

EECS 440: Machine Learning Fall 2020 Written Problems Week 1 due 9/11 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.

1. For three random variables A , B and C , show with a counterexample that the statement “ A is independent of B ” does not imply the statement “ A is independent of B given C .” (9/11)
2. Points are sampled uniformly at random from the interval $(0,1)^2$ so that they lie on the line $x+y=1$. Determine the expected squared distance between any two sampled points. (9/11)
3. Describe two learning tasks that might be suitable for machine learning approaches. For each task, write down the goal, a possible performance measure, what examples you might get and what a suitable hypothesis space might be. What learning setting (supervised, unsupervised, etc.) seems most appropriate for each task? What example representation seems most appropriate? Be original---don't write about tasks discussed in class or described in the texts. Preferably select tasks from your research area. Describe any aspect of the task(s) that may not fit well with the learning settings and representations we have discussed. (9/11)
4. Explain in your own words: (i) why memorization should not be considered a valid learning approach, (ii) why tabula rasa learning is impossible, and (iii) why picking a good example representation is important for learning. Try to use good, intuitive examples from human learning to motivate your arguments. (9/11)