Homework 3

- Due Feb 27 at 11:59pm
- Points 80
- Questions 27
- Available Feb 19 at 12am - Feb 28 at 11:59pm
- Time Limit None
- Allowed Attempts 3

Instructions

This homework consists of a collection of multiple choice questions. The goal of this homework is to revisit what we have seen so far in the semester.

More than one answer may be correct. You should select **all** the correct answers to get full points.

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

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	150 minutes	78 out of 80

ⓘ Correct answers are hidden.

Score for this attempt: 78 out of 80
Submitted Feb 27 at 10:23pm
This attempt took 150 minutes.

Question 1
3 / 3 pts

Suppose we have a binary classification problem with n features. Among these features, m of them can take three values, namely A, B, and C. All the other features are binary; that is they can take two values: `True` or `False`. Suppose we define a concept class C_1 consisting of all binary classifiers

over this feature space. What is the size of this concept class?

- [x] $2^{(3^m 2^{n-m})}$
- [ ] $3^{2(n+m)}$
- [ ] 2^{2^n}
- [ ] $3^m 2^{2^{n-m}}$

Question 2
3 / 3 pts

The concept class C_1 defined above is learnable in the mistake bound model.

- True
- False

Question 3
3 / 3 pts

Define C_2 as the set of all disjunctions with n Boolean features if we do not allow any negations. What is the size of this set?

- 2^{2^n}
- [ ]
- n^2
- [x] 2^n
- n
- [ ]

Question 4
3 / 3 pts

The concept class C_2 defined above is learnable in the mistake bound model.

- (•) True
- ( ) False

Question 5
3 / 3 pts

An m-of-n function is defined as follows: Select a *fixed* subset of Boolean variables of size n. The function returns true for inputs where m of these chosen variables are true.

Which of the following statements are correct about m-of-n functions?

- [x] m-of-n functions can be represented by linear classifiers
- [ ] Every Boolean function can be represented as a m-of-n function
- [x] Every disjunction without negations can be represented as a m-of-n function
- [x] Every conjunction without negations can be represented as a m-of-n function

Question 6
3 / 3 pts

Suppose we know that

$$P(X = A) = \frac{1}{16},$$
$$P(X = B) = \frac{1}{16},$$
$$P(X = C) = \frac{1}{8},$$
$$P(X = D) = \frac{1}{4},$$
$$P(X = E) = \frac{1}{2}$$

Select all statements that are correct.

- ( ) These probabilities give the minimum possible value of the entropy of X
- (•) Entropy(X) = 1.875 (exactly)
- ( ) These probabilities give the maximum possible value of the entropy of X

Entropy(X) = 1.311 (approximately)

Question 7

3 / 3 pts

Which of the following statements about decision trees are correct?

- [x] Every Boolean function can be represented as a decision tree
- [ ] Every Boolean function can be represented by a unique decision tree
- [ ] Decision trees represent only linearly separable functions
- [x] Real valued features have to be discretized to use them with decision trees

Question 8

3 / 3 pts

Which of the following statements about the ID3 algorithm are correct?

- [x] It will always find a decision tree that will fit any training set.
- [ ] It is an online algorithm.
- [ ] It assumes that the training set is chosen uniformly at random from the instance space.
- [x] It is a batch algorithm.

Question 9

3 / 3 pts

Suppose we have four features (x1, x2, x3, x4) and a label y that can be either A or B. We have the following training set:

x1	x2	x3	x4	y
0	0	0	1	A
0	1	1	0	A
1	0	0	0	A
1	0	1	1	B

Which of the following statements is true?

- [x] The entropy of the label is $2 - \frac{3}{4}\log_2 3$

- [ ] This data is not linearly separable.
- [x] This data is linearly separable
- [ ] The entropy of the label is $2 + \frac{3}{4}\log_2 3$

Question 10

3 / 3 pts

A learning algorithm is said to overfit its training data if:

- [x] Its training error is less than its generalization error
- [ ] Its hypothesis space contains the true concept function
- [ ] Its training error is more than its generalization error.
- [ ] Its hypothesis space is too small to express the data.

Question 11

3 / 3 pts

Use the following data with features x1, x2 and labels y and select all statements that are correct.

x1	x2	y
0	0	1
0	1	-1
1	0	-1
1	1	1

- [ ] This function can be represented by a linear threshold unit.
- [x] This function can not be represented by a linear threshold unit.
- [x] This function can be represented by a decision tree.
- [ ] This function can not be represented by a decision tree.

Question 12

3 / 3 pts

Consider the following dataset with four features (x1, x2, x3, x4) and a label y:

x1	x2	x3	x4	y
1	0	1	1	0
1	1	0	0	1
0	0	0	1	0
1	1	1	1	1

Which feature has the highest information gain?

- [ ] x1
- [x] x2
- [ ] x3
- [ ] x4

Question 13
3 / 3 pts

Consider the Boolean disjunction with two input features that is represented by the following data set:

x1	x2	y
0	0	0
0	1	0
1	0	0
1	1	1

What is the margin of this data set?

- [ ] $\frac{1}{2}$
- [ ] 1

$\frac{1}{\sqrt{2}}$

$\frac{1}{2\sqrt{2}}$

PartialQuestion 14
2 / 4 pts

Which of the following Boolean functions with variables x_1, x_2, x_3, x_4 are linearly separable?

- the XOR function
- $x_1 \vee \neg x_2$
- x_4
- $x_1 \wedge \neg x_2$
- Label is true when an even number of x's are true.
- Label is true when an odd number of x's are true.

Question 15
3 / 3 pts

You have a dataset on which you ran the Perceptron algorithm. You find that the algorithm doesn't stop making mistakes. Which of the following *may* help?

- Transform the data using a non-linear feature transformation.
- Nothing will help.
- Delete examples where the algorithm makes mistakes and try again.
- Change the labels for examples where the algorithm makes mistakes and try again.

Question 16
3 / 3 pts

Which of the following linear threshold units is equivalent to the following Boolean function: $\neg x_1 \vee x_2 \vee x_3$?

- [x] $sgn(-2x_1 + 2x_2 + 2x_3 + 1)$
- [ ] $sgn(-x_1 - 2x_2 - 2x_3 - 1)$
- [ ] $sgn(x_1 + x_2 + x_3 - 1)$
- [ ] $sgn(-x_1 - x_2 - x_3 + 1)$

Question 17

3 / 3 pts

According to the Perceptron mistake bound, what is the maximum number mistakes that the Perceptron algorithm make on a disjunction in n dimensions?

- [ ] 1
- [ ] $O(2^n)$
- [ ] $O(\log(n))$
- [ ]
- [x] $O(n^2)$

Question 18

3 / 3 pts

Consider the following two datasets, each with two-dimensional examples and binary labels.

Dataset A		Dataset B	
(x1, x2)	y	(x1, x2)	y
(1,1)	+1	(1,1)	-1
(1,2)	-1	(1,2)	-1
(2,2)	-1	(2,2)	+1
(2,3)	-1	(2,3)	+1

Select all assertions below that are valid.

- [x] Both datasets are linearly separable
- [x] It is possible to add more labeled points to either dataset to render it linearly *inseparable*
- [ ] The entropies of both datasets are the same
- [ ] Exactly one of the two datasets is linearly separable

Question 19

1 / 1 pts

Mistake-bound algorithms only exist for Boolean functions.

- ( ) True
- (•) False

Question 20

3 / 3 pts

Which of the following statements about the perceptron update is true?

- [ ] For a negative example, after the update, the dot product of the new weights and the example will be higher than it was before.
- [x] For a negative example, after the update, the dot product of the new weights and the example will be lower than it was before.
- [ ] For a positive example, after the update, the dot product of the new weights and the example will be lower than it was before.
- [x] For a positive example, after the update, the dot product of the new weights and the example will be higher than it was before.

Question 21

3 / 3 pts

Consider a single round of the Perceptron update. Say your current weight vector (which includes a bias weight) is $[1, 1, 1, 1]$ and the current example is $[-1, 2, 2, 1]$ with a label -1. What will the

weight vector be after the round is complete? Assume that the learning rate is 1.

- [ ] [0, 3,3,2]
- [ ] [1,1,1,1]
- [x] [2,-1,-1,0]
- [ ] [0, 1, 1, 0]

Question 22

3 / 3 pts

You have been hired as a machine learning consultant by a local company. You are asked to build a model that decides whether a customer who received an email promotion will make a purchase or not. What can you say about this problem?

- [x] It is a binary classification problem
- [ ] It is a multi-class classification problem
- [ ] It is a regression problem
- [ ] There is not enough information yet.

Question 23

3 / 3 pts

Which of the following statements are true about the least mean square regression?

- [x] Gradient descent can eventually converge to the optimum weights if the algorithm runs long enough.
- [ ] Stochastic gradient descent will converge to a better optimum than gradient descent
- [ ] Gradient descent will never converge to the optimum weights
- [x] The weights can be obtained analytically without requiring an optimization algorithm.

Question 24

3 / 3 pts

How many mistakes will the Halving algorithm make on disjunctions with n Boolean features where *every* variable has to be negated?

- [ ] $O(\log n)$

$O(2^n)$

$O(n)$

$O(n^2)$

Question 25

3 / 3 pts

How many k-disjunctions are possible with n Boolean variables if we do not allow negations?

$\binom{n}{k}$

3^k

2^k

$2^k \binom{n}{k}$

Question 26

3 / 3 pts

Which of the following statements is true about the original Perceptron algorithm?

It learns a linear classifier

It is a mistake bound algorithm

It is an online algorithm

It is a batch algorithm

Question 27

3 / 3 pts

Which of the following assumptions are used to formalize the PAC model of learnability?

Training examples are drawn independently of each other

Examples are presented in a sequence to the learning algorithm

Training examples could be generated by an adversary

- [x] Future examples will be drawn from the same fixed distribution as the training examples

Quiz Score: 78 out of 80