CS 5805: Machine Learning I

Spring 2025

Homework 3 Solution Sketches

by Quyet Do

Q1 (5x2=10 points)

Point Distribution:

For each part, a correct answer and reasonable explanation are needed to get 2 points. Otherwise, 0 points.

1. **False.** If we had two rows with the same feature values and different classes (this will happen with real data), the entropy cannot be brought down to zero.

2. **False.** ID3 uses a greedy approach, selecting the best attribute at each step. This does not guarantee finding the globally optimal decision tree, as it may get stuck in local optima.

3. **False.** Feature selection aims to choose features that *do* provide discrimination about the class to help the algorithm learn effectively. Although it is important to mitigate bias and overfitting, eliminating features that provide discrimination is not the way to accomplish this goal.

4. **False.** Continuing to grow a decision tree until all instances are correctly classified will lead to severe overfitting. The tree will perfectly fit the training data but will likely perform poorly on unseen data. Pruning is necessary to prevent this.

5. **False.** In a single decision tree, each point in the feature space must follow exactly one path to one leaf. However, using rule pruning, we're able to prune different rules in different branches. By doing this, we will create a situation where there are two paths to the same point, potentially assigning different class labels. This cannot be supported by a pruned decision tree.

Q2 (10+10=20 points)

Point Distribution:

a) 10 points

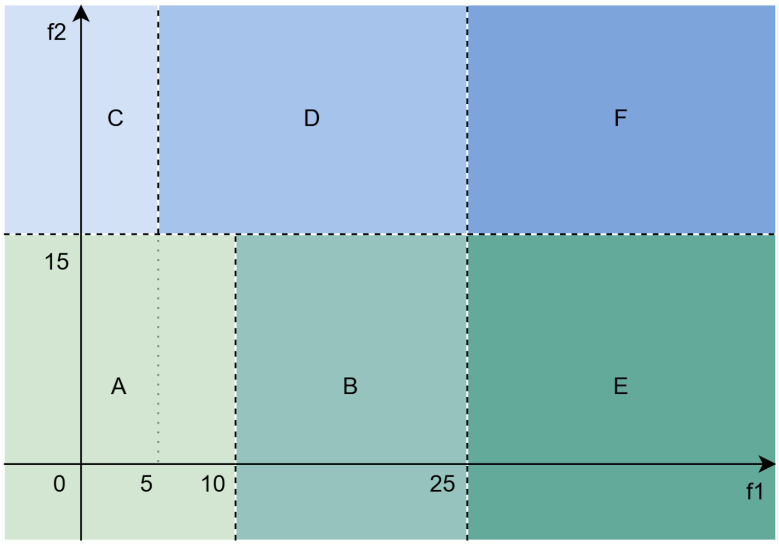
Criteria	Points	Description
Axis	2	Properly annotate the x and y axes
Decision boundaries	5	Correctly divide the plane into regions
Class label	3	Correctly label regions in the 2D plane

b) 10 points

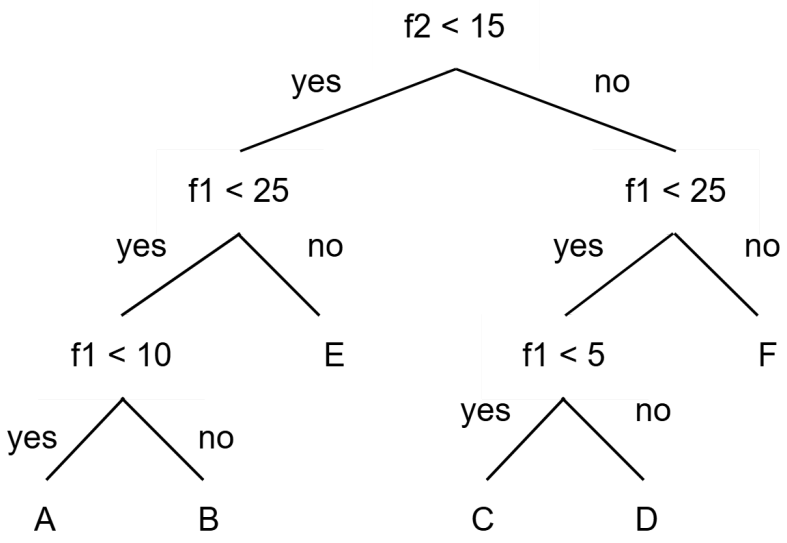
The new tree should have around n_nodes=5 nodes.

Criteria	Points	Description
Node	10/n_nodes	Correctly split the node, with correct condition and proper branch annotation (yes/no)

a) The decision boundaries defined by the tree on a 2D plane are as below.



b) Another decision tree that is syntactically different but defines the same decision boundaries. (**Note that there are many other solutions**)



Q3 (15 points)

Point Distribution:

a) Root node: 11 points

Criteria	Points	Description
Entropy	10/5	Show correct calculation (either step-by-step or composition) of Entropy of dataset and Information Gain of 4 attributes. 5 calculations in total
Splitting attribute	1	Correctly determine the splitting attribute as f3

b) Full tree: 4 points. 1 point each split.

Calculations for the Root Node (11 points)

Entropy of the target variable (class) before any splits:
Count: c1 (4), c2 (4)
P(c1) = 4/8 = 0.5
P(c2) = 4/8 = 0.5
Entropy(class) = - (0.5 * log2(0.5) + 0.5 * log2(0.5)) = -(-0.5 + -0.5) = 1

Information gain for each feature to determine the best root split.
f1:

f1 = a: c1 (2), c2 (2)
f1 = b: c1 (2), c2 (2)
Entropy(Class|f1=a) = -(2/4 * log2(2/4) + 2/4 * log2(2/4)) = 1
Entropy(Class|f1=b) = -(2/4 * log2(2/4) + 2/4 * log2(2/4)) = 1
Entropy(Class|f1) = (4/8) * 1 + (4/8) * 1 = 1
Information Gain(f1) = 1 - 1 = 0

f2:

f2 = 0: c1 (2), c2 (2)
f2 = 1: c1 (2), c2 (2)
Entropy(Class|f2=0) = -(2/4 * log2(2/4) + 2/4 * log2(2/4)) = 1
Entropy(Class|f2=1) = -(2/4 * log2(2/4) + 2/4 * log2(2/4)) = 1
Entropy(Class|f2) = (4/8) * 1 + (4/8) * 1 = 1
Information Gain(f2) = 1 - 1 = 0

f3:

f3 = a: c1 (0), c1 (2)
f3 = c: c1 (4), c2 (2)
Entropy(Class|f3=a) = -(0/2 * log2(0) + 2/2 * log2(1)) = 0.
Entropy(Class|f3=c) = -(4/6 * log2(4/6) + 2/6 * log2(2/6)) = 0.918
Entropy(Class|f3) = (2/8) * 0 + (6/8) * 0.918 = 0.689
Information Gain(f3) = 1 - 0.689 = 0.311

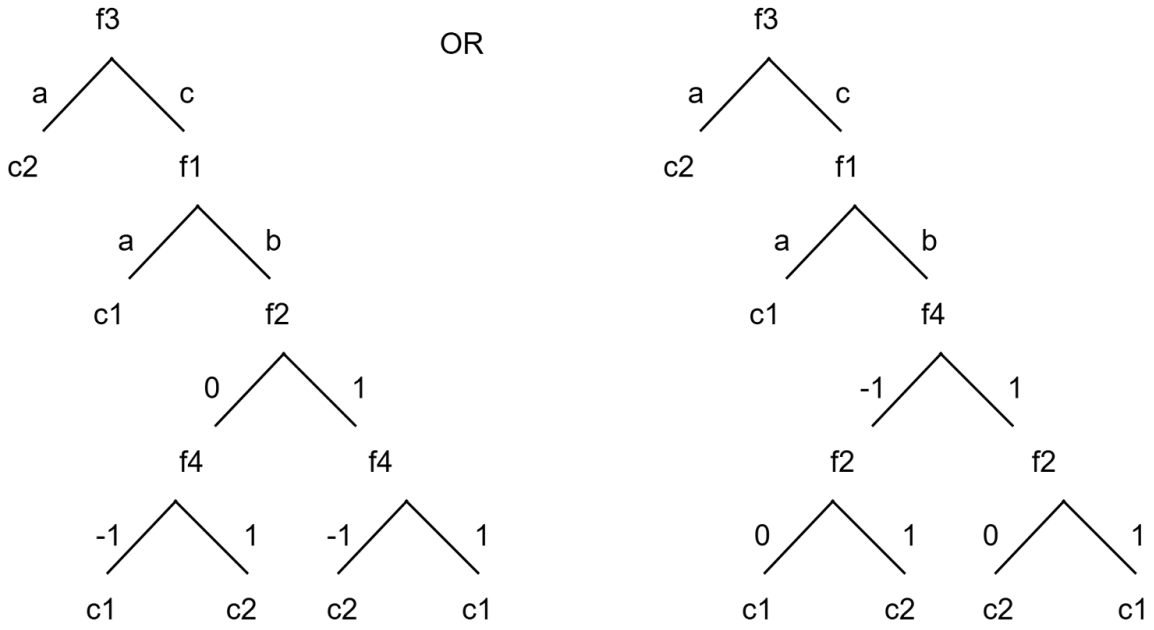
f4:

f4 = -1: c1 (2), c1 (3)
f4 = 1: c1 (2), c1 (1)

Entropy(Class|f4=a) = -(2/5 * log2(2/5) + 2/3 * log2(3/5)) = 0.971
Entropy(Class|f4=c) = -(2/3 * log2(2/3) + 1/3 * log2(1/3)) = 0.918
Entropy(Class|f4) = (5/8) * 0.971 + (3/8) * 0.918 = 0.952
Information Gain(f4) = 1 - 0.952 = 0.048

f3 has the largest information gain, so it will be selected as the splitting attribute.

The full tree (4 points)



Note: In the last two splits, f2 and f4 are interchangeable because they have the same information gain.

Q4, Q5, Q6: CS5805-HW3-Code