Text Areas : PlainText

Possible Answers :

C,F,B,D,A,E

Deep Learning

Section Id :	64065359426
Section Number :	3
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	15
Number of Questions to be attempted :	15
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 :	No
Section Maximum Duration :	0
Section Minimum Duration :	0
Section Time In :	Minutes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653124102
Question Shuffling Allowed :	No

Question Number : 42 Question Id : 640653829604 Question Type : MCQ

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 :

6406532786528. YES

6406532786529. NO

Sub-Section Number : 2

Sub-Section Id : 640653124103

Question Shuffling Allowed : Yes

Question Number : 43 Question Id : 640653829605 Question Type : SA

Correct Marks : 3

Question Label : Short Answer Question

Consider a dataset of 100 points $x_1, x_2, \cdots, x_{100}$.

First 50 points are $x_1 = x_2 = \cdots = x_{50} = \begin{bmatrix} a \\ a \end{bmatrix}$ and next 50 points are $x_{51} = x_{52} = \cdots = x_{100} = \begin{bmatrix} -a \\ -a \end{bmatrix}$, where $a > 0$. The first 50 data points belong to the positive class (denoted as 1) and the next 50 data points belong to the negative class (denoted by 0). Suppose that the perceptron learning algorithm is used to find the decision boundary that separates these data points with the following rule,

$$f(x) = \begin{cases} 1 & \text{if } w^T x \geq 0 \\ 0 & \text{if } w^T x < 0 \end{cases}$$

The algorithm checks the data points in order. How often do the weights get updated until convergence? The weights do not include bias. If the algorithm does not converge, enter the answer as -1

① If wts are not given, it means wt = zero vector

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

∴ For 1st 50 points:

Text Areas : PlainText

Possible Answers :

① $\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} a \\ a \end{bmatrix} = 0 \geq 0 \Rightarrow$ +ve.

Question Number : 44 Question Id : 640653829607 Question Type : SA

Correct Marks : 3

Question Label : Short Answer Question

For pts. 51 to 100:

② $\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} -a \\ -a \end{bmatrix} = 0 \geq 0 \Rightarrow$ wrong classification

∴ update wt vector

$\Rightarrow w_{t+1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} -a \\ -a \end{bmatrix} = \begin{bmatrix} a \\ a \end{bmatrix}$

Ans ↓ 1

③ Check pt. 52.

$\begin{bmatrix} a \\ a \end{bmatrix}\begin{bmatrix} -a \\ -a \end{bmatrix} = -2a^2 < 0 \Rightarrow$ Correct class

Consider a single McCulloch-Pitts (MP) neuron with four binary inputs x_1, x_2, x_3, and x_4. The neuron produces an output y based on a threshold function. The MP neuron uses the following decision rule

$$\hat{y} = \begin{cases} 1, & \text{if } x_1 + x_2 + x_3 + x_4 > \theta \\ 0, & \text{otherwise} \end{cases}$$

Given the following input combinations and their corresponding outputs:

Inputs: $x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1$ Output: $y = 1$

Inputs: $x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0$ Output: $y = 0$

Inputs: $x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 1$ Output: $y = 1$

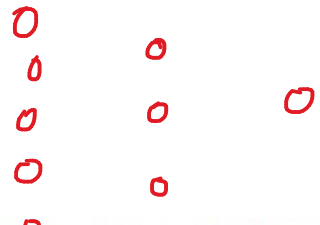
What minimum threshold value is required for the neuron to produce an output of 1? If the threshold can not be determined using the given information, enter the answer as -1.

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 : 640653829613 Question Type : SA
Correct Marks : 3
Question Label : Short Answer Question

Consider a feedforward neural network with one hidden layer trained using backpropagation for a binary classification task with classes labeled as 1 and 0. The network architecture is structured as follows:

- Input layer consisting of 5 neurons
- Hidden layer containing 3 neurons
- Output layer comprising 1 neuron



During the backpropagation process, the derivative of the sigmoid activation function $\sigma(z)$ with respect to its argument z is given by:

$$\sigma'(z) = \sigma(z) \cdot (1 - \sigma(z))$$

Sigmoid.

If the loss function utilized for binary classification is the binary cross-entropy loss, and both the hidden layer and output layer use the sigmoid activation function, the cross-entropy loss is represented by:

$$L(y, \hat{y}) = -y \log_2(\hat{y}) - (1 - y) \log_2(1 - \hat{y})$$

CE

Here, $\hat{y} = P(y = 1|\mathbf{x})$.

Given that the true label y for a data point $\mathbf{x}$ is 1 and the predicted value $\hat{y}$ is 0.9, and the activation at the hidden layer is represented by $h_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, what is the value of $\frac{\partial L}{\partial W_{200}}$? Here, W_{200} denotes the weight connecting the first neuron of the hidden layer to the output layer neuron.

$y = 1$

$\hat{y} = 0.9$

$h_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a} \cdot \frac{\partial a}{\partial w} \rightarrow [h_1]^T$$

$$[\hat{y} - y] = 0.9 - 1 = -0.1$$

$$-0.1 \times [1 \ 2 \ 1]$$

$-0.1 \times 1 =$ -0.1

Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Range
Text Areas : PlainText
Possible Answers :

Sub-Section Number : 3
Sub-Section Id : 640653124104
Question Shuffling Allowed : Yes

Question Number : 46 Question Id : 640653829606 Question Type : MCQ
Correct Marks : 3
Question Label : Multiple Choice Question

If the weight vector w_t, obtained in the t'th iteration, misclassifies the next data point $\mathbf{x}$ in the perceptron algorithm and is subsequently updated to w_{t+1}, can we be certain that $\mathbf{x}$ will be classified correctly by w_{t+1}?

Options :

6406532786531. YES

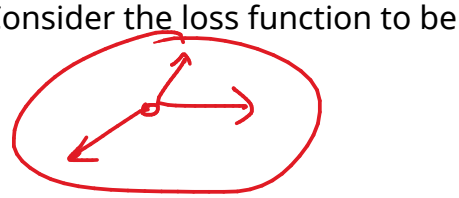
6406532786532. NO

Question Number : 47 Question Id : 640653829610 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

As per gradient descent, we should move towards 180 degrees with respect to gradient direction. What will happen if we move between 90 and 180 degrees? Consider the loss function to be convex.



Options :

6406532786542. The loss function will increase.

6406532786543. The loss function will decrease, although not to the maximum possible extent.

6406532786544. The loss function will remain the same.

6406532786545. Can't say. It depends on other parameters of the convex function.

Question Number : 48 Question Id : 640653829611 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following best describes the objective of gradient descent in relation to the Taylor series expansion of a function f(x) around a given point x_0?

Options :

6406532786546. Gradient descent aims to minimize the first-order term of the Taylor series expansion to approximate the function f(x) globally.

6406532786547. Gradient descent seeks to minimize the first-order term of the Taylor series expansion to efficiently navigate the local neighborhood around the given point x_0.

6406532786548. Gradient descent utilizes the entire Taylor series expansion to approximate the function f(x) and find its global minimum.

Sub-Section Number : 4

Sub-Section Id : 640653124105

Question Shuffling Allowed : Yes

Question Number : 49 Question Id : 640653829609 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Any boolean function of n inputs can be represented exactly by a network of perceptrons containing at least _____ hidden layer(s) with at least _____ perceptrons (each) and one output layer containing one perceptron.

Options :

6406532786538. n, n

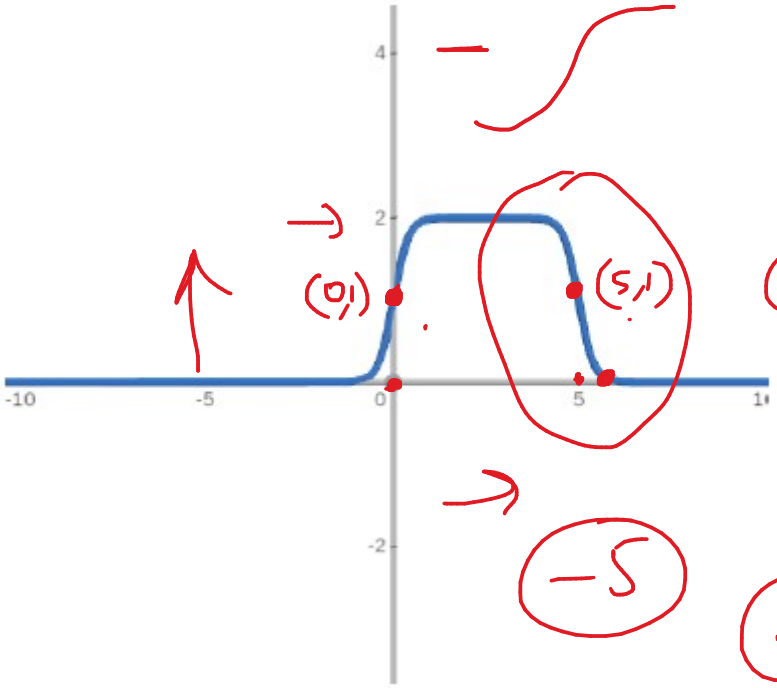
6406532786539. $n, 2^n$

6406532786540. $1, n$

6406532786541. $1, 2^n$

Sub-Section Number : 5
Sub-Section Id : 640653124106
Question Shuffling Allowed : Yes

Question Number : 50 Question Id : 640653829612 Question Type : MCQ
Correct Marks : 5
Question Label : Multiple Choice Question
Consider the plot shown below:



Which of the following could be the correct equation for this plot?

Options :

6406532786549. $f(x) = \frac{1}{1+e^{-5x+25}} - \frac{1}{1+e^{-5x}}$

6406532786550. $f(x) = \frac{2}{1+e^{-5x}} - \frac{2}{1+e^{-5x+25}}$

6406532786551. $f(x) = \frac{2}{1+e^{-5x+25}} - \frac{2}{1+e^{-5x}}$

6406532786552. $f(x) = \frac{1}{1+e^{-5x}} - \frac{1}{1+e^{-5x+25}}$

Sub-Section Number : 6

Sub-Section Id : 640653124107

Question Shuffling Allowed : Yes

Question Number : 51 Question Id : 640653829608 Question Type : MSQ

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the MP neuron model and its applicability to representing boolean functions. Select the correct statements:

Options :

6406532786534. The MP neuron model can represent a wide range of boolean functions (not all) by appropriately adjusting its weights and thresholds.

6406532786535. The MP neuron model can approximate arbitrary boolean functions, including non-linear ones.

6406532786536. The MP neuron model can accurately represent the XOR function by adjusting its weights and thresholds.

6406532786537. The representation power of the MP neuron model increases when multiple neurons are combined in a network architecture.

Sub-Section Number : 7

Sub-Section Id : 640653124108

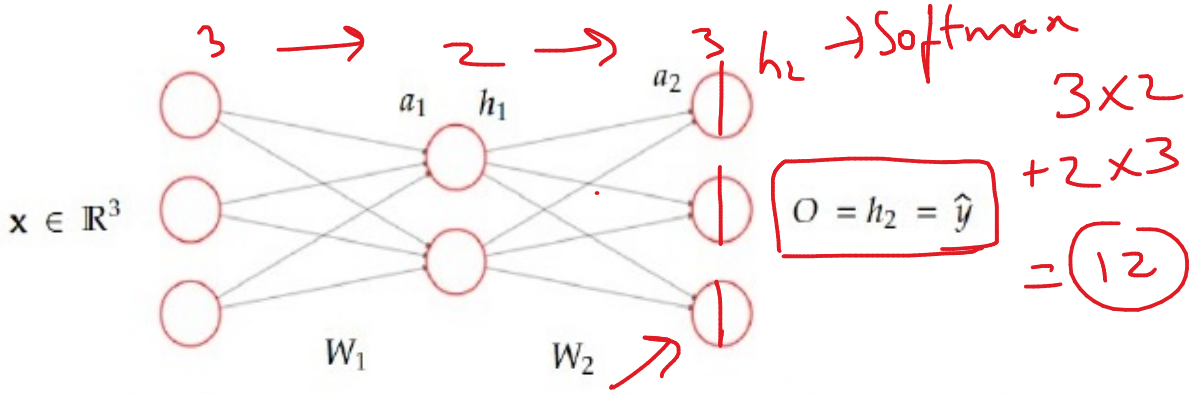
Question Shuffling Allowed : No

Question Id : 640653829614 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix

Question Numbers : (52 to 54)

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 $\mathbf{W}_1$ 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 : 52 Question Id : 640653829615 Question Type : SA
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 :

12

Question Number : 53 Question Id : 640653829616 Question Type : SA
Correct Marks : 2
Question Label : Short Answer Question

The true label for a data point $\mathbf{x}$ is given by $\begin{bmatrix}0\\1\\0\end{bmatrix}$. Assume that $h_1 = \begin{bmatrix}0.5\\1\end{bmatrix}$ and $W_2 = \begin{bmatrix}1 & 0\\1 & 1\\1 & 0\end{bmatrix}$.

If the network has no bias, find the loss value.

Use natural logrithm for loss calculation.

Enter your answer correct to two decimal places.

$a_2 = W_2 \times h_2$

$= \begin{bmatrix}1 & 0\\1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}0.5\\1\end{bmatrix}$

$\Rightarrow a_2 = \begin{bmatrix}0.5\\1.5\\0.5\end{bmatrix}$

$\Rightarrow h_2 = \begin{bmatrix}0.212\\0.576\\0.212\end{bmatrix} = \hat{y} = O_2$

Response Type : Numeric $L = -\ln(0.576) = 0.552$
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Range
Text Areas : PlainText
Possible Answers :

Question Number : 54 Question Id : 640653829617 Question Type : SA
Correct Marks : 2
Question Label : Short Answer Question

The true label for a data point $\mathbf{x}$ is given by $\begin{bmatrix}0\\1\\0\end{bmatrix}$. Assume that $h_1 = \begin{bmatrix}0.5\\1\end{bmatrix}$ and $W_2 = \begin{bmatrix}1 & 0\\1 & 1\\1 & 0\end{bmatrix}$. If the network has no bias, compute the gradient of loss with respect to $\hat{y}$ and enter the sum of the elements.

$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}(-\ln(\hat{y})) = -\frac{1}{\hat{y}}$

$\Rightarrow \frac{\partial L}{\partial \hat{y}} = \frac{-1}{0.576}$

$= -1.736$

Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Range
Text Areas : PlainText
Possible Answers :

Sub-Section Number : 8
Sub-Section Id : 640653124109
Question Shuffling Allowed : Yes

Question Number : 55 Question Id : 640653829618 Question Type : MCQ
Correct Marks : 4
Question Label : Multiple Choice Question

Suppose you are training a machine learning model using mini-batch gradient descent. You start with a learning rate (η) of 0.1 and a batch size of 50. After every epoch, you decrease the learning rate by 10%. If the first epoch completes 200 mini-batch updates, what will be the learning rate at the beginning of the second epoch?

$\eta = 0.1$
$MB = 50$

$0.1 \times 0.9 = 0.09$

Options :
6406532786557. 0.09

6406532786558. 0.06

6406532786559. 0.001

6406532786560. 0.0001

Question Number : 56 Question Id : 640653829619 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

Assume you are using mini-batch gradient descent with a total dataset size of 1,000 and a batch size of 20. If you process 10,000 mini-batches, how many epochs have been completed?

Options :

6406532786561. 5 epochs

6406532786562. 10 epochs

6406532786563. 20 epochs

6406532786564. 200 epochs

$\eta = 1000 \quad mB = 20$

$mBE = \frac{1000}{20} = 50$

$\frac{10,000}{50} = 200$ Epochs.

Question Number : 57 Question Id : 640653829620 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

If the batch size in mini-batch gradient descent is increased from 32 to 128 while keeping all other parameters constant, what is the likely impact on the variance of the gradient estimates per update?

Options :

6406532786565. Increases the variance

6406532786566. Decreases the variance

6406532786567. No impact on the variance

6406532786568. Initially decreases then increases the variance

BGD MBGD SGD → Variance increases

128 ← 32

LOWEST HIGHEST

Question Number : 58 Question Id : 640653829621 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

What do contour lines on a contour plot represent?

GD = BGD → All pts.

Options :

6406532786569. The areas where the function value increases most rapidly

6406532786570. Lines connecting points where the function has the same output value

6406532786571. The maximum and minimum values of a function

6406532786572. The gradient of the function at those points

Data Viz

Section Id : 64065359427