EECS 440: Machine Learning Fall 2020 Written Problems Week 10 and 11

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.

37. Consider a regression problem with examples described by 2 continuous attributes, x and y. Each example is sampled according to the uniform distribution on $(-1,1)^2$ and labeled with $f(x,y)=1-x^2-y^2$. A learner's hypothesis class is $h(x,y)=ax+by+c$. Calculate its bias and variance as a function of x and y if the learner sees an arbitrarily large training sample. (11/13)
38. Using MATLAB/python/R etc, find the (x,y) with the largest bias and the (x,y) with the largest variance for samples of size 10 in the previous setup. Can you intuitively justify your findings? (11/13)
39. Show that the VC dimension of the concept class of m-dimensional hyperplanes is at least m. A concept in this class is a hyperplane $f(\mathbf{x})=\mathbf{w}\cdot\mathbf{x}+b$. It classifies an example as positive if $f(\mathbf{x})\geq 0$. (11/20)
40. Show that the concept class of 3CNF Boolean formulae is efficiently PAC learnable. A 3CNF formula is a conjunction where each conjunct is a disjunction of at most 3 literals. (11/20)

All done! ☺