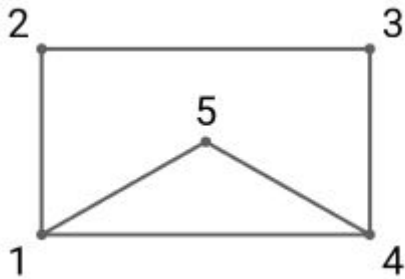For the 2D line drawing, assign consistent labels to all edges and junctions. Enter the labels of the junctions 1, 2, 3, 4 in the text box, in that order. Or enter NIL if the drawing has no consistent label assignment.

Enter a comma separated list of junction labels, or enter NIL.

NO SPACES, TABS, DOTS, BRACKETS OR EXTRANEOUS CHARACTERS.

Answer format: X9,Y9,Z9,W9

Response Type : Alphanumeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Answers Case Sensitive : No

Text Areas : PlainText

Possible Answers :

NIL

Deep Learning

Section Id :	64065338352
Section Number :	5
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	13
Number of Questions to be attempted :	13
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 :	64065380600
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 107 Question Id : 640653564731 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 <u>TOP</u> FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406531887677. ✔ YES

6406531887678. ✖ NO

|	
Sub-Section Number :	2
Sub-Section Id :	64065380601
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 108 Question Id : 640653564732 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

Suppose that we implement a three input Boolean function using the Mc-Culloch Pitts (MP) neuron. The graph below shows the Number of Correctly Classified (NCC) data points for various values of threshold θ. The threshold θ is incremented by 1 from 0 to 5. Assume that the neuron does not have any inhibitory input. This graph represents which of the following Boolean functions?

$$\hat{y} = \begin{cases} 1, & \text{if } \sum x_i \geq \theta \\ 0, & \text{otherwise} \end{cases}$$
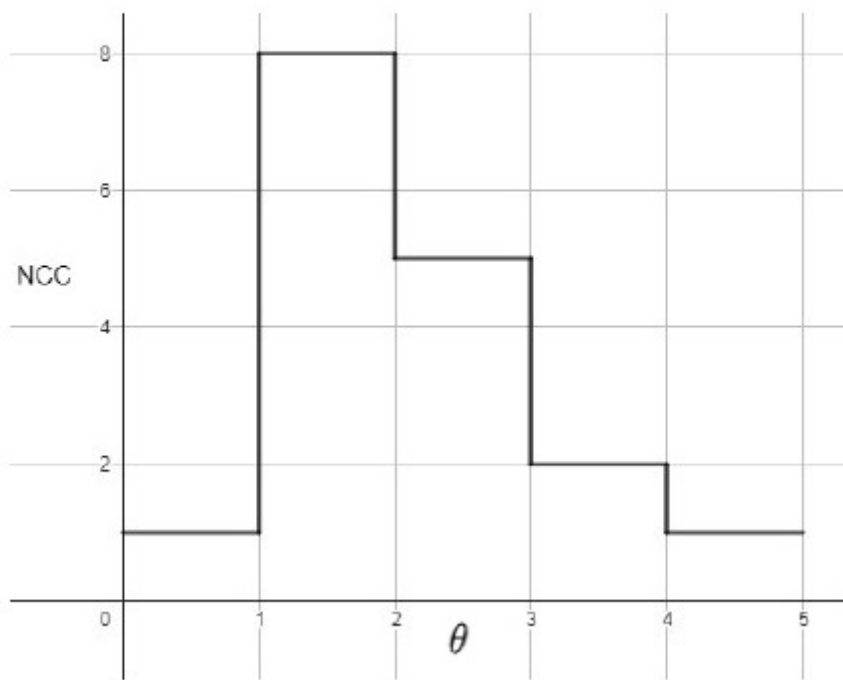
Options :

6406531887679. ✹ NOR

6406531887680. ✹ AND

6406531887681. ✔ OR

6406531887682. ✹ NAND

6406531887683. ✹ None of these

Question Number : 109 Question Id : 640653564743 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 two sentences

• A man was sitting at the bank of the river and gazing at stars in the sky

• A man went to the bank to check his current balance

Suppose we get the word representation for the word **bank** in both sentences using CBOW model which was trained as shown in the image below. The model was trained by building a vocabulary that contains unique words in the sentences. Then the statement that the word representation for the word **bank** will be different based on its context is
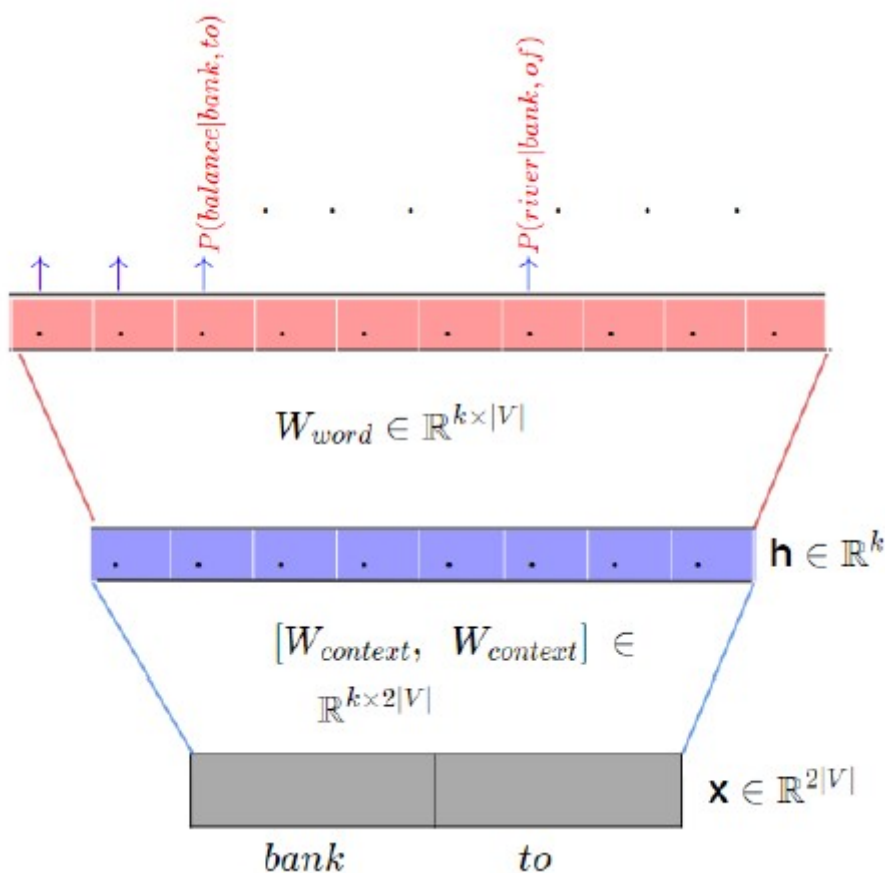
Options :

6406531887707. �֍ TRUE

6406531887708. ✔ FALSE

6406531887709. ✖ Insufficient information

Sub-Section Number : 3

Sub-Section Id : 64065380602

Question Shuffling Allowed : Yes

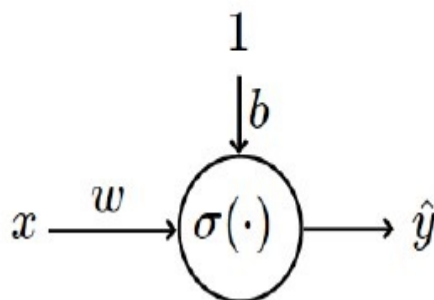**Is Section Default? :** null

Question Number : 110 Question Id : 640653564733 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

Consider a sigmoid neuron shown below. The input to the neuron is a real number. Suppose that input $x = 10$, output $y = 1$ and the parameters, $w = 0.1, b = 0.1$. Update the parameters once using GD with $\eta=1$. Enter the sum of updated parameter values. The loss function is $L = \frac{1}{2}(y - \hat{y})^2$.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.68 to 0.72

Question Number : 111 Question Id : 640653564744 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

A CNN model takes an input image of size 100×100. The first convolution layer uses 30 kernels(filters) each of size 3×3, the output from the first convolution layer is passed as an input to the second convolution layer. The second convolution layer uses 20 kernels each of size 5×5. How many parameters (including bias) are there in the network? Assume stride=1,padding=0 if required.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText
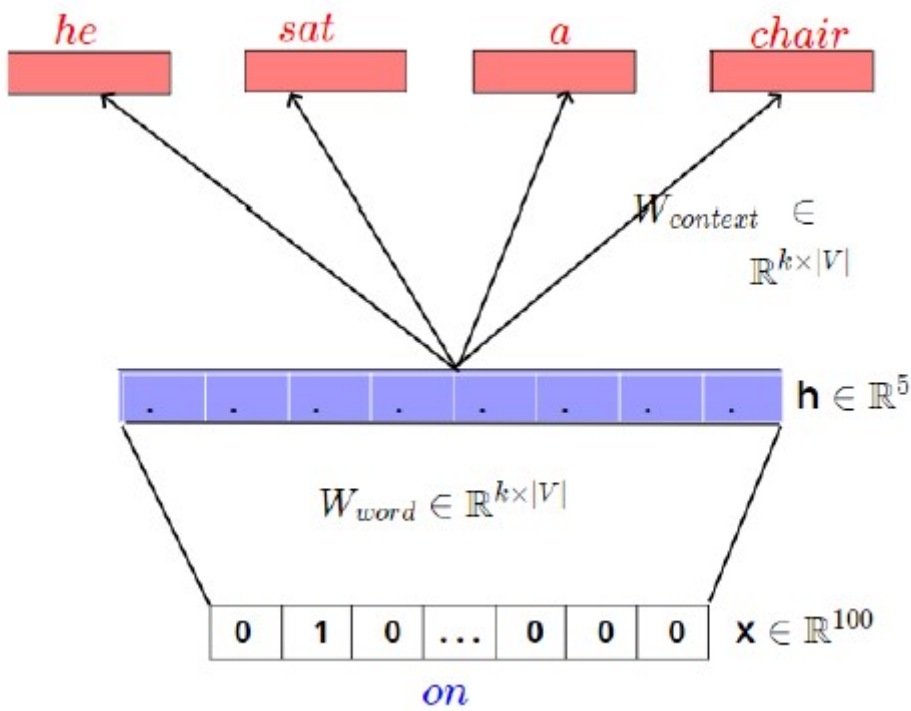
Possible Answers :

15320

Question Number : 112 Question Id : 640653564745 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

Consider a skip-gram model shown below. Each word in the vocabulary is represented as one-hot vector of size 100×1. The embedding dimension h is 5×1. Enter the number of parameters (exclude bias) in the entire network

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2500

Sub-Section Number :	4
Sub-Section Id :	64065380603
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653564734 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 115)

Question Label : Comprehension

A neural network contains an input layer $h_0 = x$, five hidden layers $(h_1, h_2, \cdots, h_5)$ and an output layer O. All the hidden layers use $Relu$ activation and the output layer uses softmax activation.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 113 Question Id : 640653564735 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 the input $x \in \mathbb{R}^{900}$ and all the hidden layers contains 10 neurons each. The output layer contains 20 neurons. How many parameters 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 :

9670

Question Number : 114 Question Id : 640653564736 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 that all the 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 : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

2.9 to 3.1

Question Number : 115 Question Id : 640653564737 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

Assume that we use backpropagation to compute the gradient of the loss w.r.t. all the parameters. Then the statement that the gradients w.r.t. all the parameters are zero is

Options :

6406531887687. ✳ TRUE

6406531887688. ✔ FALSE

6406531887689. ✳ Need more information to validate the statement

|	
Sub-Section Number :	5
Sub-Section Id :	64065380604
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 116 Question Id : 640653564738 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 plot below shows contours of a function $f(w, b)$. Choose the correct statements
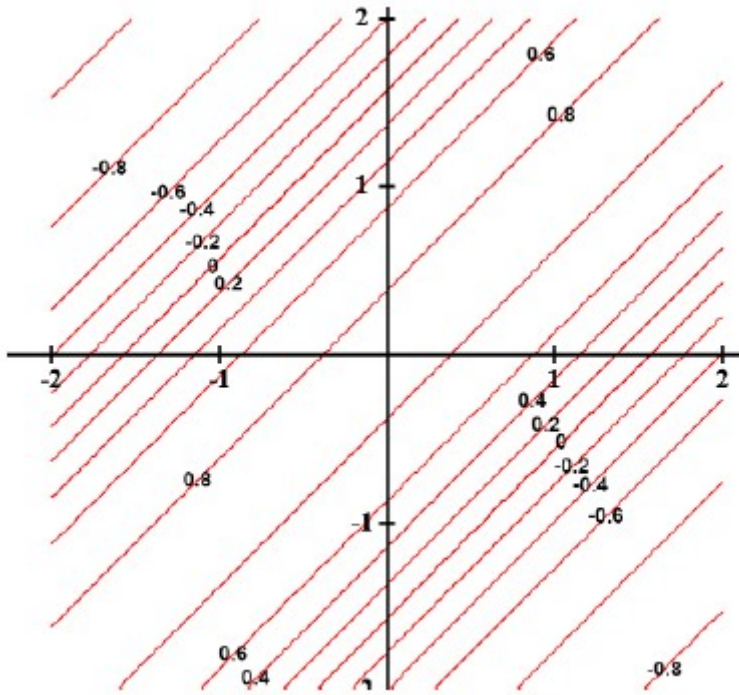
Options :

6406531887690. ✔ There might be two flat minima

6406531887691. ✖ There might be two flat maxima

6406531887692. ✔ There is one flat maxima

6406531887693. ✖ There is one flat minima

Question Number : 117 Question Id : 640653564739 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

Suppose that a model produces zero training error without using any regularization technique. What happens if we use L_2 regularization and retrain the model, in general?

Options :

6406531887694. ✔ This might increase training error

6406531887695. ✔ This might decrease test error

6406531887696.

✔ Reduce the complexity of the model by driving less important weights to close to zero

6406531887697. ✖ This might decrease training error

Question Number : 118 Question Id : 640653564750 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

Suppose we divide the available training samples into mini batches of size 32 to train a model with mini-batch gradient descent. Assume that we have 33 different machines to train the model. One out of 33 machines acts as a master machine. The actual weight update happens in the master machine. The master machine can send one sample for the rest of the machines along with a copy of the model in its current state to compute the gradients. We call this entire set-up parallelization. Which of the following deep learning architectures can be trained in parallel then?

Options :

6406531887716. ✔ Fully connected Feed forword neural network

6406531887717. ✔ Convolutional Neural network

6406531887718. ✔ Reccurent Neural Network

6406531887719. ✔ Transformers

Sub-Section Number :	6
Sub-Section Id :	64065380605
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653564740 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 : (119 to 120)

Question Label : Comprehension

A text corpus contained the following two sentences.

• In science if you know what you are doing you should not be doing it

• In engineering if you do not know what you are doing you should not be doing it

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 119 Question Id : 640653564741 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

Build a vocabulary V and enter its size $|V|$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

14

Question Number : 120 Question Id : 640653564742 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

Suppose we build a co-occurrence matrix of size $m \times n$, where each row corresponds to a word in the vocabulary and the columns corresponds to the context of the word. Which of the following could be a valid size of the co-occurrence matrix (select all correct answers)?

Options :

6406531887699. ✖ 32 × 32

6406531887700. � 32 × 14

6406531887701. ✱ 14 × 32

6406531887702. ✔ 14 × 14

6406531887703. ✔ 14 × 13

6406531887704. ✱ 13 × 14

6406531887705. ✔ 14 × 7

6406531887706. ✱ 7 × 14

Sub-Section Number :	7
Sub-Section Id :	64065380606
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653564746 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 : (121 to 122)

Question Label : Comprehension

Consider a simple RNN for a binary sequence classification problem.
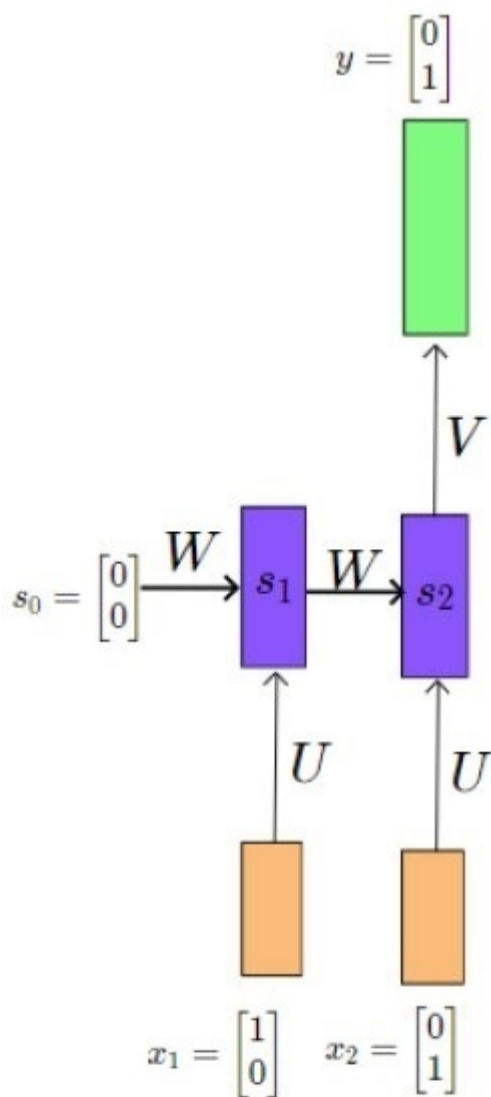
Suppose the weight matrices U, V, W are initialized as follows

$$W = U = V \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

The state vector s_t is computed as follows

$$s_t = tanh(Ux_t + Ws_{t-1})$$

Note: In all your calculations, consider only the first two decimal places of any number (input, intermediate results..). That is, if the number is -1.0234, take it as -1.02.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 121 Question Id : 640653564747 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

What is the loss value? Use cross entropy loss with natural logarithm.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

2 to 2.1

Question Number : 122 Question Id : 640653564748 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

Compute the gradients for the weight matrix, that is ∇V and enter the sum of its (main) diagonal elements?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.80 to 0.92

Sub-Section Number :	8
Sub-Section Id :	64065380607
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 123 Question Id : 640653564749 Question Type : MCQ Is Question

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the statement "the attention coefficients in RNN based Encoder- Decoder architecture are computed for each time step of the decoder". The statement is

Options :

6406531887714. ✔ TRUE

6406531887715. ✖ FALSE

PSM

Section Id :	64065338353
Section Number :	6
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	5
Number of Questions to be attempted :	5
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 :	64065380608
Question Shuffling Allowed :	No
Is Section Default? :	null