CS461 Spring 2023 Machine Learning Principles

Vinay Varma*

February 1, 2023

Keep your explanations succinct and to the point. Answer the notebook questions in the cell right below the question.

Homework 1

Age	Gender	Car_Ownership	Will_Buy_Car
50	M	Yes	Yes
50	M	Yes	No
50	M	Yes	Yes
30	M	No	No
30	M	Yes	Yes
10	M	No	No
10	M	No	No
10	M	Yes	No

Question 1 (30 Points)

1. At root node, what is the Information Gain if we split by 'Age', 'Gender', and 'Car_Ownership' respectively? (10 points)

2. Which column is used to make the first split? (10 points)

3. Once trained, What will be the output on a new sample with Age = 50, Car_Ownership = No, Gender = M ? Can you train the same tree without the Gender column, why or why not? (10 points)

Question 1 (70 Points)

Complete the jupyter notebook (provide all your answers inside it)

*Department of Computer Science, Rutgers University, vk405@scarletmail.rutgers.edu.