Casey Pei
*CSCE 421-500*
*Texas A&M University*

**February 27, 2024**

**HOMEWORK 2 — Tree-based Models**

# 1 Math Questions

## 1.1 Information Gain (20 points)

NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.
Suppose you are given 6 training points as seen below, for a classification problem with two binary attributes $X_1$ and $X_2$ and three classes $Y \in 1, 2, 3$. You will use a decision tree learner based on information gain

1. Calculate the conditional entropy for both $X_1$ and $X_2$.

$$\nabla f(x, y) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \tag{0-1}$$

2. Calculate the information gain if we split based on 1) $X_1$ or 2) $X_2$

$$\nabla f(x, y, z) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) \tag{0-2}$$

3. Report which attribute is used for the first split. Draw the decision tree using this split.

$$\nabla f(x, y, z) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) \tag{0-3}$$

4. Conduct classification for the test example $X_1 = 0$ and $X_2 = 1$.

$$\nabla f(x, y, z) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) \tag{0-4}$$

| $X_1$ | $X_2$ | $Y$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 2 |
| 1 | 0 | 3 |
| 0 | 0 | 2 |
| 0 | 0 | 3 |

# 2 Programming Questions

Answers for these are located in the attached Casey_Pei_HW2.ipynb file.

*Submitted by Casey Pei on February 27, 2024.*