EECS 440: Machine Learning Fall 2020 Written Problems Week 3 due 9/25 11:59pm

General Instructions: Write or type your answers neatly and remember to show all relevant work. All questions are worth 10 points. Each answer should be a separate pdf, and you can turn in the pdfs on canvas in the appropriate assignment. Some questions may be very challenging; significant partial credit is available for reasonable attempts at solutions. Since each question is worth the same number of points, do not waste too much time on any one. Ask me or the TAs for help if stuck.

Some of the questions require you to write short programs to simulate things. You can use any language/software to do this, and you do not need to turn in your code.

Upload your answers to Canvas as a pdf file by 11:59pm on the due date specified after the question. You will receive a 10% bonus for a solution turned in a week or more in advance of the due date. You can use one late day each week (up to Saturday 11:59pm) with a penalty of 20%. Submissions after Saturday 11:59pm for any week will not be graded.

Each group must do their own work. Only one submission is needed from each group. Do not use any source other than the lecture notes, textbook(s) and readings on the class website to answer these questions. Only those who contributed equally to a submission should have their names and Case IDs on the submission. Those not listed as contributing will not receive points.

10. Under what circumstances might it be *beneficial* to overfit?
11. Restriction biases of learning algorithms prevent overfitting by restricting the hypothesis space, while preference biases prevent overfitting by preferring simpler concepts but not necessarily restricting the hypothesis space. Discuss the pros and cons of preference vs restriction biases.
12. Person *X* wishes to evaluate the performance of a learning algorithm on a set of *n* examples. *X* employs the following strategy: Divide the *n* examples randomly into two equal-sized disjoint sets, A and B. Then train the algorithm on A and evaluate it on B. Repeat the previous two steps for *N* iterations (*N* large), then average the *N* performance measures obtained. Is this sound empirical methodology? Explain why or why not.
13. Prove that an ROC graph must be monotonically increasing.
14. Prove that the ROC graph of a random classifier that ignores attributes and guesses each class with equal probability must be a diagonal line.