

# [CSC 5825 Fall 2017]

Due. Before the Class of Dec, 11, 2017

Full credit: 100 points with extra 10 points bonus

November 19, 2017

**Question 1.** AdaBoost algorithm (100 points) Consider the following labeled data  $(x, y) \in R^2$  ( $i$  is the example index):

$i$	$x$	$y$	Label
1	11	3	-
2	10	1	-
3	4	4	-
4	12	10	+
5	2	4	-
6	10	5	+
7	8	8	-
8	6	5	+
9	7	7	+
10	7	8	+

In this problem, you will use AdaBoost to learn a hidden function from this set of training examples. We will use two rounds of AdaBoost to learn a hypothesis for this data set. In round number  $t$ , AdaBoost chooses a weak learner that minimizes the weighted error  $\epsilon_t$ . As weak learners, you will use axis parallel lines as I demonstrated in the class. Using the Ada Boost algorithm that I covered in the class, please answer the following questions:

- (1) The first step of AdaBoost is to create an initial data weight distribution  $D_1$ . What are the initial weights given to data points with index 4 and 7 by the AdaBoost algorithm, respectively? (10 points)
- (2) Which is the hypothesis  $h_1$  that minimizes the weighted error in the first round of AdaBoost, using the distribution  $D_1$  computed in the above question? (10 points)
- (3) What is the *weighted error* of  $h_1$  computed above? (10 points)
- (4) After computing  $h_1$  in the previous questions, we proceed to round 2 of AdaBoost. We begin by recomputing data weights depending on the error of  $h_1$  and whether a point was (mis)classified by  $h_1$ . What are the weights given to data points with index 4 and 7 according to the distribution after round 1,  $D_2$ , respectively? Do not forget to normalize the new data weights so that they sum to 1. (10 points)
- (5) Which is the hypothesis  $h_2$  that minimizes the weighted error in the second round of AdaBoost, using the distribution  $D_2$  computed in the above question? (10 points)

- (6) What is the weight assigned to the hypothesis of round 2,  $h_2$ ? (10 points)
- (7) Now that we have completed two rounds of AdaBoost, it is time to create the final output hypothesis. What is the final weighted hypothesis after two rounds of AdaBoost? (10 points)
- (8) Submit your source code with sufficient comments. (30 points)

**Bonus:** complete two iterations of bagging algorithm using the similar weak learner as the AdaBoosting and give your comments (10 points).

## Submission Instructions

To earn the full credit, you must type your solutions with sufficient details either using LaTeX or Microsoft Word **embedded** with source code and plots. Please submit it through the blackboard website. You will get points off for not following these requirements.

Late homework (w/o acceptable documents) will be accepted with penalty. 20% off penalty if late for 24 hours or less. 40% off penalty if late between 24 hours and 48 hours. 60% off penalty if late between 48 hours and 72 hours. 80% off penalty if late between 72 hours and 96 hours. No homework late more than 96 hours will be accepted.