Homework 5: Boosting and Ensembles

- Due Apr 8 at 11:59pm
- Points 50
- Questions 24
- Available Apr 1 at 11:59am - Apr 9 at 11:59pm
- Time Limit None
- Allowed Attempts 2

Instructions

Most of the homework asks you to step through the AdaBoost algorithm on a certain dataset. The goal is to understand the mechanics of the algorithm, and also to understand what ensembles are.

You will have two attempts and one week to answer these. After your first attempt, you will know which questions you got wrong, but not the right answers.

This quiz was locked Apr 9 at 11:59pm.

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	42 minutes	48 out of 50

ⓘ Answers will be shown after your last attempt

Score for this attempt: 48 out of 50
Submitted Apr 8 at 11:55pm
This attempt took 42 minutes.

You will have to step through the AdaBoost algorithm using the following dataset that has three binary features (x1, x2 and x3) and a binary label y:

x1	x2	x3	y
-1	-1	-1	-1
-1	+1	-1	-1
+1	-1	-1	-1
+1	+1	-1	+1

In order to instantiate the AdaBoost algorithm, we will need a collection of weak classifiers and a weak learner that picks one of them.

We will use weak classifiers that correspond to each feature. For example, corresponding to the feature x1, the *weak classifier 1* will predict the label +1 when the value of the feature is +1, and it will predict the label -1 when the feature becomes -1. Similarly, we have *weak classifiers 2* and *3* corresponding to the features x2 and x3 respectively. Together this gives us our hypothesis space $H = \{1, 2, 3\}$.

The weak learner we will use is the one that selects the weak classifier that has the lowest weighted error. We will be using the definition of weighted error from the AdaBoost lecture. (*If two weak classifiers have equal error, the weak learner will pick the lowest numbered one.*)

Recall from the lecture that AdaBoost keeps track of a distribution over the examples at each step. In the tth round, if we denote the distribution as D_t, in our example, this consists of four elements: $D_t(0)$ for the example #0, $D_t(1)$ for example #1, $D_t(2)$ for example #2 and $D_t(3)$ for example #3.

For the weak learner, at the tth round, you will need to rank the three hypotheses in terms of their weighted error ϵ_t. To compute the weighted error of a weak classifier, you need to add up the D_t values for examples where it makes an incorrect prediction.

In this question, you will have to work through three rounds of AdaBoost. In each round, you should:

1. Pick a weak classifier from H that has the lowest weighted error. This answer should be one of 1, 2 or 3, and will define h_t for round t.
2. Write the weighted error of that weak classifier (i.e. ϵ_t)
3. Compute the value of α_t for that round, defined as

   $\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$
4. Update the distribution D_t for the next round to produce D_{t+1}. This step is described in the lecture.

To start things off, in the first round, the distribution $D_0$is the uniform distribution over the four examples.

In the end, the final ensemble is defined as the function $\text{sgn}\left(\alpha_0 h_0 + \alpha_1 h_1 + \alpha_2 h_2\right)$.

Question 1
2 / 2 pts

Which hypothesis would the weak learner described in the question above pick in round 0.

- (•) weak classifier 1
- ( ) weak classifier 2
- ( ) weak classifier 3

Question 2
2 / 2 pts

What is the weighted error of the best classifier in round 0? That is, what is ϵ_0?

(Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.25

Question 3
2 / 2 pts

What is the value of α_0 for first weak classifier? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.549

Question 4
2 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the first round. This question focuses on the first example. What is $D_1(0)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.1667

Question 5
2 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the first round. This question focuses on the second example. What is $D_1(1)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.1667

Question 6

2 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the first round. This question focuses on the third example. What is $D_1(2)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.499

Question 7

2 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the first round. This question focuses on the fourth example. What is $D_1(3)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.1667

Question 8

2 / 2 pts

Which hypothesis would the weak learner described in the question above pick in round 1?

- 1
- 2
- 3

Question 9

2 / 2 pts

What is the weighted error of the best classifier in round 1? That is, what is ϵ_1?

(Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.1667

Question 10

2 / 2 pts

What is the value of α_1 for second weak classifier? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.805

Question 11

2 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the second round. This question focuses on the first example. What is $D_2(0)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.1

Question 12

2 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the second round. This question focuses on the second example. What is $D_2(1)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.499

Question 13

2 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the second round. This question focuses on the third example. What is $D_2(2)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.3

IncorrectQuestion 14

0 / 2 pts

This question is a part of four questions that ask you to compute the updated distribution over examples after the second round. This question focuses on the fourth example. What is $D_2(3)$ after the first weak classifier is selected? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.499

Question 15
2 / 2 pts

Which hypothesis would the weak learner described in the question above pick in round 2?

- 1
- 2
- 3

Question 16
2 / 2 pts

What is the weighted error of the best classifier in round 2? That is, what is ϵ_2?

(Your answer has to be within 0.01 of the right answer to be marked as correct.)

0.1

Question 17
2 / 2 pts

What is the value of α_2 for the third weak classifier? (Your answer has to be within 0.01 of the right answer to be marked as correct.)

1.0986

Question 18
4 / 4 pts

The final classifier that your AdaBoost algorithm constructs can be used to assign labels to the training data. Now, suppose we have four test examples. Select the labels for each example below that your ***final ensemble*** hypothesis will predict.

1. (x1, x2, x3) = (-1, -1, +1): -1

2. (x1, x2, x3) = (+1, -1, +1): +1

3. (x1, x2, x3) = (-1, +1, +1): +1

4. (x1, x2, x3) = (+1, +1, +1): +1

Answer 1:

-1

+1

Answer 2:

+1

-1

Answer 3:

+1

-1

Answer 4:

+1

-1

Question 19

2 / 2 pts

Does the ensemble classifier achieve zero training error?

- (•) True
- ( ) False

Question 20

2 / 2 pts

Could any of the weak classifiers we have achieve zero training error?

- ( ) True
- (•) False

Question 21

2 / 2 pts

Could the Perceptron algorithm learn a classifier for the data above with zero training error?

- (•) True

- False

Question 22
2 / 2 pts

Could a decision tree with depth limit 1 (i.e. no more than one feature used in the tree) achieve zero training error?

- True
- False

Question 23
2 / 2 pts

Suppose you had to build a random forest with whose decision trees are limited to depth 1 (i.e. no more than one feature used). If you were allowed to use as many trees as you want, could your random forest achieve zero training error?

- True
- False

Question 24
2 / 2 pts

Could a decision tree of depth no more than three achieve zero train and test errors?

- It could achieve zero train error, but not zero test error
- It could achieve zero train error, and zero test error
- It could neither achieve zero train error, nor zero test error
- It could not achieve zero train error, but could get zero test error

Quiz Score: 48 out of 50