

CSE 417T: Homework 6

Due: December 1st (Saturday), 2018, **2pm**

Notes:

- Please check the submission instructions for Gradescope provided on the course website. You must follow those instructions exactly.
- Please note that this homework is due **by 2 PM** on the due date. Please plan your schedule accordingly. The reason is (given the two-day rule for late days) that, if needed, we can safely discuss the homework in the section on December 3rd (Monday) before the exam.
- 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.**
- Please do not directly post your answers on Piazza even if you think they might be wrong. Please try to frame the question such that you don't give the answers away. If there is specific information you want to ask about your answers, try the office hours or private posts on Piazza.
- There are 5 problems on 2 pages in this homework.

Problems:

1. (25 points) You have been hired by a biologist to learn a decision tree to determine whether a mushroom is poisonous. You have been given the following data:

Color	Stripes	Texture	Poisonous?
Purple	No	Smooth	No
Purple	No	Rough	No
Red	Yes	Smooth	No
Purple	Yes	Rough	Yes
Purple	Yes	Smooth	Yes

Use ID3 to learn a decision tree from the data (this is a written exercise – no need to code it up):

- (a) What is the root attribute of the tree? Show the computations.

- (b) Draw the decision tree obtained using ID3.
2. (15 points) Suppose your input data consists of the following (x, y) pairs:

$(3, 5); (5, 6); (7, 9); (2, 11); (3, 8)$

What value of y would you predict for a test example where $x = 3.2$ using the 3-nearest neighbors regression?

3. (20 points) (From Russell & Norvig) Construct a support vector machine that computes the XOR function. $\mathbf{x} = [x_1, x_2]$ denotes the two inputs and y denotes the output. Instead of using 1 and 0 to represent boolean variables, use +1 and -1 in this question (so the notations are consistent with our lectures). Map the input $[x_1, x_2]$ into a space consisting of x_1 and x_1x_2 . Draw the four input points in this space, and the maximal margin separator. What is the margin? Now draw the separating line back in the original Euclidean input space.
4. (20 points) The key point of the so-called “kernel trick” in SVMs is to learn a classifier that effectively separates the training data in a higher dimensional space without having to explicitly compute the representation $\Phi(\mathbf{x})$ of every point \mathbf{x} in the original input space. Instead, all the work is done through the kernel function that computes dot products $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)$.

Show how to compute the squared Euclidean distance in the projected space between any two points $\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)$ without explicitly computing the Φ mapping. Instead, write down the squared Euclidean distance using the kernel function K .

5. (20 points) Create a neural network with only one hidden layer (of any number of units) that implements XOR(AND(x_1, x_2), x_3). Draw your network, and show all weights of each unit.