Assignment 1 Decision Trees and Naïve Bayes (75 points)

CSCE 415: Machine Learning
Spring 2021

Due: 5 February 2021 11:30 PM

Submission Instructions: Turn in this assignment in a zip or tar file through Blackboard. The zip or tar file should be of the form *last_name_only.zip* Each problem should be in a separate folder before zipping/tarring the two folders.

Problem 1 (25 points): Decision Trees

Using the ID3 algorithm, build a decision tree to predict the habitability of planets based on size and orbit (Martin Azizyan). The final result should be decision tree <u>showing your calculations</u> for each internal node, showing why you selected the node. Show the decision tree with labels and show calculations. Make sure your labels and calculations are readable.

Size	Orbit	Habitable	Count
big	near	yes	20
big	far	yes	170
small	near	yes	139
small	far	yes	45
big	near	no	130
big	far	no	30
small	near	no	11
small	far	no	255

Problem 2 (50 points): Sloan Digital Sky Survey

For this problem you have file containing 10,000 records from the Sloan Digital Sky Survey. These records consist of 17 feature columns and a target column. Using Jupyter Notebook with python3, classify the observations (according to the Class field). The goal is to develop a classifier that can distinguish between different astronomical observations. Use the process

described in class and in your book to build the "best" decision tree. Make sure you write up your analysis. Turn in your Jupiter notebook file and your write up. Do Not Turn In The Data! If you change the data in anyway, it must be through your Jupiter Notebook file. Don't forget to include your visualizations.

Please turn in your files showing your work in addition to your write up in Blackboard. Again, put the two problems in different folders, zip or tar them.