

Quiz 3

Started: Oct 21 at 1:35pm

Quiz Instructions

Question 1

1 pts

About gradient descent (GD) and stochastic gradient descent (SGD) for logistic regression, which of the following statement(s) are true?

- ☐ In GD, we use the same data to compute the gradient in every epoch, so every epoch we will get the same gradient.
- ☐ In SGD, we pick a random example from the dataset to approximate the gradient. In expectation, the approximated gradient direction is the same as the true gradient.
- ☐ In GD, we need to use the whole dataset to compute the gradient.
- ☐ The computation cost of each iteration in SGD is much less than GD as it only needs to compute the gradient with respect to one data point.

Question 2

1 pts

Suppose we have a binary classification dataset with 10 examples. Given the ground-truth labels and predictions by 4 different models, which model has the best F1 score?

Ground-truth	0	0	0	1	1	1	1	1	1	1
Model A	0	0	0	0	0	0	0	0	0	0
Model B	1	1	1	1	1	1	1	1	1	1
Model C	0	1	1	0	1	1	1	1	1	1
Model D	0	0	1	0	1	1	1	1	1	1

- ☐ Model A

☐ Model B☐ Model C☐ Model D**Question 3****1 pts**

Suppose we have two Boolean variables $x_1, x_2 \in \{0, 1\}$. We define an activation function $\text{ReLU}(x) = \max\{0, x\}$. Which of the following function(s) can successfully classify XOR? Note that $\text{XOR}(x_1, x_2) = (x_1 \wedge \neg x_2) \vee (\neg x_2 \wedge x_1)$.

(A) $y = 2|x_1 + x_2 - 1| - 1$

(B) $y = |x_2 - x_1 - 1| - 1.5$

(C) $y = \text{ReLU}(x_1 - x_2) + \text{ReLU}(x_2 - x_1) - 0.5$

(D) $y = \text{ReLU}(x_1 + x_2) + \text{ReLU}(x_2 - x_1) - 0.5$

☐ A☐ B☐ C☐ D**Question 4****1 pts**

Given variables a, b, c, d, e, f , where they satisfy the followings:

$$c = \sigma(w_1a + w_2b),$$

$$d = \sigma(w_3a + w_4b),$$

$$e = \sigma(w_5a + w_6b),$$

$$f = \sigma(w_7d + w_8e),$$

$$g = \sigma(w_9c + w_{10}f),$$

where w_i are constants and $\sigma(\cdot)$ denotes sigmoid function. Which of the following statement(s) are true? (Hint: it would be helpful to draw a computational graph.)

(A)

$$\frac{\partial g}{\partial b} = \frac{\partial g}{\partial c} \cdot \frac{\partial c}{\partial b} + \frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial b}$$

(B)

$$\frac{\partial g}{\partial b} = \frac{\partial g}{\partial c} \cdot \frac{\partial c}{\partial b} + \frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial d} \cdot \frac{\partial d}{\partial b}$$

(C)

$$\frac{\partial g}{\partial a} = \frac{\partial g}{\partial c} \cdot \frac{\partial c}{\partial a}$$

(D)

$$\frac{\partial f}{\partial a} = \frac{\partial d}{\partial a} + \frac{\partial e}{\partial a}$$

☐ A

☐ B

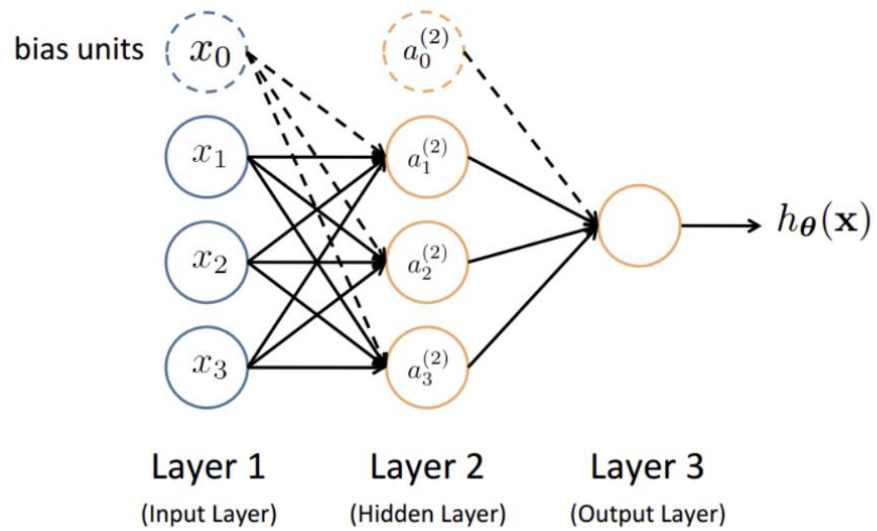
☐ C

☐ D

Question 5

1 pts

What is the total number of parameters in the following neural network:



- ☐ 9
- ☐ 12
- ☐ 16
- ☐ 25

Not saved

Submit Quiz