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	1,971 minutes	7.5 out of 10

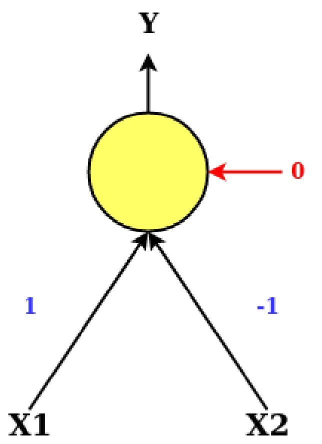
! Answers will be shown after your last attempt

Score for this attempt: **7.5** out of 10
Submitted Jan 29 at 11:52pm
This attempt took 1,971 minutes.

Question 1

1 / 1 pts

Consider the following perceptron:



X1 and **X2** are the inputs to the network. **Y** is the output of the network. The weights of the connections are shown in blue against the corresponding black arrows. The biases are shown in red. The perceptron uses the threshold activation function:

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

If the inputs to the perceptron are **X1**=1 and **X2**=1.25, the desired output is **d**=1and the weights and bias are updated using the ADALINE learning rule with a learning rate of η=0.4, then what will be the values of the weights and the bias after the 1st iteration? (Assume that **W1** is the weight associated with **X1** and **W2** is the weight associated with **X2**)

Hint: See hidden ADALINE and MADALINE slides

- **W1**=1, **W2**=-1, bias=0

- **W1**=1.5, **W2**=-0.375, bias=0.5
- **W1**=1, **W2**=-1, bias=0.5
- **W1**=0.5, **W2**=-1.625, bias=0.5
- **W1**=1.5, **W2**=0.375, bias=-0.5

Question 2

1 / 1 pts

The *true* objective that we wish to minimize while learning network parameters, when we want it to represent a specific function, is **(select all that apply):**

Hint: See slide Lec3 P125-130

- [ ] The expected empirical risk over a training set
- [x] The average (expected) divergence between the output of the network and the actual function being approximated, over the entire domain of the input
- [ ] None of these
- [ ] The average error over all training points

Question 3

1 / 1 pts

(Select all that apply) Which of the following is true of the gradient of a function $f : \mathbb{R}^d \to \mathbb{R}$, computed at any point

Hint: See slide Lec4 p10

- [ ] The gradient is a vector that points in the direction of fastest decrease of the function at that point
- [ ] You can always compute the gradient of any function at any location
- [x] The gradient is a vector that points in the direction of fastest increase of the function at that point
- [x] The gradient is a vector composed of the partial derivatives of the scalar output of a function with respect to the components of its vector input
- [x] The length of the gradient is indicative of the actual rate of change of the function, in the direction of the gradient

Question 4

1 / 1 pts

The perceptron learning algorithm is always guaranteed to converge in a finite number of steps.

Hint: Lec 3 slide 55

- [ ] True
- [x] False

It only converges in finite steps if the classes are linearly separable

Question 5

1 / 1 pts

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

Hint: See slide Lec3 p81 and p88

- [x] MADALINE utilizes ADALINE to update neuron parameters
- [ ] 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
- [x] ADALINE is used to train individual neurons, while MADALINE is used to train the entire network

Incorrect

Question 6

0 / 1 pts

(Select all that apply) How does ADALINE resolve the non-differentiability of the threshold activation?

Hint: See slide Lec3 p77-81

- [x] It uses a differentiable sigmoidal approximation to the threshold function during learning, but uses the hard threshold activation subsequently when operating on test data.
- [ ] It computes the squared error between the output of the perceptron and the target output, instead of counting errors.
- [x] It tries to minimize the error between the desired binary output and the affine combination of inputs before the threshold activation is applied.
- [ ] It ignores the threshold activation during training, and only applies it during testing.

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
- [x] Computing the second derivative f"(x) and find an x where f"(x) >0 and f'(x) = 0

Question 8

1 / 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 $f(x_1, x_2, x_3) = x_1^2 + 3x_2^2 + 2x_3^2$ at the point $(-1, \sqrt{2}, -1)$:

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.

- Negative semidefinite
- Indefinite
- Positive definite

Hessian: [[2,0,0], [0,6,0], [0,0,4]] and eigenvalues: 2, 6, 4

- Negative definite
- Positive semidefinite

Incorrect

Question 9

0 / 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 **(1.7, -1.2)** 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.

$x =$ -1.5

$y =$ -0.5

Answer 1:

-1.5

Answer 2:

-0.5

Partial

Question 10

0.5 / 1 pts

Which of the following statements are true, according to lecture 3? **(select all that apply)**

Hints: Lecture 4 discussion on derivatives, lecture 4 discussion on divergence, and lec 4 – individual neurons.

- [ ] Making the activation functions of the neurons differentiable enables us to determine how much small perturbations of network parameters influence the number of training data instances that are misclassified, and so helps us determine how to modify the parameters to reduce this number.
- [x] The actual objective of training is to minimize the average error on the training data instances.
- [x] The derivative of a function $f(x)$ with respect to a variable z tells you how much minor perturbations of z perturbs $f(x)$
- [ ] The derivative $\nabla_x f$ of a function $f(x)$ of a vector argument x, with respect to x, is the same as the gradient of $f(x)$ with respect to x.
- [ ] The derivative of a function $y = f(x)$ with respect to its input x is the ratio $\frac{dy}{dx}$ of small increments in the output that result from small increments of the input.

- [x] It is necessary for both the activations and the divergence function that quantifies the error in the output of the network to be differentiable functions in the function minimization approach to learning network parameters.

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

Quiz Score: **7.5** out of 10