Quiz 3

Graded

Student

HARRIS DOAN

Total Points

12 / 14 pts

Question 1

(no title) **4** / 4 pts

1.1 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

1.2 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

1.3 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

1.4 **(no title)** **1** / 1 pt

✔ **+ 1 pt** Correct

Question 2

(no title) **0** / 1 pt

✔ **+ 0 pts** Incorrect

Question 3

(no title) **0** / 1 pt

✔ **+ 0 pts** Incorrect

Question 4

(no title) **2** / 2 pts

✔ **+ 2 pts** Correct

Question 5

Logistic regression **4** / 4 pts

5.1 (no title) **1** / 1 pt

✔ **+ 1 pt** Correct

5.2 (no title) **1** / 1 pt

✔ **+ 1 pt** Correct

5.3 (no title) **1** / 1 pt

✔ **+ 1 pt** Correct

5.4 (no title) **1** / 1 pt

✔ **+ 1 pt** Correct

Question 6

Convexity and optimization **2** / 2 pts

6.1 (no title) **1** / 1 pt

✔ **+ 1 pt** Correct

6.2 (no title) **1** / 1 pt

✔ **+ 1 pt** Correct

Q1
4 Points

This question contains the following True/False subquestions about the Perceptron and logistic regression.

Q1.1
1 Point

Can a Perceptron model represent boolean function XNOR? The table shows the results of XNOR computation on two independent binary variables.

X_1	X_2	Y
0	0	1
1	0	0
0	1	0
1	1	1

- (●) No
- ( ) Yes

Q1.2
1 Point

An updated vector w_{new} always has lower training error compared to the previous weight vector w.

- (●) False
- ( ) True

Q1.3
1 Point

The perceptron algorithm eventually terminates on linearly separable data.

- (●) True
- ( ) False

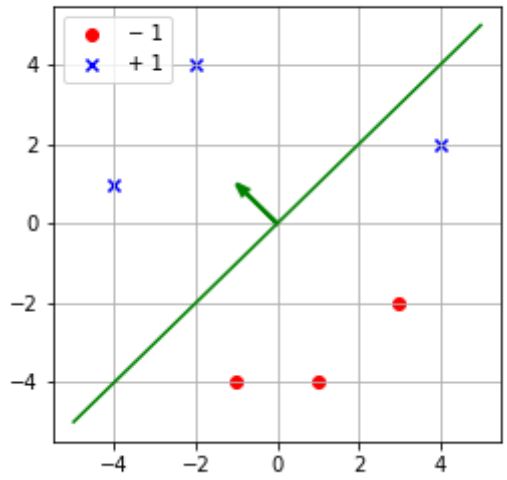
Q1.4
1 Point

On linearly separable data, the perceptron algorithm finds the separating hyperplane with the largest margin.

- (●) False
- ( ) True

Q2
1 Point

The following diagram represents data and a hyperplane.



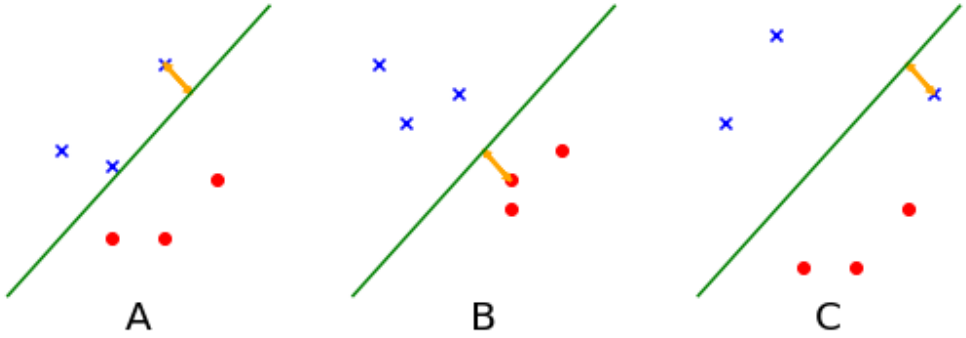
Blue crosses (x) and red circles (o) correspond to labels $y = +1$ and $y = -1$, respectively. The current hyperplane is defined by its normal vector $w = (-1, 1)$ (green arrow). Suppose we are in process of learning a perceptron on the dataset, and we use the algorithm described in the lecture.

What will the updated vector w be after the current iteration?

- ○ $(2, 0)$
- ○ $(-2, -2)$
- ○ $(3, 3)$
- ◉ $(2, -1)$

Q3
1 Point

Which of the following figures correctly labels the margin (the orange arrow) *with respect to the hyperplane*?



- C
- None of the above
- B
- A

Q4
2 Points

Given the following three points in $\mathbb{R}^3$:

$X = (3, 1, 4),$
$Y = (1, 5, 6),$
$Z = (4, 9, 13),$

which one of the following vectors is the normal vector of the plane formed by points X, Y, Z?

- $(2, 1, 1)$
- $(1, 1, -1)$
- $(1, -1, 2)$
- $(-1, 2, 1)$

Q5 Logistic regression

4 Points

For this question, $\mathbf{x}$ is the feature vector, x, x_i's are scalar features, and y is the label.

Q5.1

1 Point

Suppose we build a logistic regression classifier $P(y = 1|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x} + b)$ for a binary classification task. Which one of the following statements about logistic regression is correct?

- ○ In general, logistic regression and perceptron learn the same parameters $\mathbf{w}, b$ for the decision boundary.
- ○ Logistic regression is guaranteed to maximize the training accuracy.
- ○ The decision boundary of logistic regression is nonlinear.
- ◉ Logistic regression minimizes the negative log likelihood of the training data.

Q5.2

1 Point

Given a binary logistic regression model $P(y = 1|x) = \sigma(2x - 3)$ where σ is the sigmoid function. What is the decision boundary?

- ○ $x = -2$
- ○ $x = 0$
- ○ $2x = -3$
- ◉ $2x = 3$

Q5.3
1 Point

Given a logistic regression model $P(y=1|\mathbf{x}) = \sigma(1 + 0.7x_1 + 0.2x_2 - x_3)$ where σ is the sigmoid function, increasing the value of the third feature (keeping the value of the other features fixed) decreases the model's output probability for $y = 1$.

- (●) True
- ( ) False

Q5.4
1 Point

Given a logistic regression model $P(y=1|\mathbf{x}) = \sigma(1 + 0.7x_1 + 0.2x_2 - x_3)$ where σ is the sigmoid function, when all features take the value 0, the model predicts $y = 1$.

- ( ) False
- (●) True

Q6 Convexity and optimization

2 Points

Q6.1

1 Point

The following function $f(x) = 10x^4$ is convex.

- ○ False
- ◉ True

Q6.2

1 Point

Which of the following statements about gradient descent is correct?

- ○ Gradient descent, using an appropriate step size, will converge to the global minimum of a nonconvex function.
- ○ Gradient descent, using any step size, will converge to the local minimum of a function.
- ◉ Gradient descent, using an appropriate step size, will converge to the local minimum of a function.
- ○ Gradient descent is used to maximize a function.