

Y2,W3,L5,T4

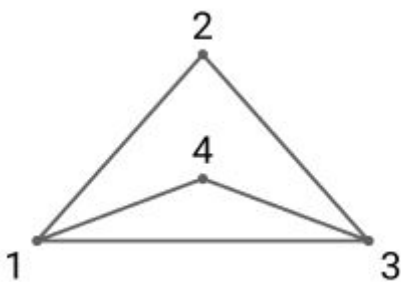
Question Number : 107 Question Id : 640653566988 Question Type : SA Calculator : None

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

Correct Marks : 1

Question Label : Short Answer Question

For the 2D line drawing, assign consistent labels to all edges and junctions. Enter the labels of the junctions 1, 2, 3 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 EXTRANEIOUS CHARACTERS.

Answer format: X9,Y9,Z9

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 :

64065338432

Section Number :	5
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	14
Number of Questions to be attempted :	14
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 :	64065381160
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 108 Question Id : 640653566989 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 :

6406531894836. ✓ YES

6406531894837. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065381161
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 109 Question Id : 640653566990 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 implement AND boolean function using McCulloch Pitts (MP) Neuron. The neuron takes in 10 inputs x_0, x_1, \dots, x_9 and one additional inhibitory input x_{10} . The inhibitory input is by default set to 0. A correct value of θ is one that classifies all the input data points correctly.

$$\hat{y} = \begin{cases} 1, & \text{if } \sum_{i=0}^9 x_i \geq \theta \\ 0, & \text{otherwise} \end{cases}$$

choose the correct statement(s) from the following statements.

Options :

6406531894838. ✔ Setting $\theta = 10$ divides the whole input space R^{10} into two regions and therefore makes zero classification error

6406531894839. ✖ Setting $\theta = 10$ divides the whole input space R^{10} into 10 regions and therefore makes zero classification error

6406531894840. ✔ Setting inhibitory input to 1 doesn't divide the whole space R^{10} and makes only one misclassification

6406531894841. ✖ Setting inhibitory input to 1 divides the whole space R^{10} into two and makes only one misclassification

Sub-Section Number : 3
Sub-Section Id : 64065381162
Question Shuffling Allowed : No
Is Section Default? : null

Question Id : 640653566992 **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 : (110 to 112)

Question Label : Comprehension

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

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 110 **Question Id :** 640653566993 **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 contain 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 : 111 Question Id : 640653566994 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 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 : 112 Question Id : 640653566995 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 and hence none of the parameters get updated is

Options :

6406531894848. ✖ TRUE

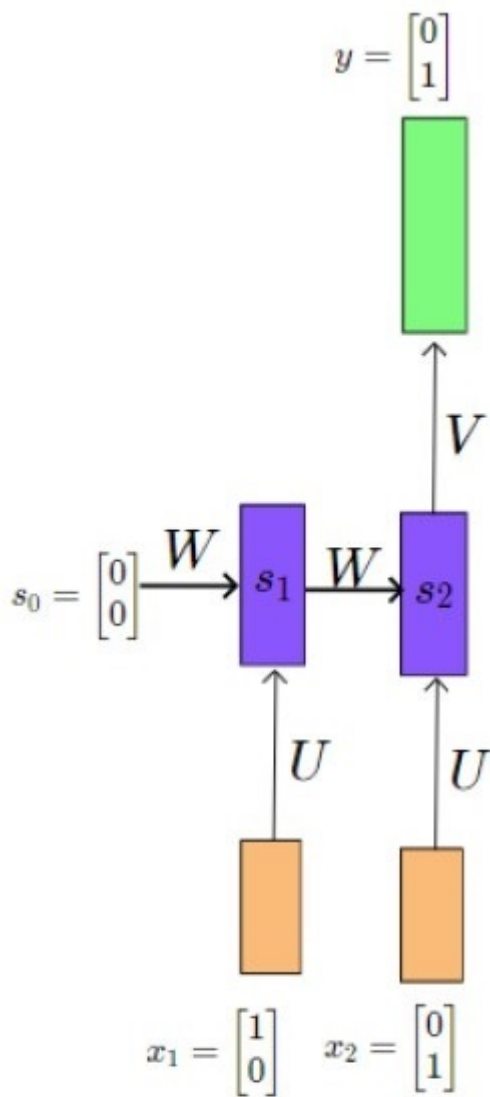
6406531894849. ✔ FALSE

6406531894850. ✖ Need more information to validate the statement

**Question Id : 640653567005 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 114)**

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 (such as inputs, 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 : 113 Question Id : 640653567006 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

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 :

0.008 to 0.018

Question Number : 114 Question Id : 640653567007 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 :

0.01 to 0.03

Sub-Section Number : 4

Sub-Section Id : 64065381163

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 115 Question Id : 640653566997 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

Which of the following statements are true according to the following equation

$$E((y - \hat{f}(x))^2)$$

where y is an observed value and $\hat{f}(x)$ is an estimate (predicted value).

Options :

6406531894852. ✓ The training error can be made zero with sufficiently complex models

6406531894853. ✖ The training error cannot be made zero even with sufficiently complex models due to the presence of irreducible error

6406531894854. ✖ The test error can be made zero with sufficiently complex models

Sub-Section Number :	5
Sub-Section Id :	64065381164
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 116 Question Id : 640653566991 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

Consider a sigmoid neuron shown in the figure A, The equation corresponding to plot in the figure A is

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

for some values of $w \in \mathbb{R}$ and $b \in \mathbb{R}$. The figure B is obtained by changing (increase/decrease) the values of w and b to get the sigmoid as in figure B. To obtain the plot in figure B,

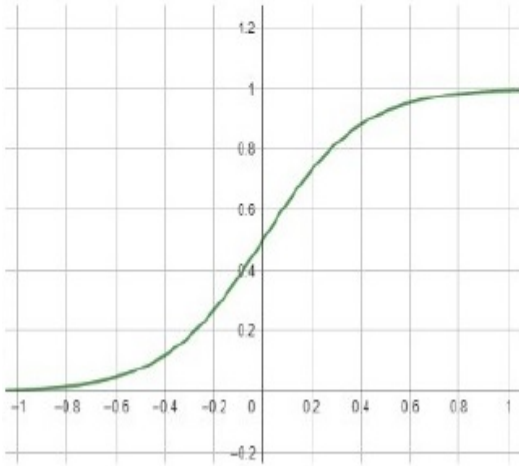


Fig A

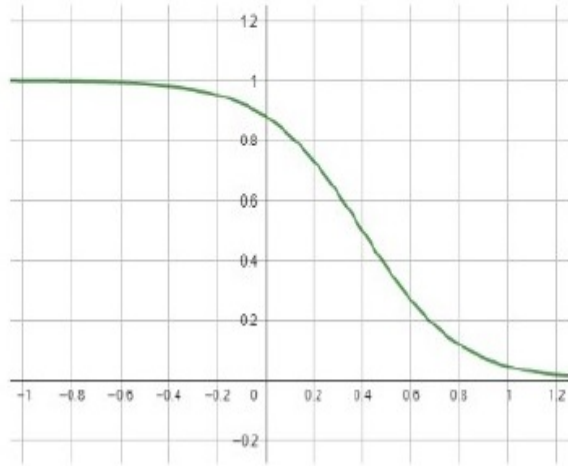


Fig B

Options :

6406531894842. ✓ w has to be decreased

6406531894843. ✗ w has to be increased

6406531894844. ✓ b has to be increased

6406531894845. ✗ b has to be decreased

Question Number : 117 Question Id : 640653566998 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 diagram below shows contours of a loss function $\hat{L}(\theta)$ where,
 $\hat{L}(\theta) = L(\theta) + \alpha\Omega(\theta)$.

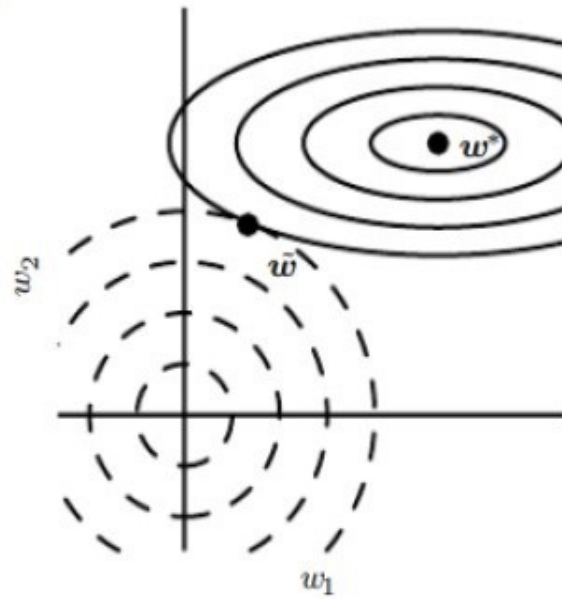


Figure 1: Contours

Here, $L(\theta)$ is an un-regularized loss function and $\Omega(\theta)$ is a regularization term. Suppose we use L_2 regularization, then choose the correct statements from the following statements (with respect to this diagram)

Options :

- 6406531894855. ✓ The dotted circles represent contours of regularization term $\Omega(\theta)$
- 6406531894856. ✗ The dotted circles represent contours of loss term $L(\theta)$
- 6406531894857. ✓ The solid circles represent the contours of loss term $L(\theta)$
- 6406531894858. ✗ The solid circles represent the contours of regularization term $\Omega(\theta)$
- 6406531894859. ✓ Increasing the value of α make the parameters sparse
- 6406531894860. ✗ Decreasing the value of α makes the parameter sparse

6406531894861. ✓ The L_2 regularization is independent of the input samples used to train the network

Question Number : 118 Question Id : 640653567003 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

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 and Skip-Gram model. Choose which of the following are correct statements. The models were trained by building a vocabulary with unique words from the sentences.

Options :

6406531894872. ✖ CBOW model produces two different representations for the word bank based on its context

6406531894873. ✖ Skip-gram model produces two different representations for the word bank based on its context

6406531894874. ✓ CBOW model gives a single representation for the word bank despite its context

6406531894875. ✓ Skip-gram model gives a single representation for the word bank despite its context

Question Number : 119 Question Id : 640653567009 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 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 :

6406531894881. ✓ Fully connected Feed forward neural network

6406531894882. ✓ Convolutional Neural network

6406531894883. ✓ Reccurent Neural Network

6406531894884. ✓ Transformers

Sub-Section Number :	6
Sub-Section Id :	64065381165
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 120 Question Id : 640653566999 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 30×30 . The first convolution layer uses 100 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 10 kernels each of size 5×5 . How many parameters 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 :

26010

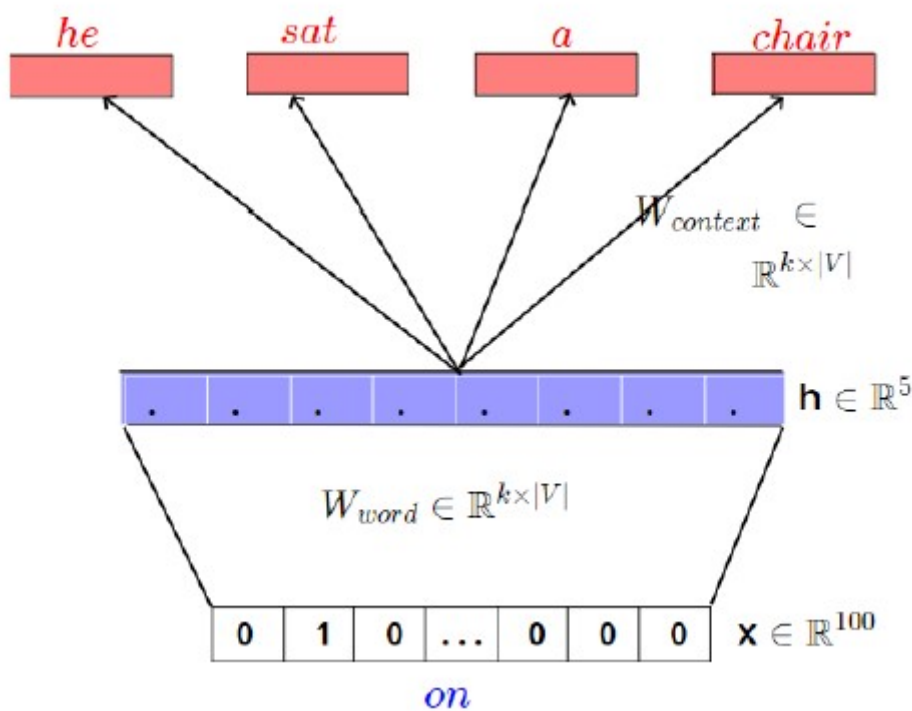
Question Number : 121 **Question Id :** 640653567004 **Question Type :** SA **Calculator :** None

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 :

Sub-Section Id :

64065381166

Question Shuffling Allowed :

Yes

Is Section Default? :

null

Question Number : 122

Question Id : 640653566996

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

A team has 25600 samples in a dataset for a classification problem. They decided to train the neural network model for 10 epochs with Nesterov Accelerated Gradient descent optimization algorithm with Cross entropy loss. Suppose they use a mini-batch with 32 samples per batch. How many times do the parameter in the network get updated?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

16000

Sub-Section Number :

8

Sub-Section Id :

64065381167

Question Shuffling Allowed :

Yes

Is Section Default? :

null

Question Number : 123

Question Id : 640653567008

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 the statement “the attention mechanism in RNN based Encoder- Decoder architecture helps the encoder to understand the context of words in a given sentence”. The statement is

Options :

6406531894879. ✖ TRUE

6406531894880. ✔ FALSE

Sub-Section Number :	9
Sub-Section Id :	64065381168
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653567000 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 : (124 to 125)

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 : 124 Question Id : 640653567001 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 : 125 Question Id : 640653567002 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 :

6406531894864. ✖ 32×32

6406531894865. ✖ 32×14

6406531894866. ✖ 14×32

6406531894867. ✔ 14×14

6406531894868. ✔ 14×13

6406531894869. ✖ 13×14

6406531894870. ✔ 14×7

6406531894871. ✖ 7×14

PSM

Section Id : 64065338433

Section Number : 6

Section type : Online