CSE 417T: Homework 1
Due: Friday, September 13th, 2019 by 9 PM

Notes:

- Please check the submission instructions for Gradescope provided on the course website. You must follow those instructions exactly.

- Please complete and submit the following two Matlab stub files for Problem 2:
  - https://classes.cec.wustl.edu/~SEAS-SVC-CSE417T/perceptron_experiment.m
  - https://classes.cec.wustl.edu/~SEAS-SVC-CSE417T/perceptron_learn.m

- Homework is due by 9 PM on the due date. Remember that you may not use more than 2 late days on any one homework and you only have a budget of 5 in total.

- Please keep in mind the collaboration policy as specified in the course syllabus. If you discuss questions with others you must write their names on your submission and if you use any outside resources you must reference them. Do not look at each others’ writeups, including code.

- There are 6 problems and 1 bonus problem on 2 pages in this homework.

- Keep in mind that problems and exercises are distinct in LFD.

- All graphs should have clearly labeled axes. The Matlab hist function should be useful.

Problems:

1. (15 points) LFD Problem 1.3

2. (25 points) Consider the following experiment on perceptron learning for random training sets of dimension 10:
   - Generate an 11-dimensional weight vector $w^*$, where the first dimension is 0 and the other 10 dimensions are sampled independently at random from the uniform (0, 1) distribution (the first dimension will serve as the threshold and we set it to 0 for convenience).
   - Generate a random training set with 100 examples, where each dimension of each training example is sampled independently at random from the uniform (-1, 1) distribution. The examples are all classified in accordance with $w^*$.
   - Run the perceptron learning algorithm, starting with a zero weight vector, on the training set you just generated. Keep track of the number of iterations it takes to learn a hypothesis that correctly separates the training data.

Write code in Matlab to perform the above experiment and then repeat it 1000 times (note that you’re generating a new $w^*$ and a new training set each time). You should complete the two stub files linked above (perceptron_experiment.m and perceptron_learn.m) for this purpose. The files have comments that explain their inputs and outputs. Submit your final versions along with your homework writeup in Gradescope (there will be a separate submission for your code).
Once you have your code working:

(a) Plot a histogram of the number of iterations the algorithm takes to learn a linear separator; you should submit this with your writeup.

(b) How does the number of iterations compare with the bound on the number of iterations derived in part (e) of the previous problem? Note that this bound will be different for each instantiation of $w^*$ and training set so in order to answer this question, you will need to analyze the distribution of differences between the bound and the number of iterations. Plot and submit a histogram of the log of this difference. Discuss your interpretation of these results.

3. (15 points) LFD Problem 1.7

4. (15 points) LFD Problem 1.8

5. (15 points) LFD Problem 1.12

6. (15 points) LFD Problem 2.3. You do not need to compute the VC dimension, $d_{VC}$, for the provided hypothesis sets.

7. (Bonus 5 points) Write a multiple choice question related to the content of this homework (LFD Chapter 1 and Section 2.1). Be sure to indicate the correct answer! See Lecture 2, Slides 3 and 4 for the rubric and guidelines to writing a good multiple choice question. If you write a great question, there’s a chance it will be included on the next exam!