Question 1

Figure 1 shows a game tree to be searched based on the minimax algorithm. The root of the tree corresponds to Max's turn.

a) Fill in the minimax value at each node.

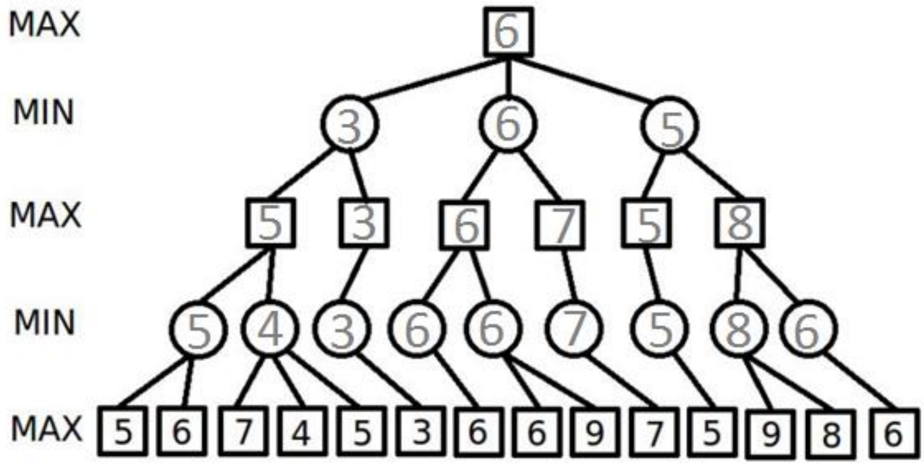


Figure 1: Game tree

b) If *alpha-beta pruning* is to be used to search the game tree, clearly circle the branches that are pruned (i.e, branches that do *not* need to be searched). Note: at each node, search the child nodes from left to right.

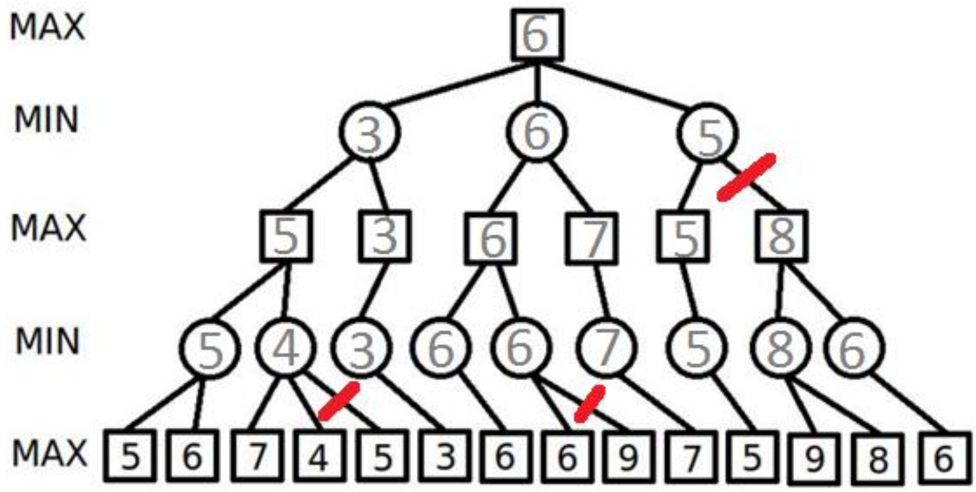


Figure 1: Game tree

Question 2 (Question 6.1 of AIMA 2ed)
Take the game of tic-tac-toe. Draw the game tree to a depth of two (shows the starting board at root and then has two more levels). *Use symmetry to avoid drawing states that are essentially equivalent.* Now annotate the "leaf nodes" (leaves of this partial tree) with utility according to the following. We define X_n as the number of rows, columns or diagonals with nothing else in them but n X's. Likewise define O_n for the noughts.

For a complete tree, the utility function assigns +1 to any position with $X_3 = 1$ (the game has been won by the X player) and -1 to any position with $O_3 = 1$ (the game has been won by the O player). All other terminal nodes would have utility zero.

Non-terminal leaf nodes (i.e., leaves of a partial tree for a certain lookahead) are given a utility by a linear relation $Eval(s) = 3X_2(s) + X_1(s) - (3O_2(s) + O_1(s))$.

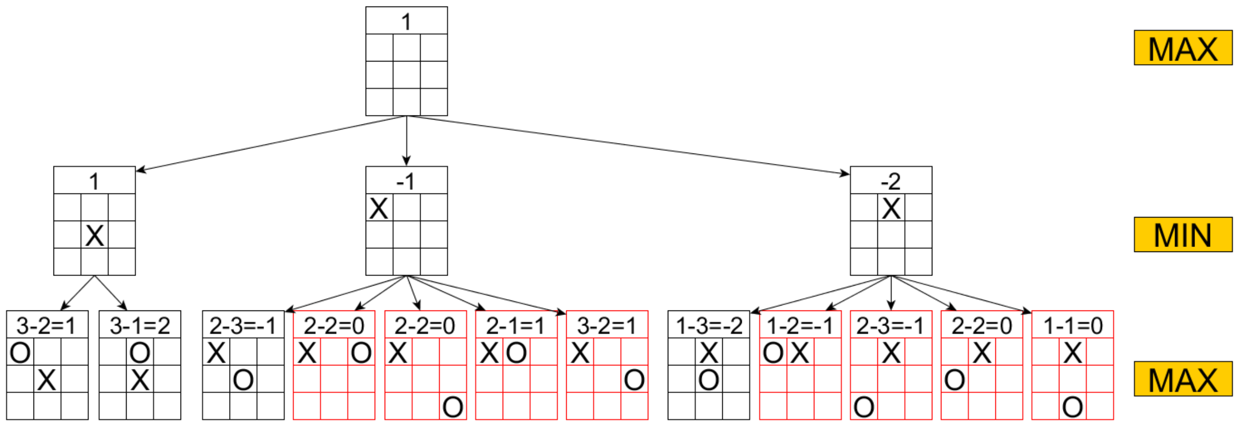
Annotate your leaf nodes with their cost and then use Min-Max analysis to annotate all other nodes.

a) What is the best starting move according to this analysis?

Centre

b) Circle the nodes that would not be evaluated using alpha-beta pruning *assuming an ordering of the leaf nodes that maximises the benefit of such pruning*.

Arranging the leaf nodes in ascending order will maximize the benefit of pruning. In this case we sort the nodes, but we'd ideally design the algorithm to generate them in the correct order (such as place in the centre, then diagonals and then edge centres). We could extend the pruning to also exclude the node with value 2, as we also know this will not change the result (as was done in the tutorial).



Question 3

Figure 2 shows a scatter plot of points $p_1, p_2, \ldots, p_{10}$ with their class labels. The plot also includes the testing point z.

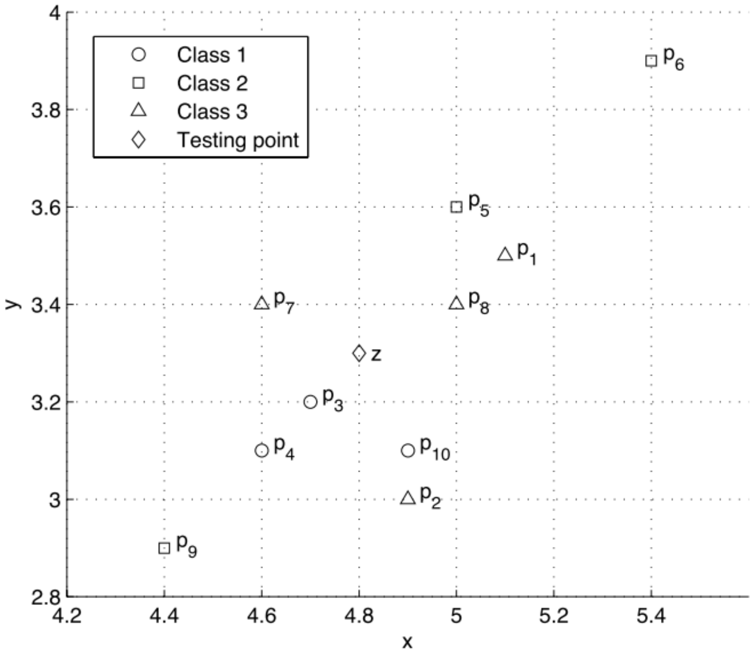


Figure 2: Scatter plot of points with class labels.

The precise coordinates of all the points are as follows:

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	z
x	5.1	4.9	4.7	4.6	5.0	5.4	4.6	5.0	4.4	4.9	4.8
y	3.5	3.0	3.2	3.1	3.6	3.9	3.4	3.4	2.9	3.1	3.3

Classify the testing point z using K nearest neighbours with $K = 1, 3, 4, 5$ and 7.

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
Dist	0.361	0.316	0.141	0.283	0.361	0.849	0.224	0.224	0.567	0.224

KNN-1 Use p_3 Class 1

KNN-3 Pick p_3,(p_7/p_8),p_{10} Class 1, Pick p_3,p_7,p_8 Class 3

KNN-4 Pick p_3,p_7,p_8,p_{10} Indeterminate (Class 1 or Class 3)

KNN-5 Pick p_3,p_7,p_8,p_{10},p_4 Class 1

KNN-7 Pick p_3,p_7,p_8,p_{10},p_4,p_2,p_1 Class 3, Pick p_3,p_7,p_8,p_{10},p_4,p_2,p_5 Indeterminate (Class 1 or Class 3)

How you resolve equidistance conflicts is up to you. General approaches include; randomly picking from the subset of conflict points, using the points with the most common class globally, resorting to KNN-1 or using KNN-(k+1).

Question 4
Using the training data $X = \{p_1, p_2, \ldots, p_{10}\}$ in Figure 2,

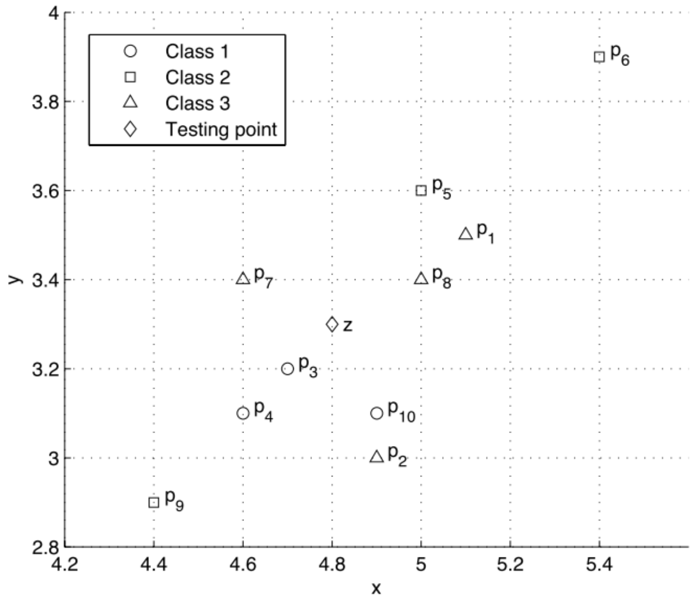
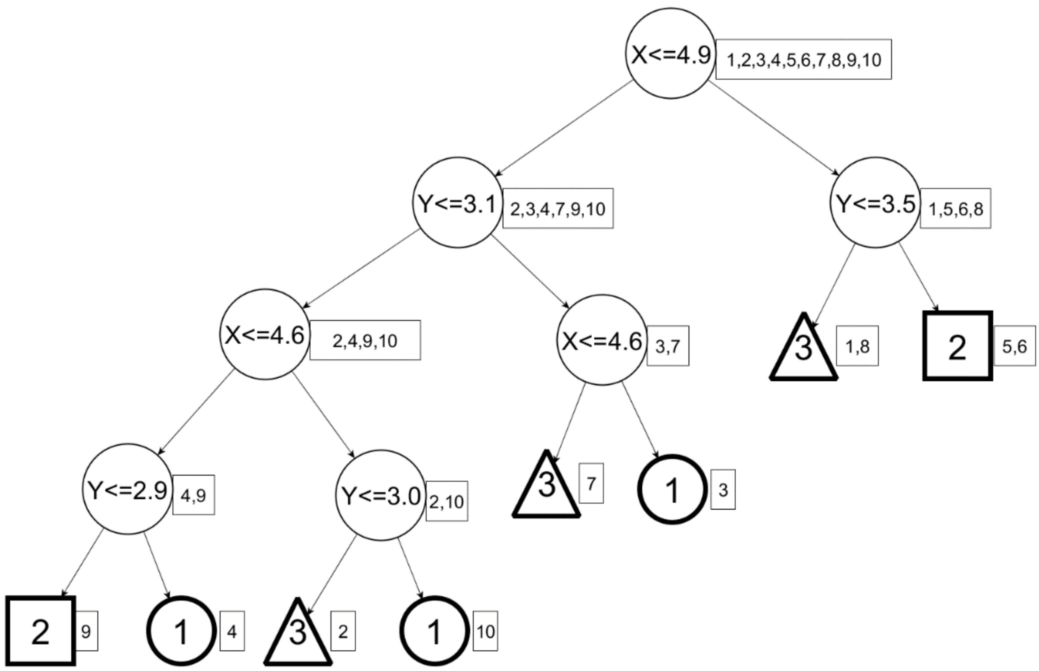


Figure 2: Scatter plot of points with class labels.

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	z
x	5.1	4.9	4.7	4.6	5.0	5.4	4.6	5.0	4.4	4.9	4.8
y	3.5	3.0	3.2	3.1	3.6	3.9	3.4	3.4	2.9	3.1	3.3

1. Construct a Kd-tree with a bucket-size of 1, i.e., each leaf node must contain at least 1 point. Indicate clearly the branching criterion at each node.



2. In Figure 2, show clearly the spatial partitioning corresponding to the con structed Kd-tree.

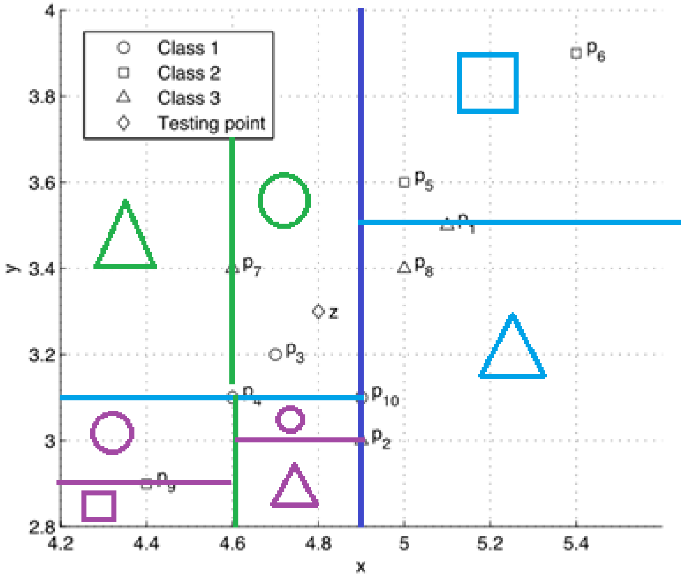
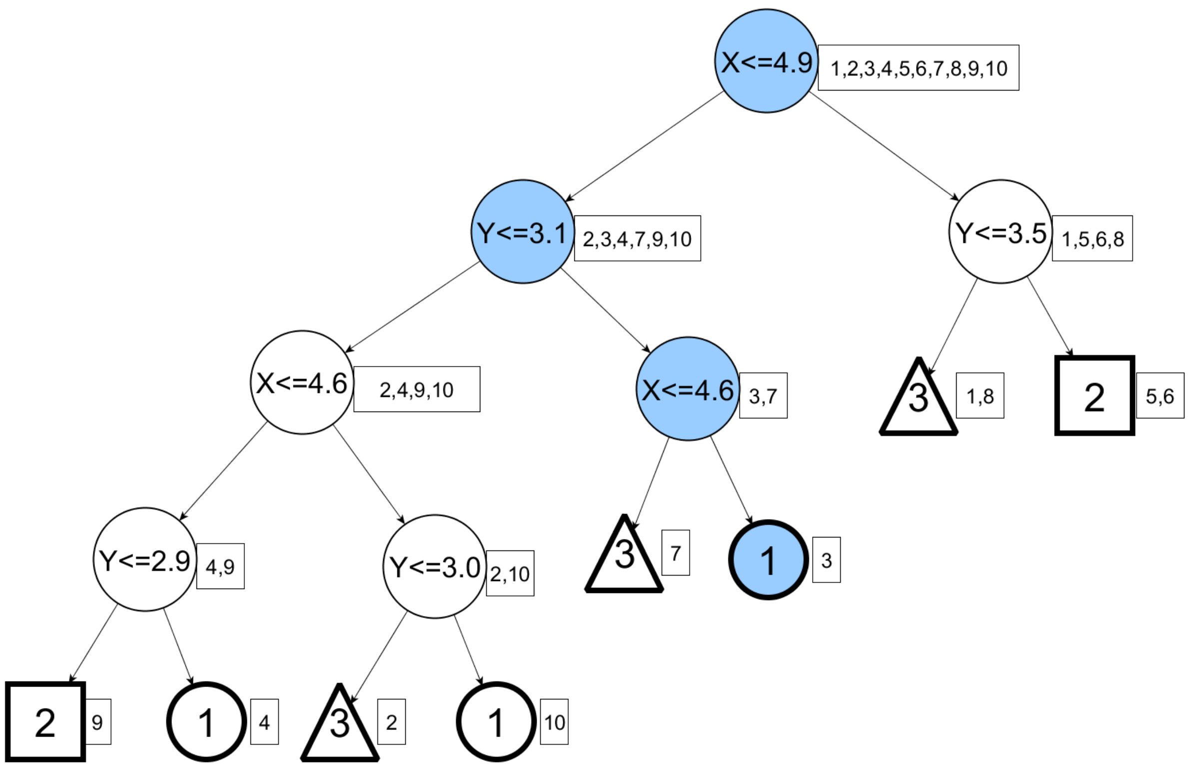


Figure 2: Scatter plot of points with class labels.

3. Search the Kd-tree for the nearest neighbour of testing point z, indicating clearly on the Kd-tree the nodes that are visited.



Question 5 (Question 18.15 in AIMA 3ed)
Suppose a 7-nearest-neighbors regression search returns $\{7, 6, 8, 4, 7, 11, 100\}$ as the 7 nearest y values for a given x value. What is the value of $\hat{y}$ (the predicted output value of x) that minimizes the L_1 loss function (sum of absolute errors) on this data? There is a common name in statistics for this value as a function of the y values; what is it? Answer the same two questions for the L_2 loss function (sum of squared errors).

L1 Loss: $\min_{y^*} \sum_{y \in Y} |y^* - y|$ this will find the *median* so,

$$y^* = 7$$

Extension: Verify your answer by calculating the loss, as well as the loss for the values to either side:

L1 Loss with $y^* = 7$: $Loss = \sum_{y \in Y} |7 - y| = 0 + 1 + 1 + 3 + 0 + 4 + 93 = 102$

L1 Loss with $y^* = 6$: $Loss = \sum_{y \in Y} |7 - y| = 1 + 0 + 2 + 2 + 1 + 5 + 94 = 105$

L1 Loss with $y^* = 8$: $Loss = \sum_{y \in Y} |7 - y| = 1 + 2 + 0 + 4 + 1 + 3 + 92 = 103$

L2 Loss: $\min_{y^*} \sum_{y \in Y} (y^* - y)^2$ this will find the *mean* so,

$$y^* = \frac{7 + 6 + 8 + 4 + 7 + 11 + 100}{7} \approx 20.43$$

Extension: Verify your answer by calculating the loss, as well as the loss for the values to either side:

L1 Loss with $y^* = 20.43$: $Loss = \sum_{y \in Y} (20.43 - y)^2 \approx 7414$

L1 Loss with $y^* = 20$: $Loss = \sum_{y \in Y} (20 - y)^2 = 1 + 0 + 2 + 2 + 1 + 5 + 94 = 7415$

L1 Loss with $y^* = 21$: $Loss = \sum_{y \in Y} (21 - y)^2 = 1 + 2 + 0 + 4 + 1 + 3 + 92 = 7416$

Question 6

A compilation of the playing conditions and outcomes of matches between tennis players Federa and Nadal is given in Table 1, where the time of the match is either Morning (M), Afternoon (A) or Night (N); the type of match is either Grand Slam (G), Master (M) or Friendly; the type of court is either Grass (G), Hard (H), Clay (C) or Mixed (M); and the outcome is a Federera (F) win or a Nadale (N) win.

Time	Match	Surface	Outcome	ID
M	M	G	F	1
A	G	C	F	2
N	F	H	F	3
A	F	M	N	4
A	M	C	N	5
A	G	G	F	6
A	G	H	F	7
A	G	H	F	8
M	M	G	F	9
A	G	C	N	10
N	F	H	F	11
N	M	M	N	12
A	M	C	N	13
A	M	G	F	14
A	G	H	F	15
A	G	C	F	16

Table 1: A record of previous tennis matches between Federa and Nadal.

Build a decision tree that can predict the outcome of a new match, given information about the time, type of match, and surface. Show all your working to demonstrate that you are correctly using information gain as the splitting criterion.

$$\mathrm{H}(\mathcal{S}) = -\sum p_i \log_2 p_i,$$

$$IG(T,a) = H(T) - \sum_{v \in vals(a)} \frac{|\{\mathbf{x} \in T | x_a = v\}|}{|T|} \cdot H(\{\mathbf{x} \in T | x_a = v\})$$

Depth 0:

$$H(Root) = -\left(\frac{11}{16}\log_2\frac{11}{16} + \frac{5}{16}\log_2\frac{5}{16}\right) \approx 0.896$$

Split time:

[[1,9],[2,4,5,6,7,8,10,13,14,15,16],[3,11,12]]

$$H(\boldsymbol{x} = M) = -(1\log_2 1 + 0) = 0$$

$$H(\boldsymbol{x} = A) = -\left(\frac{7}{11}\log_2\frac{7}{11} + \frac{4}{11}\log_2\frac{4}{11}\right) \approx 0.946$$

$$H(\boldsymbol{x} = N) = -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right) \approx 0.918$$

$$IG(Time) = 0.896 - \left(\frac{2}{16}0 + \frac{11}{16}0.946 + \frac{3}{16}0.918\right) \approx 0.074$$

Split Match:

[2,6,7,8,10,15,16],[1,5,9,12,13,14],[3,4,11]

$H(\boldsymbol{x}=G) = -\left(\frac{6}{7}\log_2\frac{6}{7}+\frac{1}{7}\log_2\frac{1}{7}\right) \approx 0.381$

$H(\boldsymbol{x}=M) = -\left(\frac{3}{6}\log_2\frac{3}{6}+\frac{3}{6}\log_2\frac{3}{6}\right) \approx 1.0$

$H(\boldsymbol{x}=F) = -\left(\frac{2}{3}\log_2\frac{2}{3}+\frac{1}{3}\log_2\frac{1}{3}\right) \approx 0.918$

$\mathrm{I}G(Match) = 0.896 - \left(\frac{7}{16}0.381+\frac{6}{16}1.0+\frac{3}{16}0.918\right) \approx 0.182$

Split Surface:

[1,6,9,14],[3,7,8,11,15],[2,5,10,13,16],[4,12]

$H(\boldsymbol{x}=G) = -(1\log_2 1+0) = 0$

$H(\boldsymbol{x}=H) = -(1\log_2 1+0) = 0$

$H(\boldsymbol{x}=C) = -\left(\frac{2}{5}\log_2\frac{2}{5}+\frac{3}{5}\log_2\frac{3}{5}\right) \approx 0.971$

$H(\boldsymbol{x}=M) = -(0+1\log_2 1) = 0.0$

$\mathrm{I}G(Surface) = 0.896 - \left(\frac{5}{16}0.971\right) \approx 0.593$

Choose Surface as it has the highest information gain

Depth 2:

All non-clay surface games are leaf nodes (entropy of zero). Need to split clay further (entropy is non-zero). The set is now [2,5,10,13,16].

$$H(Surface = Clay) = -\left(\frac{2}{5}\log_2\frac{2}{5}+\frac{3}{5}\log_2\frac{3}{5}\right) \approx 0.971$$

Split time:

[],[2,5,10,13,16],[]

Omitted as this split trivially gives us nothing

Split Match:

[2,10,16],[5,13],[]

$H(\boldsymbol{x}=G) = -\left(\frac{2}{3}\log_2\frac{2}{3}+\frac{1}{3}\log_2\frac{1}{3}\right) \approx 0.918$

$H(\boldsymbol{x}=M) = -(0+1\log_2 1) = 0.0$

$H(\boldsymbol{x}=F) = 0.0$

$\mathrm{I}G(Match) = 0.971 - \left(\frac{3}{5}0.918+0\right) \approx 0.420$

Note: there was a mistake in this part in the tutorial as I used the IG, instead of H.

Split Surface:

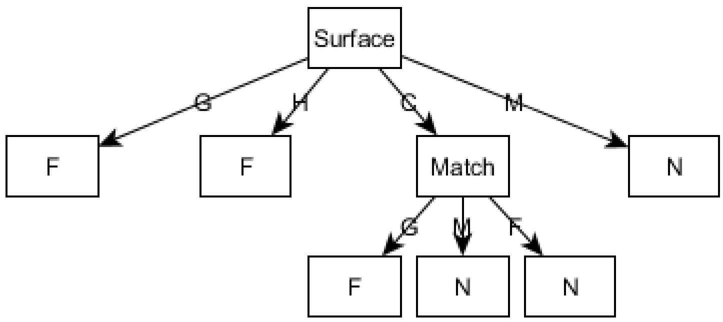
Discrete values, skip

Choose Match as it has the highest information gain

Depth 3:

Set is now [2,10,16], we cannot split this further so we assign the most probable value at this node.

Assign most probable outcome of parent to the node F as it has no elements.



I also mentioned two other useful equations:

$\log_a k = \frac{\log_b k}{\log_b a}$, eg. $\log_2 42 = \frac{\ln 42}{\ln 2} = 5.39$

And:

$H_b(S) = -\sum p_i \log_b p_i$, where b is the number of possible outcomes (2 for binary events).