GC,112

Deep Learning

Section Id :	64065330377
Section Number :	6
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	9
Number of Questions to be attempted :	9
Section Marks :	50
Display Number Panel :	Yes
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 :	64065367960
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 104 Question Id : 640653470865 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"

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 :

6406531564582. ✔ YES

6406531564583. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065367961
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 105 Question Id : 640653470866 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following statement(s) about MP neuron is (are) true?

Options :

6406531564584. ✖ Given linearly separable data points, it learns the decision rule from the data points

6406531564585. ✔ The inputs to the neuron are Boolean

6406531564586. ✖ The neuron must have inhibitory inputs to implement a Boolean function

6406531564587. ✔ The neuron always use a linear decision boundary to separate the data points

Question Number : 106 Question Id : 640653470869 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

(Multiple select) Assume that we have absolutely linearly separable data points $\mathbf{x_i} \in \mathbb{R}^d, i = 1, 2, \cdots N$. Each data point belongs to either class 1 (positive P) or class 0 (negative N) (i.e., $y \in \{0, 1\}$). The perceptron outputs($\hat{y}$) 1 if $\mathbf{w^T x} \geq 0$ else it outputs 0. $\mathbf{w}$ is a randomly initialized weight vector. Suppose we use the perceptron learning algorithm to separate the data points. Choose the right update equation(s) that finds the decision boundary that separates the data points after t iterations.

Options :

6406531564593. ✖ $\mathbf{w} = \mathbf{w} + (\hat{y} - y)\mathbf{x}$

6406531564594. ✔ $\mathbf{w} = \mathbf{w} + (y - \hat{y})\mathbf{x}$

6406531564595. ✖ $$\mathbf{w} = \begin{cases} \mathbf{w} + \mathbf{x}, & \text{if } \mathbf{w^T x} \geq 0, x \in \mathbf{P} \\ \mathbf{w} - \mathbf{x}, & \text{otherwise} \end{cases}$$

6406531564596. ✔ $$\mathbf{w} = \begin{cases} \mathbf{w} - \mathbf{x}, & \text{if } \mathbf{w^T x} \geq 0, x \in \mathbf{N} \\ \mathbf{w} + \mathbf{x}, & \text{if } \mathbf{w^T x} < 0, x \in \mathbf{P} \end{cases}$$

Sub-Section Number :	3
Sub-Section Id :	64065367962
Question Shuffling Allowed :	Yes
Is Section Default? :	null

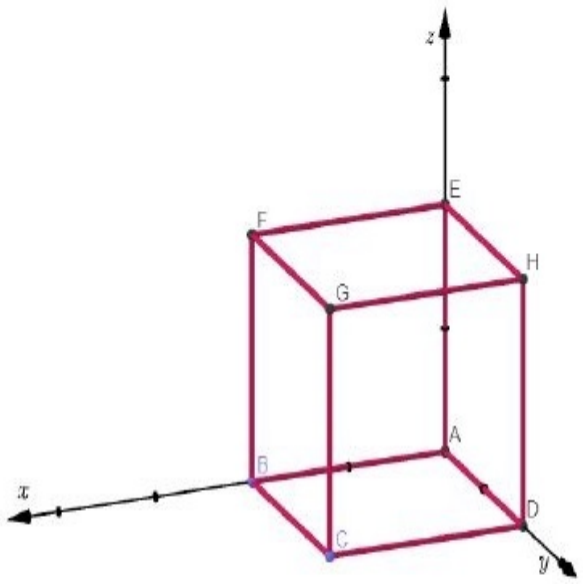
Question Number : 107 Question Id : 640653470867 Question Type : SA Calculator : None

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

Correct Marks : 6

Question Label : Short Answer Question

The coordinate of each vertex of the unit cube shown below is an input to the MP neuron. The coordinate of the vertex is represented by an ordered tuple (x, y, z). The vertex A has the coordinate of (0,0,0) and G has the coordinate of (1,1,1). The vertices (A,E,H,D) are of negative (0) class and the vertices (F,G,B,C) are



of positive (1) class. The MP neuron uses the following decision rule

$$\hat{y} = \begin{cases} 1, & \text{if } x + y + z > \theta \\ 0, & \text{otherwise} \end{cases}$$

where $\theta \in \mathbb{R}$. Suppose that we set two of the inputs (y, z) to always zero. Then find the threshold value that separates these data points correctly? If such threshold doesn't exist, then enter your answer as -1. (If it exists, then enter the answer to one decimal point, say, if your answer is -2.1031, then enter it as -2.1)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.1 to 0.9

Sub-Section Number :	4
Sub-Section Id :	64065367963
Question Shuffling Allowed :	Yes

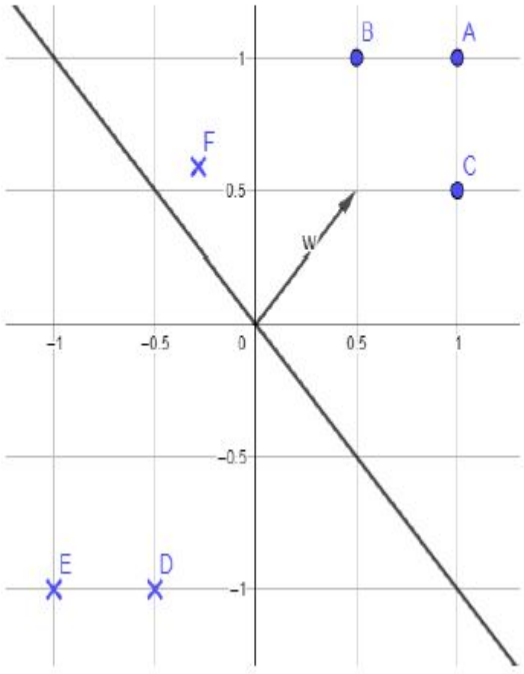
Is Section Default? : null

Question Number : 108 Question Id : 640653470868 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

The diagram below shows a decision boundary of a perceptron. The weight vector w is perpendicular to the decision boundary initially. The current decision boundary misclassifies some data points. After running the perceptron learning algorithm for a few iterations, the weight gets updated and the corresponding decision boundary separates all the data points correctly. Then the angle between the updated weight vector



and its corresponding decision boundary

Options :

6406531564589. ✖ reduces

6406531564590. ✔ remains 90^o

6406531564591. ✖ increases

6406531564592. ✖ insufficient information

Sub-Section Number : 5

Sub-Section Id : 64065367964

Question Shuffling Allowed : No

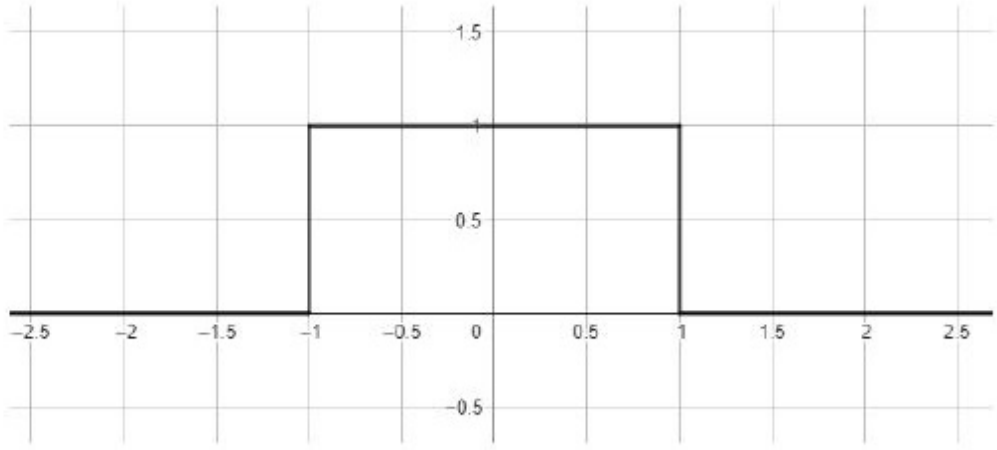
Is Section Default? : null

Question Id : 640653470870 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 : (109 to 112)

Question Label : Comprehension

Consider the function shown below. Suppose we wish to approximate this function using



a Multi Layer Perceptron (MLP) with a suitable activation function.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 109 Question Id : 640653470871 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What is the minimum number of neurons required to implement the network?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

3

Question Number : 110 Question Id : 640653470872 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 use a hard-threshold as an activation function for the neurons in the first hidden layer and add/subtract the outputs from the first hidden layer to obtain the function. The hard threshold function is defined to be

$$y = \begin{cases} 1, & \text{if } wx \geq \theta \\ 0, & \text{otherwise} \end{cases}$$

where, y is an output from neuron and θ is the threshold of neuron.

Assume we limit the number of neurons in the layer to two and let θ_1 and θ_2 represent their respective threshold values.

Find the suitable threshold values and enter their sum, i.e., $\theta_1 + \theta_2$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0

Question Number : 111 Question Id : 640653470873 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

The statement that replacing the hard-threshold by a sigmoid neuron with suitable values for w and θ achieves the same with certain degree of accuracy is

Options :

6406531564599. ✔ TRUE

6406531564600. ✖ FALSE

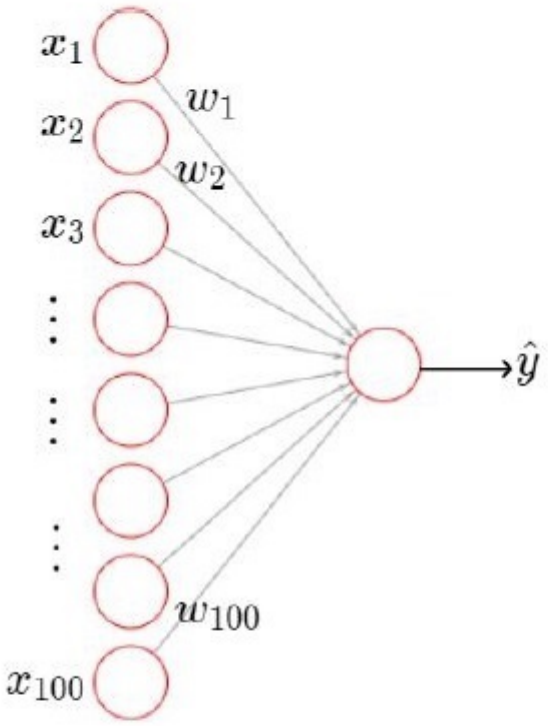
Question Number : 112 Question Id : 640653470874 Question Type : SA Calculator : None

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

Correct Marks : 6

Question Label : Short Answer Question

Consider a sigmoid neuron that takes in an input vector x of size 100, all the elements in x are 1.
The parameters (weights and bias) are initialized to 1, that is,



$w_i, b = 1, \forall i$. Suppose we randomly set 80% of the inputs to zero. Assume the true label $y = 0$ and Mean Squared Error loss function. Update the parameters by running gradient descent algorithm with the learning rate $\eta = 0.1$ for 3 iterations. Enter the sum of updated parameters. In case of insufficient information, enter your answer as -1.
Note: Truncate the gradient value to two decimal points, that is, if gradient is 0.043210001, then take it as 0.04.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

101

Sub-Section Number : 6

Sub-Section Id : 64065367965

Question Shuffling Allowed : No

Is Section Default? : null

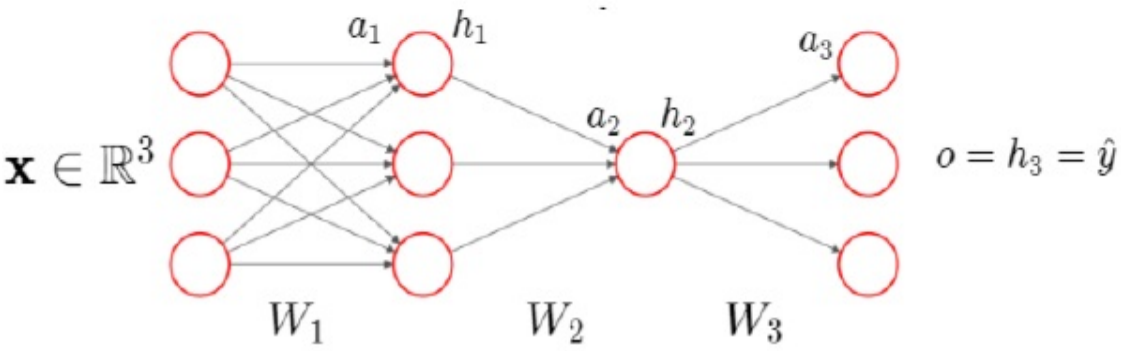
Question Id : 640653470875 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (113 to 117)

Question Label : Comprehension

Consider a feed forward neural network shown below where, $\mathbf{x}$ is an input vector. The vectors a_l, h_l correspond to pre-activation and activation at layer l. The matrices $\mathbf{W_l}$ are weights that connect neurons from layer $l-1$ to layer l. Finally, the vector $\mathbf{o}$ is an output vector $\mathbf{o} = \mathbf{h_3} = \hat{\mathbf{y}}$.

All neurons in the hidden layer use the logistic (sigmoid) activation function, and neurons in the output layer use softmax function. Further, the network minimizes cross entropy loss, $\mathcal{L} = -\sum y_i \log_e(\hat{y}_i)$.



Based on the above data, answer the given subquestions.

Sub questions

Question Number : 113 Question Id : 640653470876 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What is the total number of parameters (excluding bias) in the network?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

15

Question Number : 114 Question Id : 640653470877 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

The number of paths via which the input x_1, that is the first element of **x**, affects the loss is

Note: Path is a directed edge that connects one neuron with other neuron

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

9

Question Number : 115 Question Id : 640653470878 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

The input to the network is $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Suppose that all the weights in the network are initialized to a very large positive value, (say, $w_{ijl} \gg 5, \forall i, j, l$) and biases are set to zero.
Then enter the sum of the elements of $\hat{y}$. (If you think, it is not possible to calculate $\hat{y}$ with the given information, enter -1)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.98 to 1.00

Question Number : 116 Question Id : 640653470879 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

Assume that $h_2 = 0$, $\mathbf{W_3} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and

$\mathbf{b_3} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Suppose that the true label

$\mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Compute the gradient

of loss with respect to $\hat{\mathbf{y}}$ and enter the sum of the elements.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

-3.1 to -3.0

Question Number : 117 Question Id : 640653470880 Question Type : SA Calculator : None

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

Correct Marks : 5

Question Label : Short Answer Question

Assume that $h_2 = 0$,

$$\mathbf{W_2} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{W_3} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{b_3} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Suppose that the true label

$\mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Update the weights of

$\mathbf{W_2}$ (take $\eta = 1$) and enter its sum.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

1

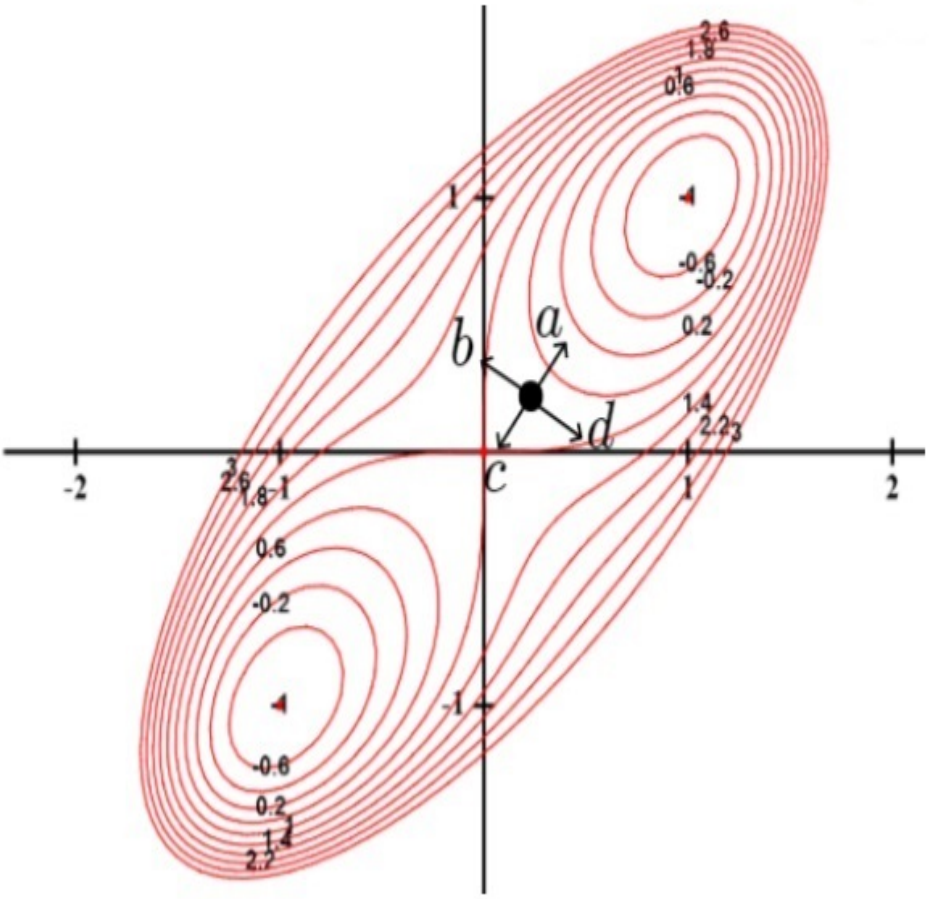
Sub-Section Number :	7
Sub-Section Id :	64065367966
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 118 Question Id : 640653470881 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5

Question Label : Multiple Choice Question

The diagram below shows the contours of a loss surface on a (w_1, w_2) plane. Suppose that the parameters are initialized at the point marked by the solid black circle. What will be direction of descent if we run the gradient descent algorithm with an appropriate learning rate?



Options :

6406531564607. ✖ d

6406531564608. ✖ b

6406531564609. ✔ a

6406531564610. ✖ c

Sub-Section Number :	8
Sub-Section Id :	64065367967
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 119 Question Id : 640653470882 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

The statement that using vannila (plain) gradient descent algorithm always avoids oscillation around local minimum while converging is

Options :

6406531564611. ✖ TRUE

6406531564612. ✔ FALSE

PSM

Section Id :	64065330378
Section Number :	7
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	4
Number of Questions to be attempted :	4
Section Marks :	50
Display Number Panel :	Yes
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 :	64065367968
Question Shuffling Allowed :	No
Is Section Default? :	null