In the planning graph, which of the following are mutex action pairs in Layer 1?

Options :

6406531515038. ✔ Putdown(D), Stack(D,A)

6406531515039. ✔ Putdown(D), Stack(D,C)

6406531515040. ✖ Putdown(D), Stack(D,B)

6406531515041. ✖ Putdown(D), Pickup(A)

6406531515042. ✖ Putdown(D), Unstack(C,B)

Question Number : 60 Question Id : 640653455618 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Selectable Option : 0

Question Label : Multiple Select Question

In the planning graph, which of the following are mutex proposition pairs in Layer 1?

Options :

6406531515043. ✔ onTable(D), on(D,A)

6406531515044. ✔ onTable(D), on(D,C)

6406531515045. ✔ onTable(D), holding(D)

6406531515046. ✖ onTable(B), clear(A)

6406531515047. ✖ on(C,B), on(C,D)

Deep Learning

Section Id :	64065329447
Section Number :	3
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	17
Number of Questions to be attempted :	17
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 :	64065365932
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 61 Question Id : 640653455622 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 SUBJECT TO BE WRITTEN.

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

Options :

6406531515050. ✔ YES

6406531515051. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065365933
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 62 Question Id : 640653455625 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

Consider two data points $\mathbf{x_1} = \begin{bmatrix} a \\ a \end{bmatrix}$ and $\mathbf{x_2} = -1 * \mathbf{x_1}$, where $a > 0$. The data point $\mathbf{x_1}$ belongs to positive class (denoted as 1) and the datapoint $\mathbf{x_2}$ belongs to 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,

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

The algorithm checks $\mathbf{x_1}$ in the first iteration and $\mathbf{x_2}$ in the second iteration and so on. How many times the weights get updated until convergence (That is, the algorithm classifies both the points correctly)? The weights do not include bias.

Options :

6406531515061. ✔ 1

6406531515062. ✖ 2

6406531515063. ✖ 4

6406531515064. ✖ It oscillates and never converges

Question Number : 63 Question Id : 640653455633 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

A team has a data set that contains 10000 samples for training a feed forward neural network. Suppose they decided to use the gradient descent algorithm to update the weights. Suppose further that they use line search algorithm for the learning rate as follows, $\eta = [0.01, 0.1, 1, 2, 10]$. How many times do the weights get updated after training the network for 10 epochs? (Note, for each weight update the loss has to decrease)

Options :

6406531515078. ✖ 50

6406531515079. ✖ 500000

6406531515080. ✖ 100000

6406531515081. ✔ 10

Sub-Section Number : 3

Sub-Section Id : 64065365934

Question Shuffling Allowed : Yes

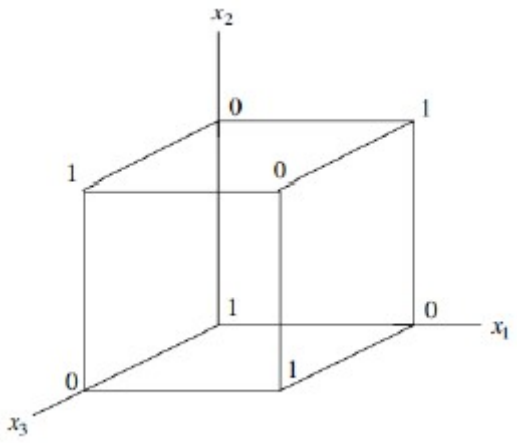
Is Section Default? : null

Question Number : 64 Question Id : 640653455624 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

The figure below represents a Boolean function using a unit cube. Each vertex shows the label or output of the function for the given input. Which of the followings is (are) true about the function?



Options :

6406531515056. ✔ The function has 8 inputs

6406531515057. ✖ The function has 12 inputs

6406531515058. ✖ The function has 24 inputs

6406531515059. ✖ The function is linearly separable

6406531515060. ✔ The function is not linearly separable

Sub-Section Number : 4

Sub-Section Id : 64065365935

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 65 Question Id : 640653455623 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following threshold θ values of MP neuron implements OR Boolean function denoted by $f(\mathrm{x})$? Assume that the number of inputs x_i to the neuron is four and the neuron does not have any inhibitory inputs.

$$f(\mathrm{x}) = \begin{cases} 1, & \text{if } \sum_{i=0}^{3} x_i > \theta \\ 0, & otherwise \end{cases}$$

Options :

6406531515052. ✔ 0

6406531515053. ✖ 1

6406531515054. ✖ 2

6406531515055. ✖ 3

Question Number : 66 Question Id : 640653455626 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Consider a sigmoid function

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

Suppose that w is restricted to take only negative 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 :

6406531515065. ✖ Increasing the value of *b* shifts the sigmoid function to the left (i.e., towards negative infinity)

6406531515066. ✔ Increasing the value of *b* shifts the sigmoid function to the right (i.e., towards positive infinity)

6406531515067. ✖ Increasing the value of *w* increases the steepness of the sigmoid function

6406531515068. ✔ Increasing the value of *w* decreases the steepness of the sigmoid function

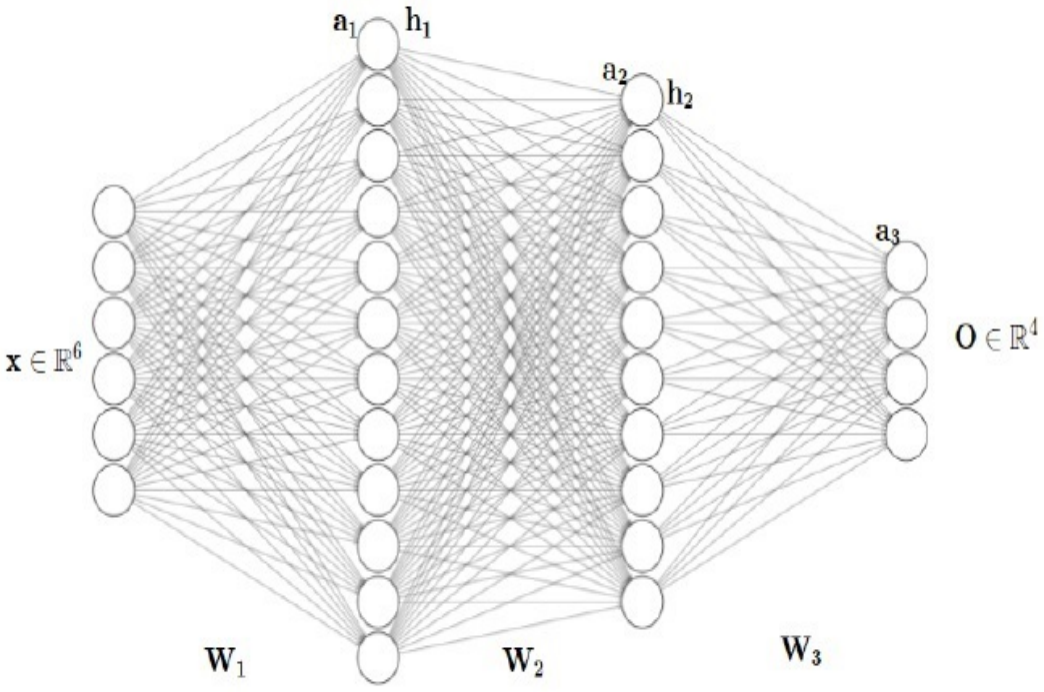
Question Number : 67 Question Id : 640653455632 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Suppose that a momentum based gradient descent algorithm is used to train the feed forward neural network shown below.
The parameter update rule for the weight matrix $\mathbf{W_3}$ at time step t is written as follows,

$$u_3^t = \beta u_3^{t-1} + (1-\beta)\nabla a_3^t (h_2^t)^T$$



$$W_3^{t+1} = W_3^t - \eta u_3^t$$

where, $0 \leq \beta < 1$.Choose the correct statements about the update rule

Options :

6406531515073. ✔ At any given time step u_3^t is a matrix

6406531515074. ✖ At any given time step u_3^t is a vector

6406531515075. ✖ The update rule for u_3^t is wrong, the second term should have been written as ∇W_3^t

6406531515076. ✔ It requires bias correction to avoid bias towards zero

6406531515077. ✖ It does not require any bias correction

Question Number : 68 Question Id : 640653455634 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

The update rule for the ADAM (Adaptive Moments) optimization algorithm is given below,

$$m_t = \beta_1 m_{t-1} + \nabla w_t$$
$$v_t = \beta_2 v_{t-1} + (\nabla w_t)^2$$
$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}$$
$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}$$
$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}}\hat{m}_t$$

where, $\hat{m}_t, \hat{v}_t$ are bias corrected version of m_t, v_t. Here, $0 \leq \beta_1 < 1$ and $0 \leq \beta_2 < 1$. Choose the correct statement(s) from the following statements.

Options :

6406531515082. ✔ m_t does not require bias correction

6406531515083. ✔ v_t does not require bias correction

6406531515084. ✔ m_t and v_t are exponential averaging of history of gradients and gradients squared, respectively

6406531515085. ✔ Both β_1 and β_2 can take values greater than one and in such a case the past gradients are given more weightage than the current gradients

Question Number : 69 Question Id : 640653455635 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Select all correct statements about the bias and variance of deep learning models.

Options :

6406531515086. ✔ A model with high bias produce high training error and high validation error

6406531515087. ✖ A model with high variance produce high training error and low validation error

6406531515088. ✔ A model with high variance produce low training error and high validation error

6406531515089. ✔ Bias and variance can be both made zero if the underlying distribution of the samples are known and also the samples are not corrupted by noise

Question Number : 70 Question Id : 640653455636 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Suppose that we use Dropout scheme to regularize the deep learning model. Assume that the dropout probability p = 0.75 for all the layers in the model and the model is trained using Stochastic Gradient Descent (SGD) optimizer. Choose the correct statements about Dropouts regularization.

Options :

6406531515090. ✖ Dropout reduces (removes) 25 percent of weights from the model during

training and inference

6406531515091. ✖ For each iteration, 75% of neurons are dropped randomly from the model

6406531515092. ✔ For each iteration, 25% of neurons are dropped randomly from the model

6406531515093. ✔ Each weight gets multiplied by a factor *p* during inference

6406531515094. ✖ Each weight gets multiplied by a factor 1 − *p* during inference

6406531515095. ✖ The Droput method can not be combined with other regularization methods

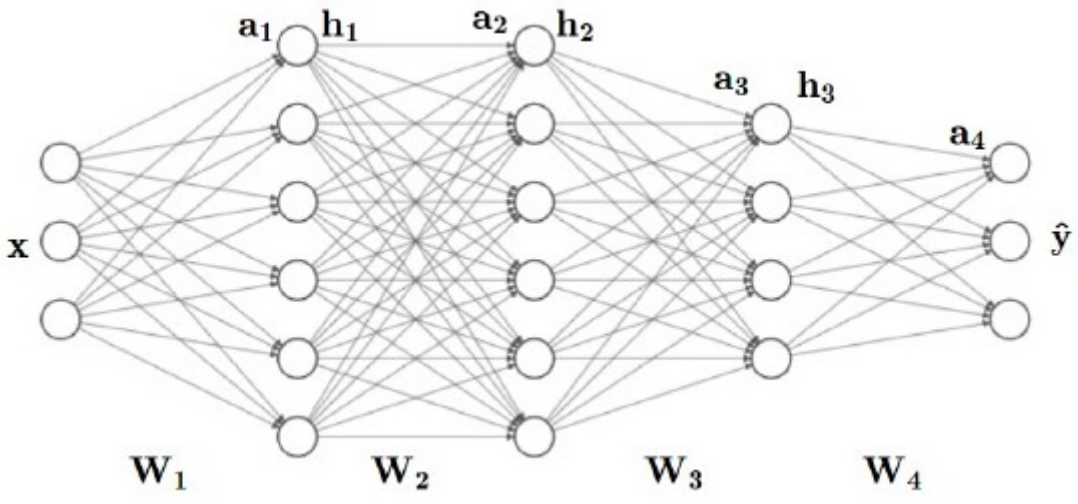
Sub-Section Number :	5
Sub-Section Id :	64065365936
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653455627 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 : (71 to 74)

Question Label : Comprehension

The diagram below shows a feed forward neural network with fully connected layers (not to scale)



Based on the above data, answer the given subquestions.

Sub questions

Question Number : 71 Question Id : 640653455628 Question Type : SA Calculator : None

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

Correct Marks : 1

Question Label : Short Answer Question

Suppose that $\mathbf{x} \in \mathbb{R}^{10\times1}$, $\mathbf{h_1} = \mathbf{h_2} \in \mathbb{R}^{100\times1}$, $\mathbf{h_3} \in \mathbb{R}^{50\times1}$ and $\mathbf{y} \in \mathbb{R}^{3\times1}$. How many parameters **(including bias)** are there in the network? (If your answer is 12345, enter it as 12345)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

16402 to 16404

Question Number : 72 Question Id : 640653455629 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

Suppose that a softmax function is used in the output layer to compute the probability for each class.

$$\hat{\mathbf{y}} = softmax(\mathbf{a_4})$$

Assume that $\mathbf{a_4} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$. What is the value of third element in $\hat{\mathbf{y}}$?(if your intermediate result is 1.2345, then take it as 1.23 to the next step)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.23 to 0.25

Question Number : 73 Question Id : 640653455630 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

Suppose that the number of classes is 1000, that is, $\hat{\mathbf{y}} \in \mathbb{R}^{1000\times 1}$. Suppose further that all the elements in $\mathbf{a_4}$ is 1. Compute the $\hat{\mathbf{y}}$ as follows and enter the sum of the elements in $\hat{\mathbf{y}}$

$$\hat{\mathbf{y}} = softmax(\mathbf{a_4})$$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.9 to 1.1

Question Number : 74 Question Id : 640653455631 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

Suppose the dimension of $\mathbf{h_3}$ is $\mathbb{R}^{4\times1}$ and $\mathbf{a_4}$ is $\mathbb{R}^{3\times1}$. Suppose further that the network uses sigmoid (logistic) neurons in the output layer. The output $\mathbf{h_3} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{a_4} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$. The true label $\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

The loss is calculated as follows,

$$L = \frac{1}{2}\sum_{k=1}^{3}(\hat{y}_k - y_k)^2$$

Compute the gradient matrix $\frac{\partial L}{\partial \mathbf{W_4}}$ and enter the sum of the elements of the matrix.
(Truncate your result to 2 decimal points)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.3 to 0.36

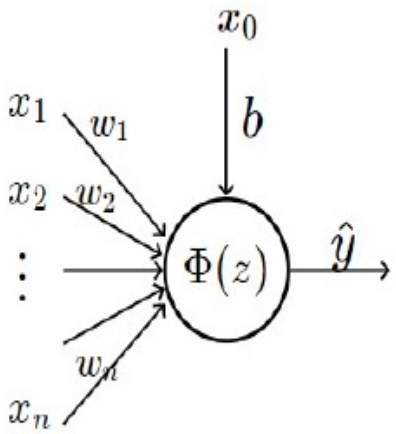
Sub-Section Number :	6
Sub-Section Id :	64065365937
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653455637 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 : (75 to 77)

Question Label : Comprehension

Consider a single neuron with the activation function $\Phi(\cdot)$ as shown in the figure below. Here, b denotes the bias, z denotes the input to the activation function and $\hat{y}$ is the output from the neuron. For all the calculations, take the true label $y = 0$ and clip the gradient



values to zero if the value is smaller than 10^{-5}. **Truncate** the intermediate results to 2 decimal digits (that is, if the intermediate output is 0.23456, take it as 0.23 to the next step).

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 75 Question Id : 640653455638 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

Suppose that all the elements in the input are one (that is, $x_0 = x_1 = \cdots = x_n = 1$) and all the weight values are 0.1 (that is, $b = w_1 = w_2 = \cdots = w_n = 0.1$). Assume the activation function is logistic, $\Phi(z) = \frac{1}{1+exp(-z)}$ and $n = 10$. Use the Mean square error loss ,

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

enter the number of weights that will get modified from their current values during the next iteration.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

10.9 to 11.1

Question Number : 76 Question Id : 640653455639 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

From the previous question, if we change all the weight values to 10, that is, $b = w_1 = w_2 = \cdots = w_n = 10$, then enter the number of weights that will get modified during the next iteration.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.00 to 0.01

Question Number : 77 Question Id : 640653455640 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

From the previous question, if we change all the weight values to 10, that is, b = w_1 = w_2 = · · · = w_n = 10 and also the activation φ(·) to Relu, then enter the number of weights that will get modified during the next iteration.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

10.9 to 11.1

Question Id : 640653455641 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 : (78 to 81)

Question Label : Comprehension

An MRI image of size 250 × 250 × 8 is feed as an input to a convolutional layer. Suppose that we want to produce an output of size 60 × 60 × 64 using a set of kernel. Suppose the kernel size is 13 × 13. Answer the given subquestions

Sub questions

Question Number : 78 Question Id : 640653455642 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Short Answer Question

What is the value of stride s?. Assume zero padding ($p = 0$)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

3.9 to 4.1

Question Number : 79 Question Id : 640653455643 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

How many parameters (including bias) are there in the convolutional layer?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

86591 to 86593

Question Number : 80 Question Id : 640653455644 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

Suppose that the input image is of size 250 × 250 × 3 × 8. Suppose further the kernel size is 13 × 13. How many parameters (including bias) are there in the layer now?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

259647 to 259649

Question Number : 81 Question Id : 640653455645 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

The statement that applying max pooling with stride 2 ($s = 2$) reduces the number of parameters by half is

Options :

6406531515102. ✖ TRUE

6406531515103. ✔ FALSE

Sub-Section Number :	7
Sub-Section Id :	64065365938
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 82 Question Id : 640653455646 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Short Answer Question

Consider the sentence " Expert is the one who knows more and more about the less and less ". What is the size of the vocabulary V ?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

10

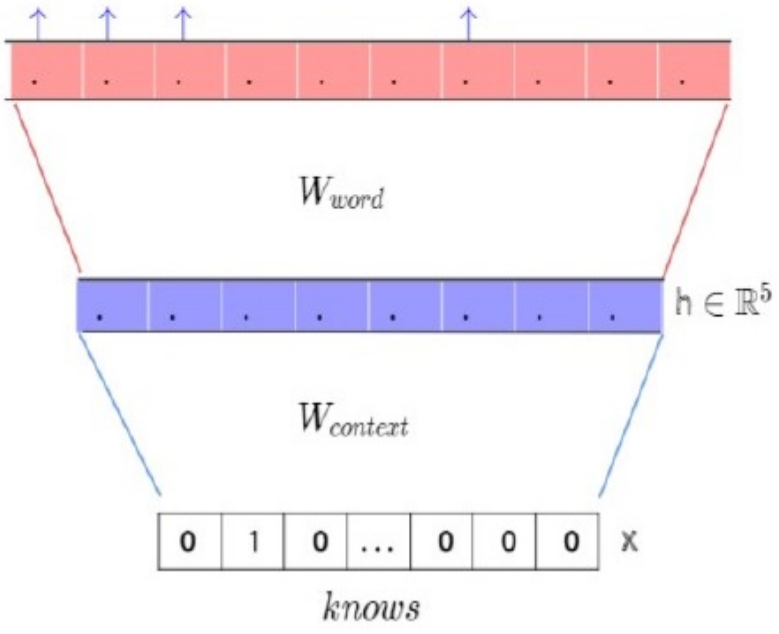
Sub-Section Number :	8
Sub-Section Id :	64065365939
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653455647 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 : (83 to 85)

Question Label : Comprehension

Suppose we use the CBOW (Continuous Bag of words) model shown below to find a distributed vector representation for all the words in the vocabulary. The size of the vocabulary $|V|$ is 15 and all the words in the vocabulary are one hot encoded **(as a column vector)** and fed as input to the network. The output layer uses softmax to produce the probability score for each word given the context. Here, $\mathbf{W}_{word}$ and $\mathbf{W}_{context}$ are weight matrices.



Based on the above data, answer the given subquestions.

Sub questions

Question Number : 83 Question Id : 640653455648 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Short Answer Question

How many number of parameters are there in the network (excluding bias)?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

149.5 to 150.5

Question Number : 84 Question Id : 640653455649 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

The word representation for the 2-nd word in the vocabulary corresponds to

Options :

6406531515106. ✔ 2-nd column of $\mathbf{W_{context}}$

6406531515107. ✖ 2-nd row of $\mathbf{W_{context}}$

6406531515108. ✖ 2-nd column of $\mathbf{W_{word}}$

6406531515109. ✖ 2-nd row of $\mathbf{W_{word}}$

Question Number : 85 Question Id : 640653455650 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

The prediction $\hat{y}_w$ is given by

$$\hat{y} = \frac{exp(u_c \cdot v_w)}{\sum_{w' \in V} exp(u_c \cdot v_{w'})}$$

where $u_c \cdot v_w$ is a dot product between a vector from $W_{context}$ and a vector from W_{word}. Then the statement that the neurons in the hidden layer do not have any non-linear activation function is

Options :

6406531515110. ✔ TRUE

6406531515111. ✖ FALSE

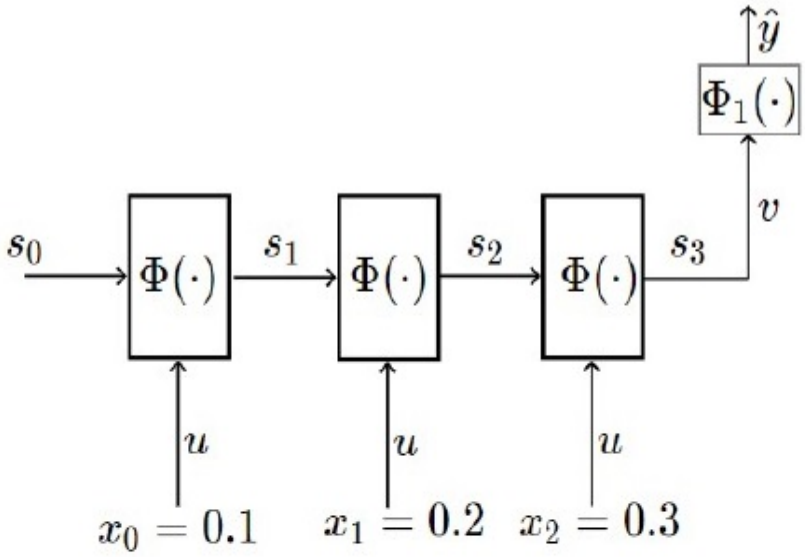
Sub-Section Number :	9
Sub-Section Id :	64065365940
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653455651 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 : (86 to 87)

Question Label : Comprehension

Consider a Recurrent Neural network shown below



where $\Phi(\cdot) = tanh(\cdot)$ and $\Phi_1(\cdot) = \sigma(\cdot)$(logistic function).The state vector is computed as follows.

$$s_t = tanh(ws_{t-1} + ux_t)$$

the initial weight values are as follows

$$u = 2, w = -2, v = 1, s_0 = 0$$

and the true label $y = 1$.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 86 Question Id : 640653455652 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 mean square error loss

$L = 0.5 * (\hat{y} - y)^2$of the network?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.065 to 0.075

Question Number : 87 Question Id : 640653455653 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

Update the parameter v using plain SGD (stochastic gradient descent) algorithm with $\eta = 1$. Enter the updated value of v.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

1.06 to 1.1

Sub-Section Number :	10
Sub-Section Id :	64065365941
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 88 Question Id : 640653455654 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following statements is(are) true about the vanilla RNN ?

Options :

6406531515114. ✖ It can’t have more than one hidden layer in it

6406531515115. ✔ It is prone to vanishing gradient problem

6406531515116. ✔ It is prone to exploding gradient problem

6406531515117. ✖ Vanishing gradient problem can be solved if ReLU activation is used across the network

Sw Testing

Section Id :	64065329448
Section Number :	4
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	18
Number of Questions to be attempted :	18
Section Marks :	100
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 :	64065365942
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 89 Question Id : 640653455655 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:SOFTWARE TESTING"