

# CSE 417T: Homework 5

Due: April 19 (Sunday), 2020, 11:30AM

## Notes:

- Please submit your homework via Gradescope and check the [submission instructions](#).
- Please note that this homework is due **by 11:30 AM** on the due date. Please plan your schedule accordingly. The reason for the 11:30am deadline is that (given the two-day rule for late days), if needed, we can safely discuss the homework in the lecture on April 21st (Tuesday) before the exam.
- Remember that you may not use more than 2 late days on this homework, and you only have a budget of 7 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 6 problems on 2 pages in this homework.

## Problems:

1. (10 points) Think about weak learners in AdaBoost for a 2-class classification problem. Suppose you're using depth 0 decision trees, which simply return the weighted majority class of the data points as the classification, as the weak learner. Imagine that, at the first iteration, 80% of the data points were positive and 20% of the data points were negative. What would the cumulative weight of positive points and the cumulative weight of negative data points be after one round of boosting)? From your result, do you think whether using depth 0 decision trees as weak learners is a good idea?
2. (10 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.  $\vec{x} = (x_1, x_2)$  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  $\Phi(\vec{x})$  of every point  $\vec{x}$  in the original input space. Instead, all the work is done through the kernel function that computes dot products  $K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i)^T \Phi(\vec{x}_j)$ .

Show how to compute the squared Euclidean distance in the projected space between any two points  $\Phi(\vec{x}_i)$ ,  $\Phi(\vec{x}_j)$  without explicitly computing the  $\Phi$  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. (You may find it useful to first simplify the Boolean formula using common Boolean identities, such as De Morgan’s law.)
6. (20 points) Read one of the articles from the folder below:

<https://wustl.box.com/s/x68vmeajbgeozlowsz640i8tvjr1ezxs>.

Answer the following questions:

- (a) Summarize the article in 2 to 3 sentences.
- (b) Rephrase the issues raised in the article using the language you learned in this course.
- (c) Propose approaches to mitigate the issues raised in the article.

This question will be graded by CJ in a loose manner. The grading will focus on 1) whether you have put thoughts into your answer, and 2) whether your answers are logical (i.e., it’s okay to give unpopular opinion, but make sure you provide your reasoning).

Example grades (not an exhausted list, but just an illustration):

You will get full points if the summary is accurate, and the proposed approach makes sense (e.g., feasible if you are given the data and resource). You will get 15 points with an accurate summary but non-negligible flaws in the proposed approach. You will get 10 points for providing inaccurate summary (e.g., incorrect mapping to the language in this course).