CSE 417T: Homework 2

Due: Thursday, October 22nd, 2020 at 11:59pm

Notes:

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

- Homework is due by 11:59pm 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 1 page in this homework.

  - **Problem 6 should only be attempted after Lecture #9: GD and SGD, nonlinear transformations given on October 15th, 2020**

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

Problems:

1. (10 points) LFD Problem 2.8

2. (15 points) The following is a lemma from Hoeffding for sampling without replacement:

   Lemma A.5. (Hoeffding, 1963). Let $\mathcal{A} = \{a_1, \cdots, a_{2N}\}$ be a set of values with $a_n \in [0, 1]$, and $\mu = \frac{1}{2N} \sum_{n=1}^{2N} a_n$ be their mean. Let $\mathcal{D} = \{z_1, \cdots, z_N\}$ be a sample of size N, sampled from $\mathcal{A}$ uniformly *without* replacement. Then

   $$P[|\frac{1}{N} \sum_{n=1}^{N} z_n - \mu| > \epsilon] \leq 2e^{-2\epsilon^2 N}$$

   Use this lemma A.5 to prove the following:

   Lemma A.4. For any h and any S,

   $$P[|E_{in}(h) - E'_{in}(h)| > \frac{\epsilon}{2} \mid S] \leq 2e^{-\frac{1}{8}\epsilon^2 N}$$

   Tip: You may find it useful to: (1) express E_{in} and E'_{in} in terms of the values in $\mathcal{A}$, and (2) Express μ in terms of E_{in} and E'_{in}.

3. (10 points) LFD Problem 2.22

4. (35 points) LFD Problem 2.24; for part (c), there is no need to submit the code you wrote to run the experiment.

5. (15 points) LFD Exercise 3.4 parts (a)-(c) only.

6. (15 points) LFD Problem 3.4; for part (c), you will need to read LFD Problem 1.5; you do not need to complete LFD Problem 1.5.

7. (Bonus 5 points) Write a multiple choice question related to the content of this homework (VC dim, BV tradeoff, linear/logistic regression, GD/SGD, nonlinear transformations). Be sure to indicate the correct answer! If you write a great question, there's a chance it will be included on the next exam!