1.

Consider the following matrix with 6 data points for two features X_1 and X_2 [By Ani Saxena]:

$$\text{Data} = \begin{bmatrix} 4 & 1 \\ 6 & 6 \\ 9 & 5 \\ 1 & 2 \\ 7 & 3 \\ 5 & 4 \end{bmatrix}$$

The labels for this data are: $\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}^T$

We wish to classify this dataset by building a decision tree of depth

a. What is the (cross)-entropy at the root of the tree?

b. What is the rule for the first split? Hint: you don't need to compute entropies for this, you should be able to eyeball the best split. You can write your answer in the form of $X_1 \geq 1$ or $X_2 \geq 3$.

c. What is the rule for the second split for each of the two nodes after the first split?

d. Imagine you are beginner in machine learning and don't know much about decision trees. Is there a (not trivial) decision tree of any depth that would give us an information gain (reduction in weighted entropy) of 0, i.e. the same cross-entropy as when we have no splits? Explain your answer.

2.

Consider the following dataset:

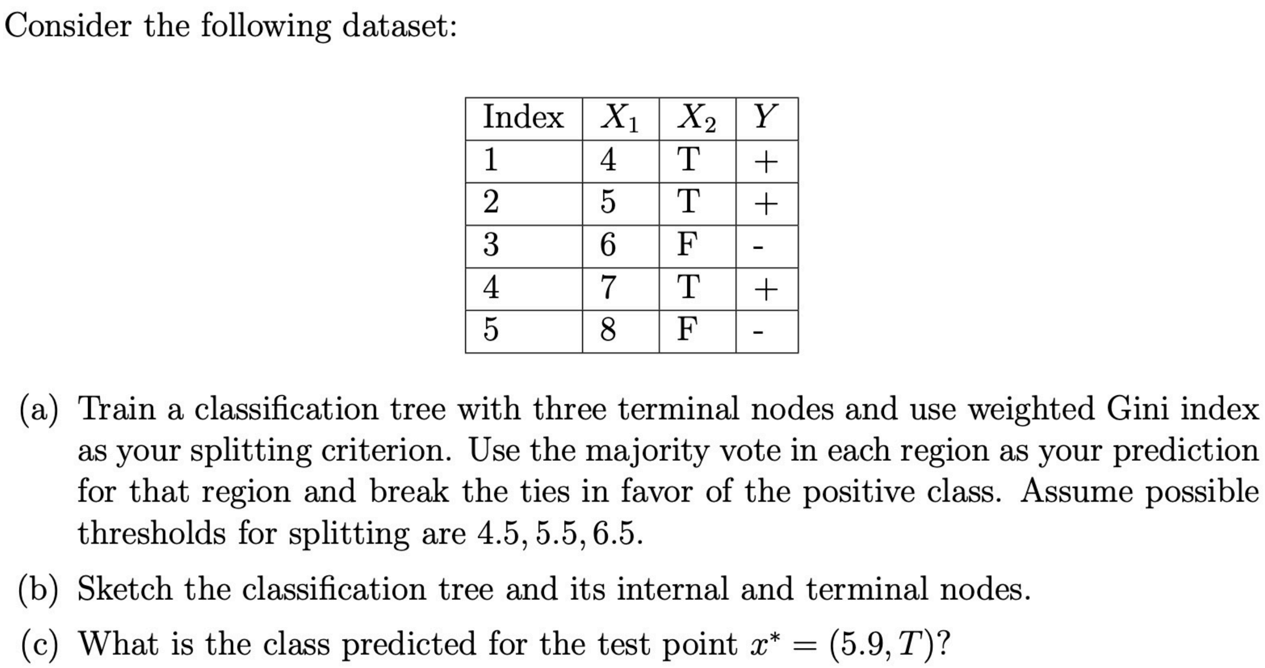
Index	X	Y
1	4	$+$
2	5	$+$
3	6	-
4	7	$+$
5	8	-

(a) Train a classification tree with three terminal nodes and use weighted Gini index as your splitting criterion. Use the majority vote in each region as your prediction for that region and break the ties in favor of the positive class. Assume possible thresholds for splitting are $4.5, 5.5, 6.5$.

(b) Sketch the classification tree and its internal and terminal nodes.

(c) What is the class predicted for the test point $x^* = 5.9$?

3.



Consider the following dataset:

Index	X_1	X_2	Y
1	4	T	+
2	5	T	+
3	6	F	-
4	7	T	+
5	8	F	-

(a) Train a classification tree with three terminal nodes and use weighted Gini index as your splitting criterion. Use the majority vote in each region as your prediction for that region and break the ties in favor of the positive class. Assume possible thresholds for splitting are $4.5, 5.5, 6.5$.

(b) Sketch the classification tree and its internal and terminal nodes.

(c) What is the class predicted for the test point $x^* = (5.9, T)$?

4.

Consider the dataset summarized in the following table with three binary attributes A, B, C, and two class labels (+, -). In this question, we will use the Gini index to create a decision tree with **two terminal nodes**.

A	B	C	Number of Instances (+, -)
T	T	T	5, 0
F	T	T	0, 20
T	F	T	20, 0
F	F	T	0, 5
T	T	F	0, 0
F	T	F	25, 0
T	F	F	0, 0
F	F	F	0, 25

(a) Calculate the Gini index for the overall collection of training examples before any splits.

b) For each attribute (A, B, and C), compute the Gini index for potential splits. Then calculate the Weighted Gini index after each split. Determine the Gini gain from splitting on each attribute. The Gini gain of a split is calculated by subtracting the Weighted Gini index after the split from the Gini index of the data without splits. Weights must be the fraction of the data in each region to all data points.

(c) Based on the Gini gain, select which attribute should be used to split the root node.

(d) Draw the decision tree and show the decision in each terminal node.

5.

Assume the following dataset with one feature and one quantitative (continuous) output, i.e. data are in the form (x_i, y_i):

$\{(0,1),(1,2),(2,-1),(3,2),(4,0),(5,4)\}$.

Build a decision tree with three leaves using recursive binary splitting and the Sum of Absolute Errors criterion whose formula for region R_m is $\sum_{j=1}^{|R_m|}|y_j - \hat{y}_{R_m}|$, where $\hat{y}_{R_m}$ is the average response in region R_m, and $|R_m|$ is the number of data points in region R_m. Assume that the feature space can only be splitted at 1.5, 2.5, and 3.5. Sketch the diagram of the decision tree.

6.

Assume the dataset with one feature and three classes:

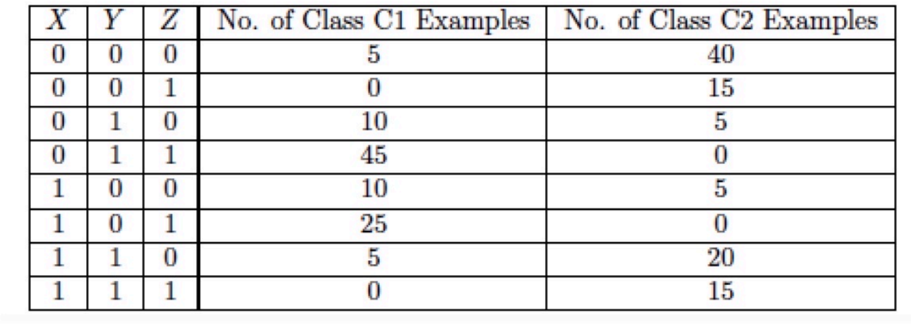
$\{(0, green), (1, green), (2, blue), (3, green), (4, red), (5, blue)\}$.

Build a decision tree with three leaves using recursive binary splitting and the Gini index whose formula for region R_m is $\sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$, where $\hat{p}_{mk}$ is the etsimated probability of class k in region R_m, where $k = 1, 2, \cdots, K$ and K is the number of classes. Assume that the feature space can only be splitted at 1.5, 2.5, and 3.5. Sketch the diagram of the decision tree. Do NOT use weighted Gini index over the regions. Simply add the Gini indices of the regions of the decision tree.

Hint: How to break possible ties? If you ever encountered ties, consider $green = 100, blue = 200, red = 300$. Break the tie in favor of the smaller number. So the tie between green and red would be broken in favor of green.

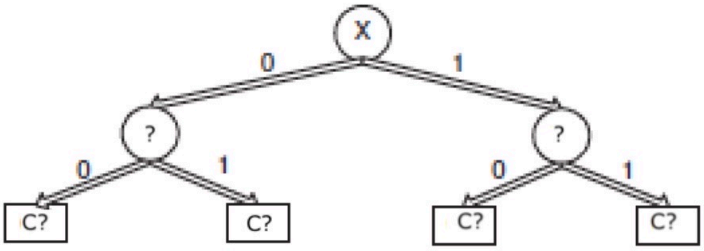
7.

Consider the following set of training examples.



X	Y	Z	No. of Class C1 Examples	No. of Class C2 Examples
0	0	0	5	40
0	0	1	0	15
0	1	0	10	5
0	1	1	45	0
1	0	0	10	5
1	0	1	25	0
1	1	0	5	20
1	1	1	0	15

Compute a two-level decision tree, and choose X as the root, as shown in the figure.



Use the classification error rate as the criterion for splitting at the second level and determine whether Y or Z should be used in the second level. To do so, fill in the following tables, assuming that $X = 0$

Y	C1	C2
0		
1		

Z	C1	C2
0		
1		

and the following tables, assuming $X = 1$

Y	C1	C2
0		
1		

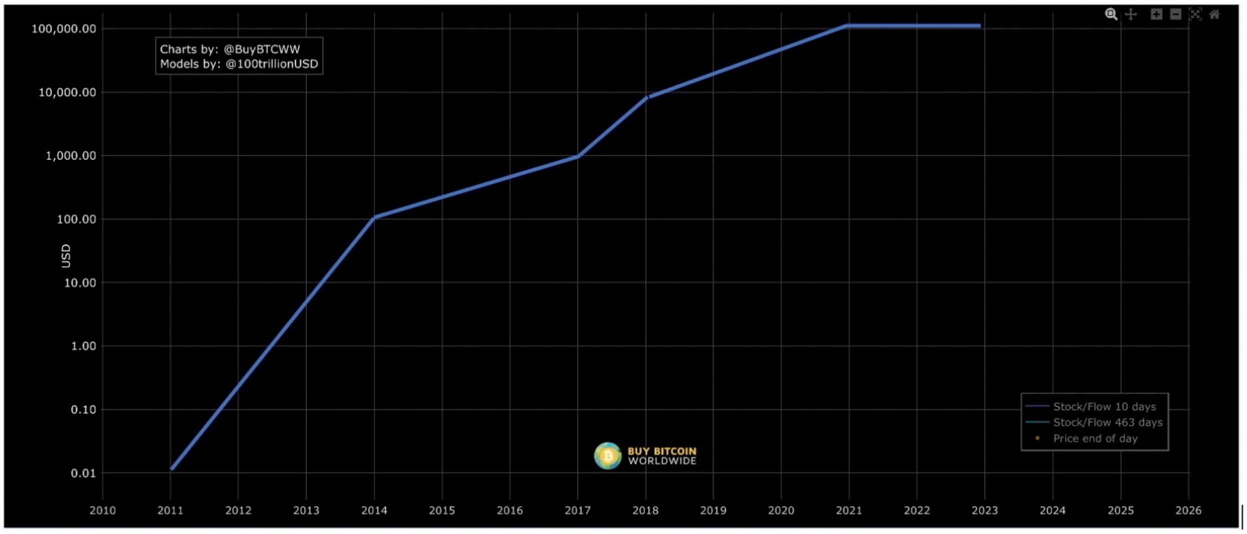
Z	C1	C2
0		
1		

The above tables assist in finding the variable splitting which provides better classification error rate for each side of the tree, i.e. for $X = 0$ and $X = 1$. Determine the classes in the leaf nodes. What is the classification error rate of the tree you found on training data?

Y

8.

The bitcoin Stock to Flow (S2F) model was created by the famous twitter user PlanB. S2F models the price of bitcoin based on its rarity. A slightly modified version of S2F is shown below. Note that the dependent variable y is $\log_{10} price$ and the independent variable is time in years since 2010 t; therefore $t \in [1, 13]$. Show this model using a decision tree and clearly determine the internal nodes and terminal nodes. Remember that this is a *model* tree, so the terminal nodes may contain *regression models.*



9.

Draw the decision trees that calculate the following Boolean functions:

(a) $Y = (X_1 \vee X_2) \wedge (\neg X_3 \wedge X_4),\ \ X_i \in \{0,1\}$. The root of the decision tree must be X_3.

(b) $Y = X_1 \otimes X_2,\ \ X_i \in \{0,1\}$, where $\otimes$ is the exclusive or (XOR) function.