CIS5200: Machine Learning Fall 2023

Homework 3

Release Date: November 1, 2023 *Due Date: November 15, 2023*

- HW3 will count for 8% of the grade. This grade will be split between the written (8 points) and programming parts. Note that this will be shorter than previous homework, hence the shorter time span and lower total grade.
- Homework solutions may be formatted using either LaTeXor handwritten, but they **must** be **unambiguously legible** if handwritten.
- You will submit your solution for the written part of HW3 as a single PDF file via Gradescope. The deadline is **11:59 PM ET**. Contact TAs on Ed if you face any issues uploading your homeworks.
- Collaboration is permitted and encouraged for this homework, though each student must understand, write, and hand in their own submission. In particular, it is acceptable for students to discuss problems with each other; it is not acceptable for students to look at another student's written Solutions when writing their own. It is also not acceptable to publicly post your (partial) solution on Ed, but you are encouraged to ask public questions on Ed. If you choose to collaborate, you must indicate on each homework with whom you collaborated.

Please refer to the notes and slides posted on the website if you need to recall the material discussed in the lectures.

1 Written Questions (8 points)

Problem 1: Decision Trees (8 points)

Recall that when building a decision tree for classification problems, we used the notion of node purity instead of classification error when deciding how to grow the decision tree. Consider 3-dimensional binary data $x \in \mathbb{R}^3$ where the label $y \in \{-1, 1\}$ is generated from the first two features as $f(x) = x_1 \wedge \neg x_2$. In other words, the entire dataset of all possible examples can be written as:

x_1	x_2	x_3	y
0	0	0	-1
0	0	1	-1
0	1	0	-1
0	1	1	-1
1	0	0	1
1	0	1	1
1	1	0	-1
1	1	1	-1

1.1 (1 point) What is the best possible classification error that a 1-leaf decision tree (i.e. only the root node predicting a constant value) can achieve?

1.2 (2 points) Consider all three possible splits for growing to a 2-leaf decision tree. Does there exist a split that achieves lower classification error than a 1-leaf decision tree? If so, what is it? If not, why is that?

1.3 (1 point) What is the entropy and Gini index of the 1-leaf decision tree?

1.4 (2 point) Consider all three possible splits for growing to a 2-leaf decision tree. Does there exist a split that achieves lower entropy than a 1-leaf decision tree? Find the split that results in the lowest possible entropy, and calculate the corresponding entropy of the split.

1.5 (2 point) Consider all three possible splits for growing to a 2-leaf decision tree. Does there exist a split that achieves lower Gini index than a 1-leaf decision tree? Find the split that results in the lowest possible Gini index, and calculate the corresponding Gini index of the split.

2 Programming Questions (15 points)

Use the link here to access the Google Colaboratory (Colab) file for this homework. Be sure to make a copy by going to "File", and "Save a copy in Drive". As with the previous homeworks, this assignment uses the PennGrader system for students to receive immediate feedback. As noted on the notebook, please be sure to change the student ID from the default '99999999' to your 8-digit PennID.

Instructions for how to submit the programming component of HW 3 to Gradescope are included in the Colab notebook. You may find this PyTorch linear algebra reference and this general PyTorch reference to be helpful in perusing the documentation and finding useful functions for your implementation.