Quiz-02

Due Jan 29 at 11:59pm **Points** 10 **Questions** 10
Available Jan 27 at 11:59pm - Jan 29 at 11:59pm **Time Limit** None
Allowed Attempts 3

Instructions

Learning in neural nets

This quiz covers topics from lectures 3 and 4, which cover the basics of learning in neural networks.

Topics in the quiz include those in the hidden slides in the slidedecks.

Take the Quiz Again

Attempt History

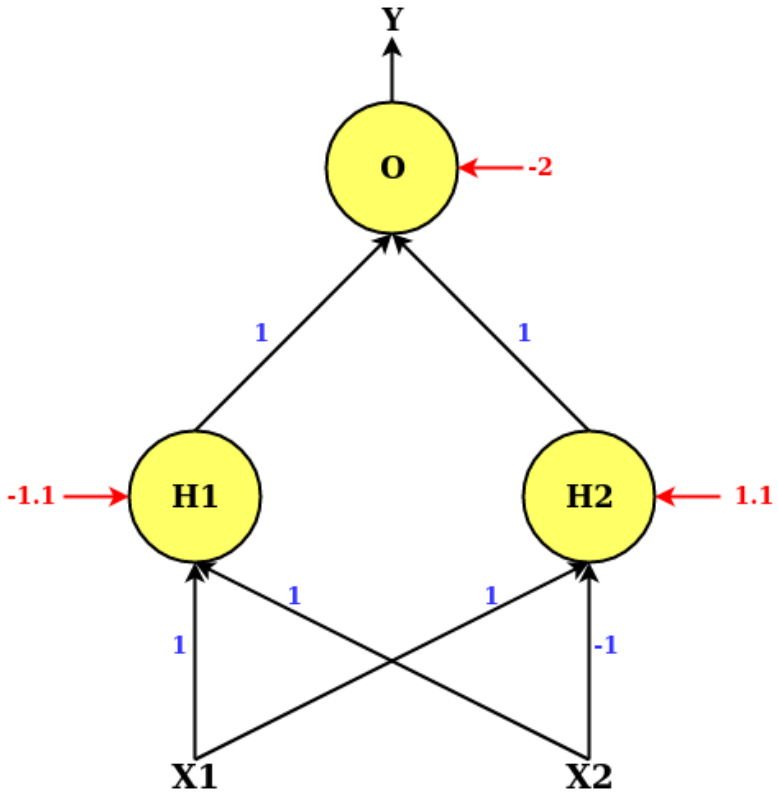
	Attempt	Time	Score
LATEST	Attempt 1	463 minutes	7 out of 10

ⓘ Answers will be shown after your last attempt

Score for this attempt: **7** out of 10
Submitted Jan 28 at 10:39pm
This attempt took 463 minutes.

Question 1 — 1 / 1 pts

Consider the following MLP and the given parameters:



X1 and **X2** are the inputs to the network. **Y** is the output of the network. **H1** and **H2** are the hidden neurons and **O** is the output neuron. The weights of the connections are shown in blue against the corresponding black arrows. The biases are shown in red. Each neuron uses the threshold activation function:

$$\phi(z) = \begin{cases} 1, \text{ if } z >= 0 \\ 0, \text{ otherwise} \end{cases}$$

If the inputs to the network are **X1**=0 and **X2**=0 and the desired output is **d**=0, which of the neurons will be updated first assuming that the MADALINE update rule is used?

Hint: See hidden ADALINE and MADALINE slides (Lec3 p77)

- The output neuron O
- All the neurons would be updated in the first step
- The first hidden neuron H1
- The second hidden neuron H2
- None of the neurons would be updated

Question 2

1 / 1 pts

(Select all that apply) In order for our NN to represent a specific function, *in practice* we will actually try to:

Hint: See slide Lec3 p130

- [ ] Minimize the weights and biases
- [x] Adjust network parameters such that the network's output matches the desired output as closely as possible, on the training instances
- [ ] Maximize the network's parameters
- [x] Minimize the empirical error on the training data

Question 3

1 / 1 pts

Gradient descent steps will always result in a decrease in the loss function we are minimizing .

Hint: See lec 4 slide 48

- ( ) True
- (•) False

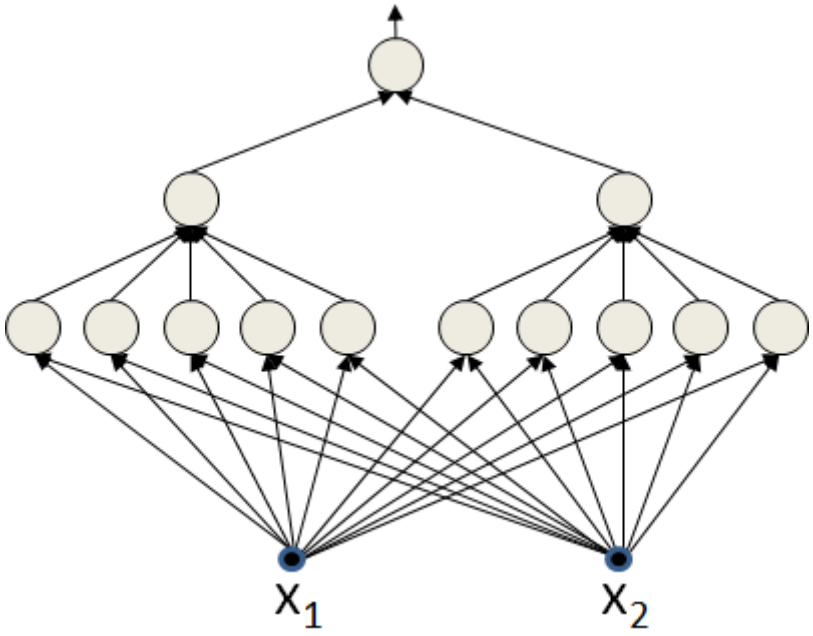
The step size might be too large

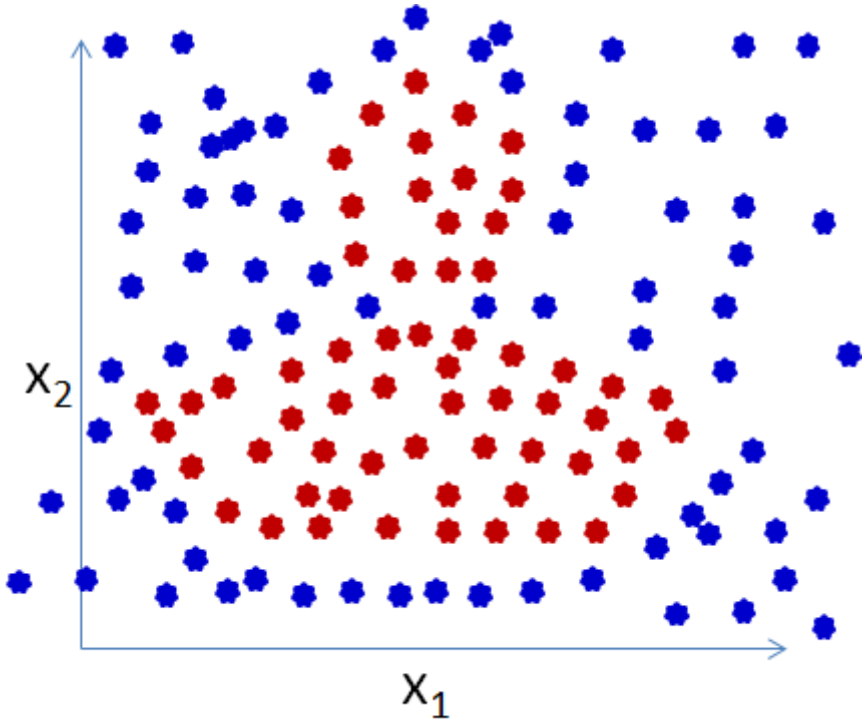
Incorrect

Question 4

0 / 1 pts

True/false: As explained in class, the perceptron learning algorithm can be used to learn the parameters of the threshold-activation MLP shown (including the weights and thresholds for every neuron), for the training data shown in the second figure (red and blue markers belong to two different classes):





Hint: See slide Lec3 p69-73

- ○ True
- ◉ False

Question 5

1 / 1 pts

(Select all that apply) For ADALINE, which of the following statements are true?

Hint: See slide Lec3 p77-80

- ☑ The calculated error is equivalent to that of a perceptron with identity activation
- ☑ Moves weights in opposite direction of the gradient of the MSE (wrt the weights)
- ☐ Is equivalent to the generalized delta rule
- ☑ Has a linear decision boundary
- ☐ Is equivalent to the perceptron learning rule

Question 6

1 / 1 pts

(Select all that apply) Which of the following statements are true?

Hint: See slide Lec3 p81 and p88

ADALINE uses a linear approximation to the perceptron that ignores the threshold activation. MADALINE, on the other hand, is greedy but exact.

MADALINE is simply ADALINE, when it utilizes parallel computation

ADALINE is used to train individual neurons, while MADALINE is used to train the entire network

MADALINE utilizes ADALINE to update neuron parameters

Question 7

1 / 1 pts

(Select all that apply) Which of the following steps could give us the minimum point of a function f(x) that is twice differentiable and defined over the reals?

Hint: See Lec4 slide 28

Computing the second derivative f”(x) and find an x where f”(x) =0 and f’(x) = 0

Computing the second derivative f”(x) and find an x where f”(x) <0 and f’(x) = 0

Computing the second derivative f”(x) and find an x where f”(x) >0 and f’(x) = 0

Computing the second derivative f”(x) and find an x where f”(x) >0 and f’(x) > 0

Incorrect

Question 8

0 / 1 pts

A matrix is said to be positive definite if all of its Eigenvalues are positive. If some are zero, but the rest are positive, it is positive semi-definite. Similarly, the matrix is negative definite if all Eigen values are negative. If some are negative, but the rest are zero, it is negative semidefinite. If it has both positive and negative Eigenvalues, it is “indefinite”.

An N-dimensional function has an NxN Hessian at any point. The Eigenvalues indicate the curvature of the function along the directions represented by the corresponding Eigenvectors of the Hessian. Negative Eigen values indicate that the function curves down, positive Eigenvalues show it curves up, and 0 Eigenvalues indicate flatness.

(Select the correct answer) The Hessian of the function $f(x_1, x_2, x_3) = x_1^2 x_2 + x_2^2 x_3 + x_3^3 + 2x_1 x_3 + x_2 x_3 + 6$ at the point (0,0,0) is :

Hint: See lec 4, slide 19, 33-34, and rewatch that portion of the lecture. You will have to work out the Hessian and compute its Eigenvalues.

- (selected) Positive semidefinite
- Positive definite
- Negative definite
- Indefinite
- Negative semidefinite

Question 9

1 / 1 pts

Suppose Alice wants to meet Bob for a secret meeting. Because it is a secret meeting, Bob didn't tell Alice the exact location where the meeting would take place. He, however, told her where to start her journey from and gave her directions to the meeting point. Unfortunately, Alice forgot the directions he gave to her. But she knows that the meeting would take place at the top of a hill close to her starting location.

Suppose the elevation of the ground that she is standing on is given by the equation $z = 20 + x^2 + y^2 - 10\cos(2\pi x) - 10\cos(2\pi y)$ where x, y are the 2-D coordinates and z is the elevation.

Alice decides to apply what she learned about function optimization in her DL class to go to the secret location. She decides to modify the gradient descent algorithm and walks in the direction of the fastest increase in elevation (instead of going opposite to the direction of fastest increase), hoping to reach the top of the hill eventually. Suppose she starts at the point **(-0.1, 0.2)** and uses a step size (learning rate) of 0.001. At what point would she end up after taking 100 such steps? Truncate your answer to 1 digit after the decimal point.

Hint: See Lec 4 slides 40-43. The answer will require simulation.

$x =$ -0.5

$y =$ 0.5

Answer 1:

-0.5

Answer 2:

0.5

Incorrect

Question 10

0 / 1 pts

In the lecture slides, what does it imply when the representation of a perceptron in the illustrations of MLPs do not contain an extra line to show the bias?

Hint: Lec 3 slides 7-8

- The neuron has no bias and operates on non-linear, affine combinations of the inputs.
- It is a lazy illustration that does not show the extra input connection for the bias, and the bias must be assumed to be present.
- The neuron is a threshold unit in which the threshold is not shown inside the circle representing the unit, because it is too small.
- **The neuron has no bias and operates on linear combinations of the inputs.**

If you got any of these wrong, please watch the portion of the lecture corresponding to the hints.

Quiz Score: **7** out of 10