COSC 410 S24 Lab 3: Nearest Neighbor and Classification Competition

This Lab is due by 11PM EST on the day of the lab (Feb 12/13 depending on your lab). You should work in groups of two or three for this lab.

Introduction

The goal of this lab, and the associated HW2, is to strengthen your understanding of k Nearest Neighbor classification and to expand on your familarity with a standard machine learning pipeline. First you'll implement a version of k-NN classification. Then, you'll explore a weather prediction data set. Finally, you'll implement a function for fitting a classification model of your choosing on this dataset. By completing the lab and the HW, you will demonstrate that you understand

- how to implement k-NN classification
- how to preprocess data in a standard ML pipeline
- how to select a good model for your task
- the ability to hyperparameter tune
- how to work with leaderboards

Save the last 15 minutes of lab to look over HW2. You should be able to answer the following questions before you leave.

- What is the difference between `HW2.ipynb` and `HW2.py`?
- What is your task?
- What is the leaderboard?
- How is HW2 graded?

Provided Files

- `Lab3.pdf`: This file
- `Lab3.ipynb`: Code scaffold for the lab
- `Lab3_train.csv`: Training examples with labels
- `Lab3_valid.csv`: Validation examples with labels
- `HW2.pdf`: Homework 2 instructions and details
- `HW2.ipynb`: Code scaffold for the homework and open-ended questions
- `HW2.py`: Code scaffold that you will need for the autograder

Good luck :)