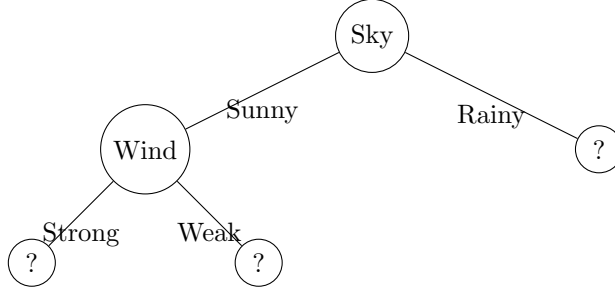
For attribute Water we have Warm($p_1 = 2, n_1 = 1$) and Cool ($p_2 = 1, n_2 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Water) &= 0.811 - \frac{3}{4} Entropy(S_{Sky=Sunny}; Water=Warm) \\
&\quad - \frac{1}{4} Entropy(S_{Sky=Sunny}; Water=Cool) \\
&= 0.811 - \frac{3}{4} Entropy(2, 1) - \frac{1}{4} Entropy(1, 0) \\
&= 0.811 - \frac{3}{4} \left(-\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \right) - \frac{1}{4} \left(-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} \right) \\
&= 0.811 - 0.689 \\
&= 0.123
\end{aligned} \tag{50}$$

For attribute Forecast we have Same($p_1 = 2, n_1 = 1$) and Weak ($p_2 = 1, n_2 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Forecast) &= 0.811 - \frac{3}{4} Entropy(S_{Sky=Sunny}; Forecast=Same) \\
&\quad - \frac{1}{4} Entropy(S_{Sky=Sunny}; Forecast=Change) \\
&= 0.811 - \frac{3}{4} Entropy(2, 1) - \frac{1}{4} Entropy(1, 0) \\
&= 0.811 - \frac{3}{4} \left(-\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \right) - \frac{1}{4} \left(-\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} \right) \\
&= 0.811 - 0.689 \\
&= 0.123
\end{aligned} \tag{51}$$

Therefore the attribute with the greatest gain when Sky=Sunny is Wind. The Wind attribute will now form our new sub-tree at Sky = Sunny



Now we have to calculate the gains for Sky=Sunny and Wind=Strong. Starting with the Entropy(S) calculation. We can see from the data that there is 3 positive data rows and 0 negative data rows for Sky=Sunny and Wind=Strong

$$\begin{aligned}
Entropy(p, n) &= -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \\
&= -\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} \\
&= 0.0
\end{aligned} \tag{52}$$

For attribute AirTemp we have only the value Warm ($p_1 = 3, n_1 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny; Wind=Strong}, AirTemp) &= 0.811 - \frac{3}{3} Entropy(S_{Sky=Sunny; Wind=Strong}; AirTemp=Warm) \\
&= 0.0 - \frac{3}{3} Entropy(3, 0) \\
&= 0.0 - 1 \left(-\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{53}$$

For attribute Humidity we have the values Normal ($p_1 = 1, n_1 = 0$) and High ($p_2 = 2, n_2 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny;Wind=Strong}, Humidity) &= 0.0 - \frac{1}{3}Entropy(S_{Sky=Sunny;Wind=Strong, Humidity=Normal}) \\
&\quad - \frac{2}{3}Entropy(S_{Sky=Sunny;Wind=Strong, Humidity=High}) \\
&= 0.0 - \frac{1}{3}Entropy(1, 0) - \frac{2}{3}Entropy(2, 0) \\
&= 0.0 - \frac{1}{3}(-\frac{1}{1}\log_2 \frac{1}{1} - \frac{0}{1}\log_2 \frac{0}{1}) - \frac{2}{3}(-\frac{2}{2}\log_2 \frac{2}{2} - \frac{0}{2}\log_2 \frac{0}{2}) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{54}$$

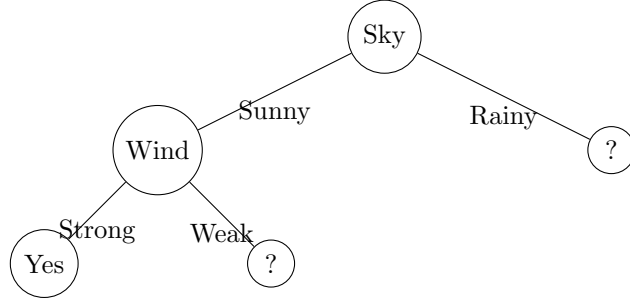
For attribute Water we have values Cool ($p_1 = 1, n_1 = 0$) and Warm($p_2 = 2, n_2 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny;Wind=Strong}, Water) &= 0.0 - \frac{1}{3}Entropy(S_{Sky=Sunny;Wind=Strong, Water=Cool}) \\
&\quad - \frac{2}{3}Entropy(S_{Sky=Sunny;Wind=Strong, Water=Warm}) \\
&= 0.0 - \frac{1}{3}Entropy(1, 0) - \frac{2}{3}Entropy(2, 0) \\
&= 0.0 - \frac{1}{3}(-\frac{1}{1}\log_2 \frac{1}{1} - \frac{0}{1}\log_2 \frac{0}{1}) - \frac{2}{3}(-\frac{2}{2}\log_2 \frac{2}{2} - \frac{0}{2}\log_2 \frac{0}{2}) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{55}$$

For attribute Forecast we have values Change ($p_1 = 1, n_1 = 0$) and Same($p_2 = 2, n_2 = 0$)

$$\begin{aligned}
Gain(S_{Sky=Sunny;Wind=Strong}, Forecast) &= 0.0 - \frac{1}{3}Entropy(S_{Sky=Sunny;Wind=Strong, Forecast=Change}) \\
&\quad - \frac{2}{3}Entropy(S_{Sky=Sunny;Wind=Strong, Forecast=Same}) \\
&= 0.0 - \frac{1}{3}Entropy(1, 0) - \frac{2}{3}Entropy(2, 0) \\
&= 0.0 - \frac{1}{3}(-\frac{1}{1}\log_2 \frac{1}{1} - \frac{0}{1}\log_2 \frac{0}{1}) - \frac{2}{3}(-\frac{2}{2}\log_2 \frac{2}{2} - \frac{0}{2}\log_2 \frac{0}{2}) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{56}$$

We can see that the gains for all the other attributes, when Sky = Sunny and Wind = Strong, are 0. This means that this branch of the tree is complete giving:



We can classify the leaf node to be yes because all training rows when Sky = Sunny and Wind = Strong has a yes value for EnjoySport.

Now we need to classify the branch Sky = Sunny and Wind = Weak. There are no positive and one negative data rows for Sky = Sunny and Wind = Weak

$$\begin{aligned}
 Entropy(S_{Sky=Sunny, Wind=Weak}) &= Entropy(p, n) \\
 &= -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \\
 &= -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \\
 &= 0
 \end{aligned} \tag{57}$$

For the attribute AirTemp we have Warm($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
 Gain(S_{Sky=Sunny; Wind=Weak}, AirTemp) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny; Wind=Weak; Airtemp=Warm}) \\
 &= 0.0 - \frac{1}{1} Entropy(0, 1) \\
 &= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
 &= 0.0 - 0.0 \\
 &= 0
 \end{aligned} \tag{58}$$

For the attribute Humidity we have Normal($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
 Gain(S_{Sky=Sunny; Wind=Weak}, Humidity) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny; Wind=Weak; Humidity=Normal}) \\
 &= 0.0 - \frac{1}{1} Entropy(0, 1) \\
 &= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
 &= 0.0 - 0.0 \\
 &= 0
 \end{aligned} \tag{59}$$

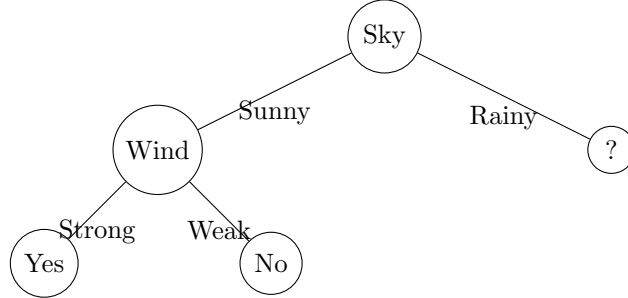
For the attribute Water we have Warm($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Wind=Weak, Water) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny}, Wind=Weak, Water=Warm) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{60}$$

For the attribute Forecast we have Same($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Sunny}, Wind=Weak, Forecast) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny}, Wind=Weak, Forecast=Same) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{61}$$

Now we can see that the gains of all the attributes, when Sky = Sunny and Wind = Weak, are 0. Additionally we can see that this node would return No, because it is the only option according to our data for Sky = Sunny and Wind = Weak. This gives us:



The only branch that still needs classifying is the Sky = Rainy branch. Now we will calculate the gains for the other attributes in this case. For Sky = Rainy, we have only one data row. Which means we have 0 positive and 1 negative data row.

$$\begin{aligned}
Entropy(S_{Sky=Rainy}) &= Entropy(p, n) \\
&= -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \\
&= -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \\
&= 0
\end{aligned} \tag{62}$$

For the attribute AirTemp we have Cold($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, AirTemp) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny}; AirTemp=Cold) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{63}$$

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

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Humidity) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny}; Humidity=High) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{64}$$

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

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Wind) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny}; Wind=Strong) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{65}$$

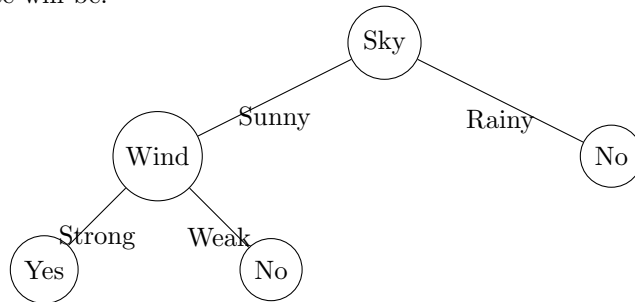
For the attribute Water we have Warm($p_1 = 0, n_1 = 1$)

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Water) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny}; Water=Warm) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned} \tag{66}$$

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

$$\begin{aligned}
Gain(S_{Sky=Rainy}, Forecast) &= 0.0 - \frac{1}{1} Entropy(S_{Sky=Sunny; Forecast=Change}) \\
&= 0.0 - \frac{1}{1} Entropy(0, 1) \\
&= 0.0 - \frac{1}{1} \left(-\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} \right) \\
&= 0.0 - 0.0 \\
&= 0
\end{aligned}
\tag{67}$$

We can see from the gains for the attributes when Sky = Rainy, that we do not create a new sub-tree but make this a leaf node. From the only applicable data row we can see that this node will return No. Therefore the final decision tree will be:



References

- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*.
University of Washington. (2012). *Nearest neighbor*. Retrieved from
http://courses.washington.edu/css490/2012.Winter/lecture_slides/06c_nearest_neighbor.pdf