Take E as the fulcrum node and compute the missing values in the savings list given below.

Construct the savings tour. Enter the path representation of the tour starting from city E.

AB	AC	AD	AF	BC	BD	BF	CD	CF	DF
88	57	123	117	5	?	?	58	48	72

Enter a comma separated list of city names.

NO SPACES, TABS, DOTS, BRACKETS, PARENTHESIS OR UNWANTED CHARACTERS.

Answer format: E,X,Y,Z

Response Type : Alphanumeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Set

Answers Case Sensitive : No

Text Areas : PlainText

Possible Answers :

E,C,F,A,D,B

Deep Learning

Section Id :	64065351480
Section Number :	3
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	12
Number of Questions to be attempted :	12
Section Marks :	50
Display Number Panel :	Yes

Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653108402
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 42 Question Id : 640653740160 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DEGREE LEVEL : DEEP LEARNING (COMPUTER BASED EXAM)"

ARE YOU SURE YOU HAVE TO WRITE EXAM FOR THIS SUBJECT?

CROSS CHECK YOUR HALL TICKET TO CONFIRM THE SUBJECTS TO BE WRITTEN.

(IF IT IS NOT THE CORRECT SUBJECT, PLS CHECK THE SECTION AT THE TOP FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406532475512. YES

6406532475513. NO

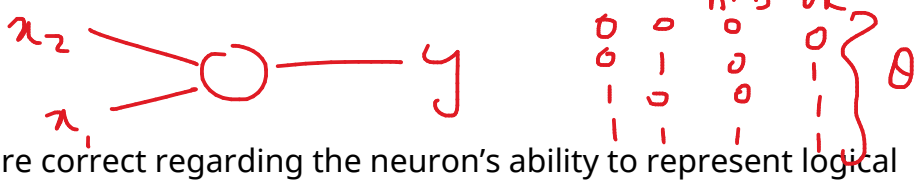
Sub-Section Number :	2
Sub-Section Id :	640653108403
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 43 Question Id : 640653740161 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider a neuron with binary inputs x_1 and x_2, and an output y. The neuron computes the weighted sum of its inputs and produces an output according to a threshold. The threshold is denoted as θ. The activation function is such that $y = 1$ if the weighted sum is greater than or equal to θ, otherwise $y = 0$.



Which of the following statements are correct regarding the neuron's ability to represent logical AND and OR functions?

Options :

6406532475514. The neuron can implement the AND function by setting appropriate weights and a threshold.

6406532475515. The neuron can implement the OR function by setting appropriate weights and a threshold.

6406532475516. There exists a single set of weights and a threshold that allows the same neuron to correctly implement both the AND and OR functions simultaneously.

6406532475517. Neurons are limited to implementing either the AND or the OR function and cannot represent both simultaneously.

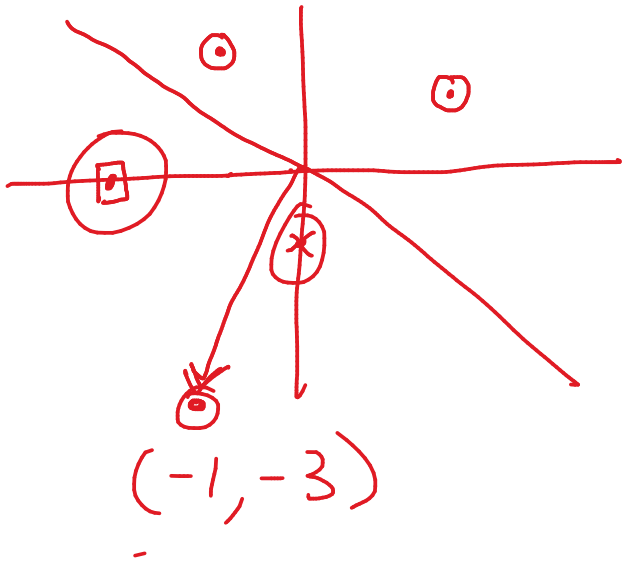
Sub-Section Number :	3
Sub-Section Id :	640653108404
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 44 Question Id : 640653740162 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Short Answer Question

Suppose we have a perceptron with two inputs, x_1 and x_2. This perceptron undergoes training on a small dataset containing three points: (−1, 2) labeled as class 0, (0,−1) labeled as class 1, and (2, 1) labeled as class 0. The weights of the perceptron are initialized to zeros, and the model is trained until it reaches convergence. Given this scenario, what would be the assigned output class by the trained perceptron for the new point (−2, 0)?



Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

Question Number : 45 Question Id : 640653740173 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Short Answer Question

You are training a neural network for sentiment analysis on a dataset of 10,000 text reviews. The dataset is divided into 80% for training and 20% for testing. You decide to use Minibatch Gradient Descent with a batch size of 32. If you perform a total of 100 epochs, how many parameter updates will be performed in total?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

Sub-Section Number : 4

Sub-Section Id : 640653108405

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 46 Question Id : 640653740163 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Let's assume a continuous function f(x_1,x_2) is approximated using 3D tower function with a 3 hidden layer neural network using 100 towers. How many minimum number of neurons will we need to approximate the function f(x_1,x_2) ?

Options :

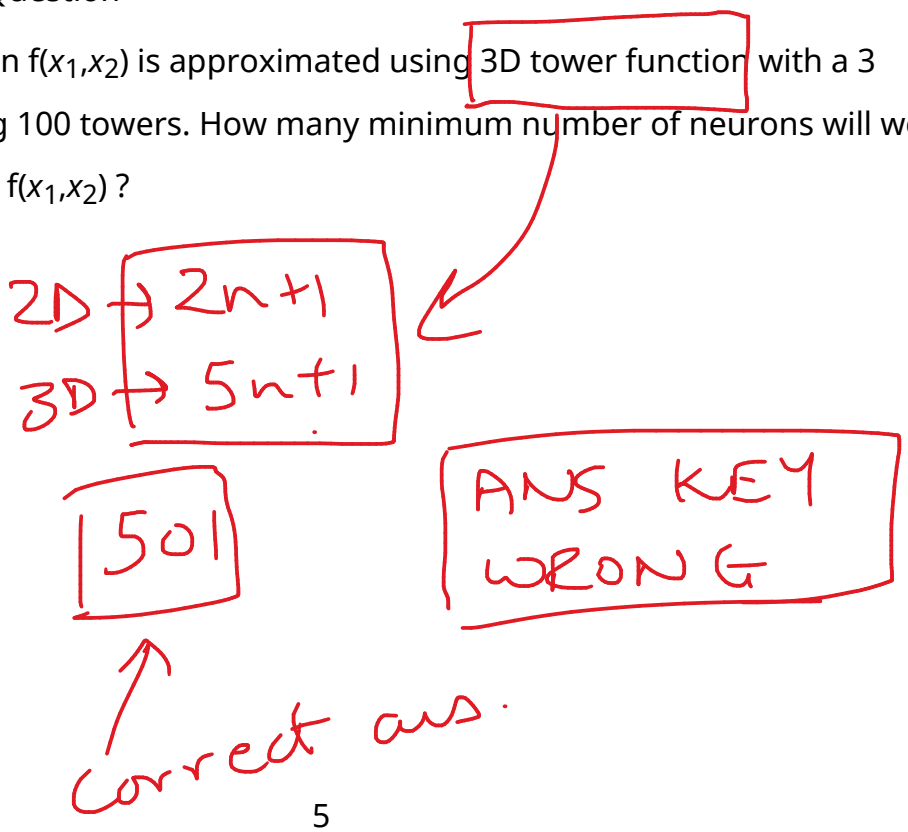
6406532475519. 201

6406532475520. 300

6406532475521. 801

6406532475522. 701

6406532475523. 800



Sub-Section Number : 5

Sub-Section Id : 640653108406

Question Shuffling Allowed : Yes

Is Section Default? : null

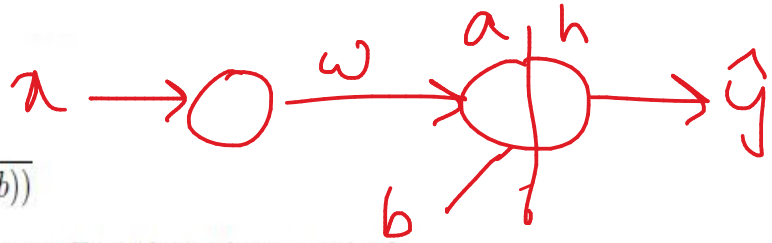
Question Number : 47 Question Id : 640653740164 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Short Answer Question

The logistic sigmoid neuron $\sigma(x)$ is defined as follows

$$\sigma(x) = \frac{1}{1 + exp(-(wx + b))}$$



where $w, b \in \mathbb{R}$ are learnable parameters.Take Mean Square Error loss where required

$$L = 0.5 * (\hat{y} - y)^2$$

Suppose we use the sigmoid function to fit the pair $x = 0, y = 1$, where x is an input and y is the ground truth. Suppose that w is initialized to $w = 2$ and b is initialized to $b = 1$.The prediction $\hat{y}$ by the model for the current w, b is, $\hat{y} = 0.731$. Update the parameter once by keeping $\eta = 10$ and compute the loss. Enter the new loss value.

Note: Enter the loss value to three significant digits. That is, if your answer is 0.06134, then enter it as 0.061

$x=0 \quad y=1$
$w=2 \quad b=1$
$\hat{y} = 0.731$
$\eta = 10$

$w_{t+1} = w_t - \eta \nabla w$
$b_{t+1} = b_t - \eta \nabla b$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

$$\Rightarrow \frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a} \cdot \frac{\partial a}{\partial w}$$

① $L = \frac{1}{2}(\hat{y} - y)^2 \Rightarrow \frac{\partial L}{\partial \hat{y}} = \hat{y} - y = -0.269$

② $\hat{y} = h = \sigma(a) \Rightarrow \frac{\partial \hat{y}}{\partial a} = (h)(1-h) = 0.731 \times 0.269 = 0.1966$

Question Number : 48 Question Id : 640653740174 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

③ $a = wx + b \Rightarrow \frac{\partial a}{\partial w} = x = 0$

$\Rightarrow \frac{\partial L}{\partial w} = 0$

Question Label : Short Answer Question

Consider a quadratic loss function given by $L(w) = (w - 4)^2$, where w represents the model parameter. You are using momentum-based gradient descent to minimize this loss. The momentum parameter is set to 0.9, the learning rate is 0.1, and the initial parameter value is w = 0. What is the updated weight value after one iteration ?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a} \cdot \frac{\partial a}{\partial b}$$

④ $a = wx + b \Rightarrow \frac{\partial a}{\partial b} = 1$

$\Rightarrow \frac{\partial L}{\partial b} = -0.269 \times 0.1966 \times 1 = -0.053$

$$\therefore w_{t+1} = 2 - 10(0) \Rightarrow w_{t+1} = 2$$

$$b_{t+1} = 1 - 10(-0.053) \Rightarrow b_{t+1} = 1.53$$

$$a_{t+1} = wx + b = b = 1.53$$

$$h_{t+1} = \hat{y}_{t+1} = \sigma(a_{t+1}) = \sigma(1.53) = 0.822$$

$$L_{t+1} = 0.5(\hat{y}_{t+1} - y)^2 = 0.5(0.822 - 1)^2$$

$$= 0.5(-0.178)^2 = \boxed{0.0158}$$

Sub-Section Number : 6

Sub-Section Id : 640653108407

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 49 Question Id : 640653740165 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Multiple Choice Question

Consider a sigmoid function

$$f(x) = \frac{1}{1+e^{-(wx+b)}}$$

Suppose that w is restricted to take only positive values ($w > 0$). Suppose further that we define the steepness of the curve as absolute value of the slope. Then, select all the correct statements about the function

Options :

6406532475525. Increasing the value of b shifts the sigmoid function to the left (i.e., towards negative infinity)

6406532475526. Increasing the value of b shifts the sigmoid function to the right (i.e., towards positive infinity)

6406532475527. Decreasing the value of w increases the steepness of the sigmoid function

6406532475528. Increasing the value of w decreases the steepness of the sigmoid function

Question Number : 50 Question Id : 640653740176 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Multiple Choice Question

Which of the following is true, given the optimal learning rate?

Options :

6406532475540. Batch gradient descent is always guaranteed to converge to the global optimum of a loss function.

6406532475541. Stochastic gradient descent is always guaranteed to converge to the global optimum of a loss function.

6406532475542. For convex loss functions, stochastic gradient descent is guaranteed to eventually converge to the global optimum while batch gradient descent is not.

6406532475543. For convex loss functions, both stochastic gradient descent and batch gradient descent will eventually converge to the global optimum.

6406532475544. For convex loss functions, neither stochastic gradient descent nor batch gradient descent are guaranteed to converge to the global optimum.

6406532475545. For convex loss functions, batch gradient descent is guaranteed to eventually converge to the global optimum while stochastic gradient descent is not.

Sub-Section Number :	7
Sub-Section Id :	640653108408
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653740166 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (51 to 53)

Question Label : Comprehension

A neural network contains an input layer $\mathbf{h_0} = \mathbf{x}$, three hidden layers $(\mathbf{h_1}, \mathbf{h_2}, \text{and}, \mathbf{h_3})$ and an output layer $\mathbf{O}$. All the hidden layers use *Sigmoid* activation and the output layer uses softmax activation. Suppose the input $\mathbf{x} \in \mathbb{R}^{400}$ and all the hidden layers contain 10 neurons each. The output layer contains 5 neurons.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 51 Question Id : 640653740167 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Short Answer Question

How many parameters (including biases) are there in the entire network?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

Question Number : 52 Question Id : 640653740168 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Short Answer Question

Suppose that all elements in the input vector are zero and the corresponding true label is also 0.

Further, suppose that all the parameters are initialized to zero.

What is the loss value if cross-entropy loss is used? Use natural logarithm ln.

Response Type : Alphanumeric

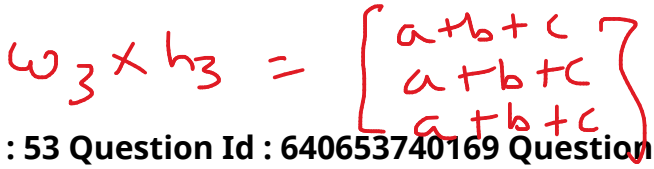
Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

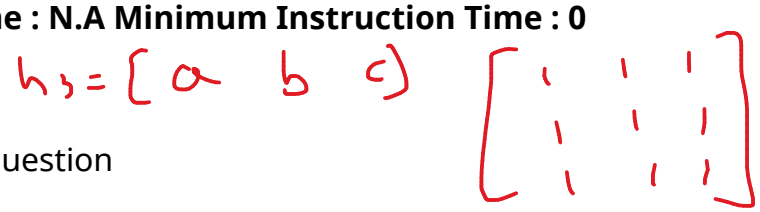
Possible Answers :



Question Number : 53 Question Id : 640653740169 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4



Question Label : Short Answer Question

Assuming that all weights between layers $\mathbf{h_3}$ and **O** are initialized to one, with no bias associated with any neuron, what would be the computed cross-entropy loss for a given single data point? If the provided information is insufficient, please enter −1.

Response Type : Alphanumeric

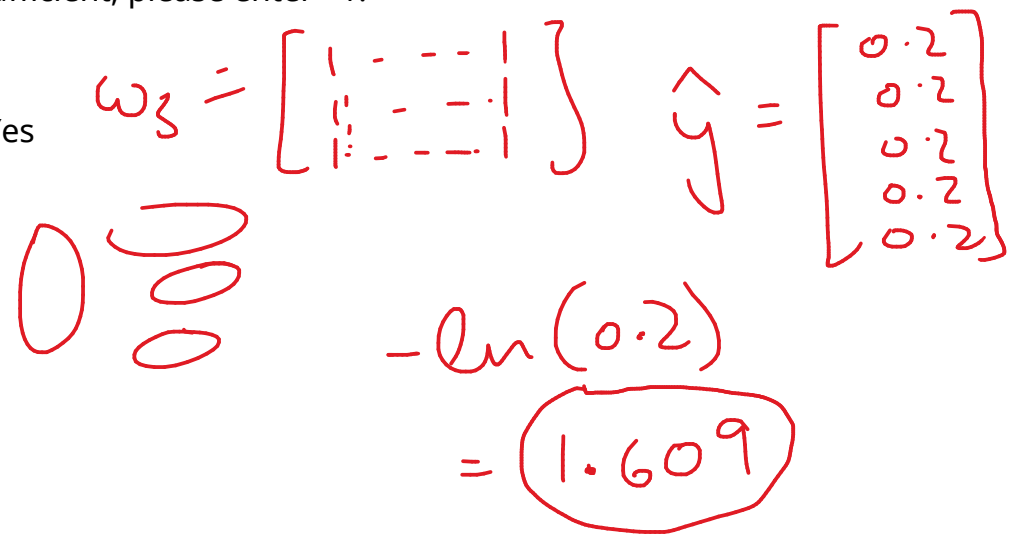
Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :



Sub-Section Number :	8
Sub-Section Id :	640653108409
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653740170 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (54 to 55)

Question Label : Comprehension

Consider a neural network with two hidden layers and one output layer, as shown below. Here, a_k is a pre-activation and h_k is the output from the $k-$th layer. All the neurons in the hidden layers use a sigmoid activation function. Suppose that the neural network is used for a classification problem with 3 classes (such that any given input belongs to exactly one class). Suppose further that the number of neurons in each hidden layer is 3 and we have only three neurons in the output layer. Suppose that the weights are initialized as follows

$$W_1 = W_2 = W_3 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

The true label for the input $h_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is one hot encoded as $y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

Assume that the output layer uses softmax activation and the neurons in the network has no bias associated with it.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 54 Question Id : 640653740171 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Short Answer Question

Compute the cross entropy loss.

Note:If you think the given information is not sufficient to calculate the loss, then enter -1 as answer.

Response Type : Alphanumeric

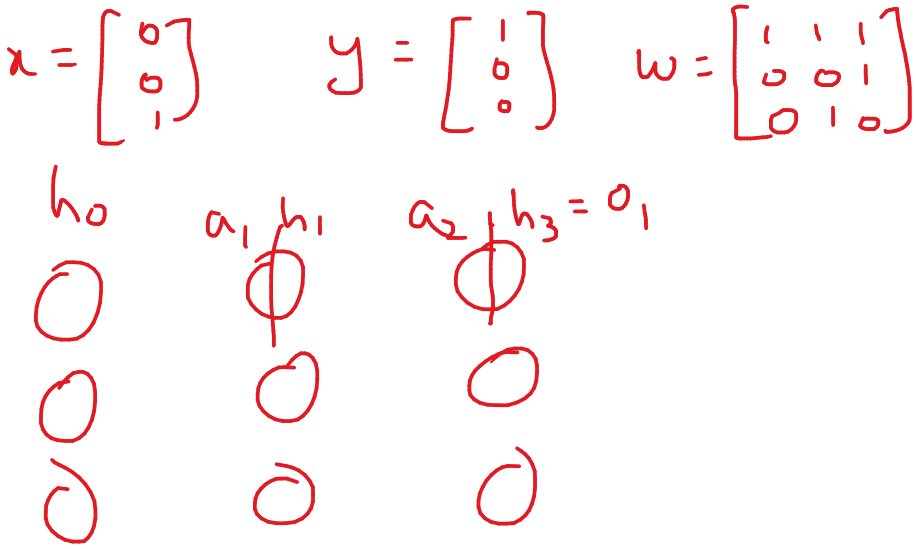
Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :



$$\omega_1 \times h_0 = a_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \Rightarrow h_1 = \begin{bmatrix} 0.731 \\ 0.731 \\ 0.5 \end{bmatrix}$$

$$a_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.731 \\ 0.731 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1.962 \\ 0.5 \\ 0.731 \end{bmatrix} \Rightarrow h_2 = \begin{bmatrix} 0.877 \\ 0.623 \\ 0.675 \end{bmatrix}$$

$$a_3 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.877 \\ 0.623 \\ 0.675 \end{bmatrix} = \begin{bmatrix} 2.175 \\ 0.675 \\ 0.623 \end{bmatrix}$$

$$h_{30} = \frac{e^{a30}}{e^{a30} + e^{a31} + e^{a32}} \Rightarrow h_3 = \begin{bmatrix} 0.697 \\ 0.156 \\ 0.148 \end{bmatrix}$$

$$\text{Loss} = -\ln(0.697) = \underline{\underline{0.361}}$$

Question Number : 55 Question Id : 640653740172 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Short Answer Question

Use cross entropy loss and compute the gradient of w_{10} (that is, ∇w_{10}) of W_3. (More precisely, ∇w_{310})

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a} \cdot \frac{\partial a}{\partial w}$$

$$[\hat{y} - y] \cdot [h_2]^T$$

Response Type : Alphanumeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

(1) $\hat{y} - y = \begin{bmatrix} 0.697 - 1 \\ 0.156 - 0 \\ 0.148 - 0 \end{bmatrix} = \begin{bmatrix} -0.303 \\ 0.156 \\ 0.148 \end{bmatrix}$

$$\begin{bmatrix} -0.303 \\ 0.156 \\ 0.148 \end{bmatrix} \begin{bmatrix} 0.877 & 0.623 & 0.675 \end{bmatrix}$$

Sub-Section Number : 9

Sub-Section Id : 640653108410

Question Shuffling Allowed : Yes

Is Section Default? : null

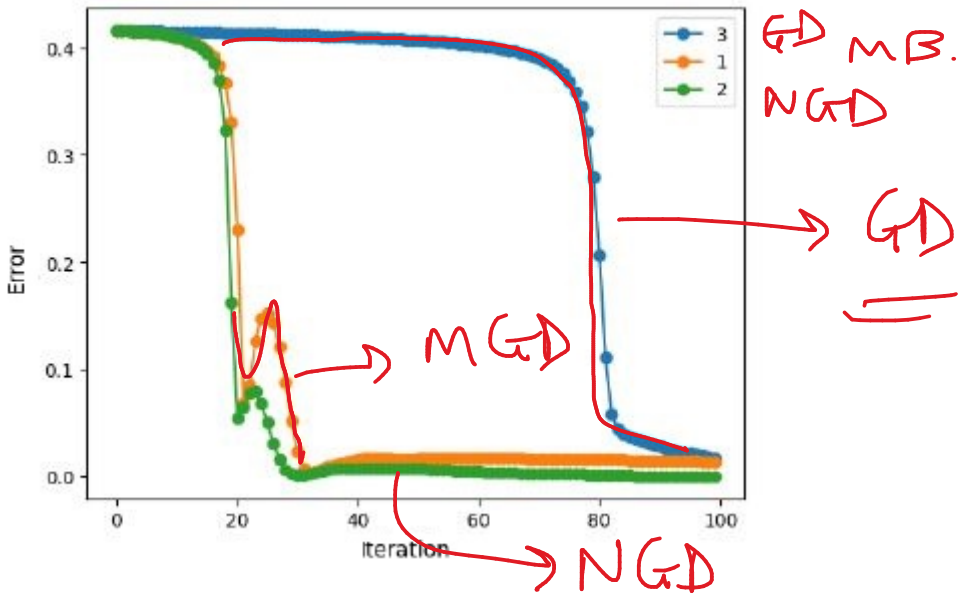
$$\Rightarrow \frac{\partial L}{\partial w_{310}} = 0.156 \times 0.877 = 0.1368$$

Question Number : 56 Question Id : 640653740175 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Consider the following image:



As per your understanding of optimization algorithms, which of the following mappings will be correct (assume optimal learning rate)?

Options :

6406532475536. 1: Gradient Descent

2: Momentum based Gradient Descent

3: Nesterov Accelerated Gradient Descent

6406532475537. 1: Momentum based Gradient Descent

2: Gradient Descent

3: Nesterov Accelerated Gradient Descent

6406532475538. 1: Gradient Descent

2: Nesterov Accelerated Gradient Descent

3: Momentum based Gradient Descent

6406532475539. 1: Momentum based Gradient Descent

2: Nesterov Accelerated Gradient Descent

3: Gradient Descent

Programming in C

Section Id : 64065351481

Section Number : 4