

Quizizz

Quiz Name: Quiz-2 (MLP, Gradient Descent, and Backprop)

Student: Konda Mopuri

Date: Thu Sep 16 2021 - 11:19 am

92%

Accuracy % (correct / total)

14

Total Questions

13

✓ Correct

1

✗ Incorrect

0

⊘ Unattempted

Questions

No.	Question	Time Taken	Score	Response
1	A multi layer neural network with ReLU activation functions in the hidden layers, if the neurons operate in linear region it can be approximated to a single layer neural network. [1 Mark]	16 secs	850	✓ True
2	Consider a 2-layered MLP as shown in the figure. Hidden neurons have ReLU activation function and the output neuron is linear. Loss function is defined as $L = 1/2 (y-d)^2$. Let the input be $x=(1,2)$ and the corresponding target $d=1$. What is the predicted value by the network, loss value for the prediction, and derivative of the loss with respect to the weight 3 (in the o/p layer)? [4 Marks] (comma separate the two quantities in the answer)	316 secs	0	3, 2, 2

No.	Question	Time Taken	Score	Response
3	From the class discussion on Gradient Descent, discuss what happens if η value is (i) large +ve, (ii) small +ve, and (iii) -ve quantity. Provide one sentence for each case. [3 Marks]	143 secs	1000	(i) large +ve value makes the optimization to oscillate about the local minima without smoothly converging, (ii) small +ve value progresses slowly towards the local minima taking too many steps for optimization, (iii) -ve value takes the optimization in the undesired direction in which the loss increases and hence the optimization does not converge on small loss value.
4	What happens if we use linear activation functions in the hidden layers of an MLP? How does it affect the capacity of the resulting model? [2 Marks]	85 secs	1000	All the hidden layers of the MLP can essentially be represented with a single layer. Despite using multiple layers in the architecture, it does not improve the model capacity because of the linear activation functions.
5	Choose all the statements that are True about the Universal Approximation Theorem (UAT). [3 Marks]	17 secs	850	✓ UAT only discusses the theoretical possibility for the approximating neural network.
6	Since the gradient always gives the useful direction, during the optimization using gradient descent we will always reach the lowest point of the loss surface. [1 Mark]	9 secs	940	✓ False
7	When using the Back-Propagation learning algorithm in an MLP, what exactly is back-propagated, and from where to where? [2 Marks]	58 secs	1000	It is the gradient of loss that is back-propagated. From the final (deepest) layer of the neural network via all the intermediate layers until the input layer.

No.	Question	Time Taken	Score	Response
8	What is learning rate in Back-Propagation training? [2 Marks] (Answer in about 2 sentences)	218 secs	1000	Learning rate is the step size used in the parameter updation using the gradient descent algorithm. In other words, the amount by which the parameters of the neural network change in every iteration is proportional to the learning rate.
9	Backpropagaion learning is based on gradient descent along error surface. [1 Mark]	13 secs	890	✓ True
10	Select all the True statements about Gradient Descent. [3 Marks]	16 secs	960	✓ It is a supervised ML technique Learning rate is a hyper parameter in gradient descent that is used for parameter updating
11	What are (pick all) general limitations of back propagation rule? Pick the right choice from the given options. [3 Marks]	14 secs	970	✓ local minima problemslow convergence scaling
12	In a given MLP, the number of nodes in the input layer is 20 and the hidden layer is 10. The maximum number of connections from the input layer to the hidden layer are ? [1 Mark]	9 secs	920	✓ 200
13	For a binary classification problem, which of the following architectures could be used? [2 Mark]	3 secs	980	✓ both
14	Pick the right order of steps for the usage of a gradient descent algorithm. [2 Marks] 1. Calculate error among the actual target and the anticipated target 2. Reiterate until you find the excellent weights of the network 3. Pass a samples and get values from the output layer 4. Initialize random weight and bias 5. Go to every neurons which contributes to the error and update its respective values to lessen the error	28 secs	900	✓ 4, 3, 1, 5, 2