Deep Learning

Section Id :	64065357854
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
Section Negative Marks :	0
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 :	640653120923
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 50 Question Id : 640653821467 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 :

6406532756982. ✔ YES

6406532756983. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	640653120924
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 51 Question Id : 640653821468 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

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 :

6406532756984. ✔ The MP neuron model can represent a wide range of boolean functions (not all) by appropriately adjusting its weights and thresholds.

6406532756985. ✖ The MP neuron model can approximate arbitrary boolean functions, including non-linear ones.

6406532756986. ✖ The MP neuron model can accurately represent the XOR function by adjusting its weights and thresholds.

6406532756987. ✔ The representation power of the MP neuron model increases when multiple neurons are combined in a network architecture.

Sub-Section Number : 3

Sub-Section Id : 640653120925

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 52 Question Id : 640653821469 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 sigmoid neurons do we require to construct a tower function using single hidden layer to approximate a 2 dimensional continuous function ?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

4

Sub-Section Number : 4

Sub-Section Id : 640653120926

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 53 Question Id : 640653821470 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 a feedforward neural network with one hidden layer trained using backpropagation for a binary classification task. The network has the following architecture:

- Input layer with 15 neurons
- Hidden layer with 25 neurons
- Output layer with 1 neuron

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

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

If the loss function used for binary classification is the binary cross-entropy loss, and the activation fuction at hidden layer and output layer is sigmoid. The output of the neural network is denoted as $\hat{y}$, and the true label is denoted as y, what is the expression for $\frac{\partial L}{\partial w_j}$, where w_j represents the weights connecting the jth neuron of hidden layer to the output layer? Assume that the output of jth neuron of hidden layer is h_j and no biases in the network.

Options :

6406532756989. ✖ $\frac{\partial L}{\partial w_j} = -(y - \hat{y}) \cdot \sigma'(w_j h_j) \cdot h_j$

6406532756990. ✖ $\frac{\partial L}{\partial w_j} = -(y - \hat{y}) \cdot \sigma'\left(\sum_{i=1}^{25} w_i\right) \cdot h_j$

6406532756991. ✖ $\frac{\partial L}{\partial w_j} = -(y - \hat{y}) \cdot \sigma'(h_j) \cdot h_j$

6406532756992. ✔ $\frac{\partial L}{\partial w_j} = -(y - \hat{y}) \cdot \sigma'\left(\sum_{i=1}^{25} w_i h_i\right) \cdot h_j$

Question Number : 54 Question Id : 640653821471 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

In the context of mini-batch gradient descent, if reducing the size of the minibatch to half makes your model take thrice as many epochs to reach convergence, how does this affect the total number of parameter updates compared to using the original mini-batch size? Assume everything else remains constant.

Options :

6406532756993. ✖ The number of updates required increases by 2/3 times.

6406532756994. ✔ The number of updates required increases by six times.

6406532756995. ✖ The number of updates required increases by 3/2 times.

6406532756996. ✖ The number of updates required decreases by 2/3 times.

6406532756997. ✖ The number of updates required decreases by six times.

6406532756998. ✖ The number of updates required decreases by 3/2 times.

Question Number : 55 Question Id : 640653821472 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 about AdaMax optimisation algorithm are True?

Options :

6406532756999. ✖ It updates the learning rate using the average of past gradients.

6406532757000. ✖ It updates learning rates adaptively based on past gradients.

6406532757001. ✔ It uses the maximum of past gradients instead of the root mean square of past gradients.

6406532757002. ✖ It does not involve any adaptive learning rate mechanism.

Sub-Section Number : 5

Sub-Section Id : 640653120927

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 56 Question Id : 640653821473 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Choose all the correct options from below.

Options :

6406532757003. ✖ In Triangular Cyclic Learning rate scheme,learning rate changes abruptly after certain number of iteration.

6406532757004. ✔ In Cosine Annealing learning rate changes abruptly after certain number of iterations.

6406532757005. ✖ In Step Decay learning rate scheme we halve the learning rate after an epoch if the validation error is less than what it was for previous epoch.

6406532757006. ✔ In Warm Start method we start with high initial learning rate and then decay it.

Sub-Section Number : 6

Sub-Section Id : 640653120928

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 57 Question Id : 640653821474 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 an ensemble consisting of 5 models, where each model has an individual error rate of 50% on the test set. Assuming a simple majority voting scheme, what is the probability that a given instance is misclassified by the ensemble?

Assume that the models are independent and their errors are uncorrelated.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.45 to 0.55

Question Number : 58 Question Id : 640653821475 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

Given the true outputs for a dataset, y = [1,2,3,4], and the predictions from three different models (A, B, and C) trained on different subsets of the training data, as follows:

Model A predictions: A = [1.1,1.9,3.1,3.9]

Model B predictions: B = [0.9,2.1,2.9,4.1]

Model C predictions: C = [2,2,3,4]

Calculate the average variance for these models evaluated in a regression task. Assume that each model uses the same parameters but is trained on different subsets of the training data. (Enter your answer up to 3 decimal places.)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.05 to 0.07

Question Number : 59 Question Id : 640653821476 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 Convolutional Neural Network (CNN) architecture for image classification with the following layers:

1. Convolutional layer with 24 filters of size 3 × 3, with a stride of 1 and no padding.
2. Max pooling layer with a pool size of 2 × 2 and a stride of 2.
3. Convolutional layer with 56 filters of size 4 × 4, with a stride of 1 and no padding.
4. Max pooling layer with a pool size of 2 × 2 and a stride of 2.
5. Fully connected layer with 108 neurons.
6. Output layer with 10 neurons (for 10 classes) using softmax activation.

If the input image size is 84 × 84 × 3, and the network has no bias term, how many parameters are there in the CNN?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

41612

Question Number : 60 Question Id : 640653821477 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 you have a vocabulary of 5,000 unique words and you want to train a CBOW model with a window size of 3 (on each side) and with an embedding dimension of 500. How many parameters (weights) will the embedding layer have?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2500000

Question Number : 61 Question Id : 640653821478 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 your CBOW model outputs a probability distribution over a vocabulary of 10,000 words for a given context. If the correct target word is word number 200, and the model's predicted probability for this word is 0.3, calculate the cross-entropy loss for this prediction.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

1.1 to 1.3

Question Number : 62 Question Id : 640653821479 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

In a Skip-gram model with a window size of 2 (on each side), how many **unique** pairs of target and context words will be generated for the following sentence:

'Finish easy questions as early as possible'

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

19

Question Number : 63 Question Id : 640653821480 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

In a Skip-gram model with a vocabulary of 20000 words, if you choose a negative sampling rate of 5 negative samples for each positive sample, how many **total** samples will be generated for a sentence with 10 words and a window size of 2 (on each side)?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

204

Sub-Section Number :	7
Sub-Section Id :	640653120929
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653821481 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 : (64 to 66)

Question Label : Comprehension

Given a scenario where there are 8 possible letters represented as one-hot encoded vectors of length 8, the RNN employs the following formulas for the state vector and output at time step t:

$$s_t = \sigma(Ux_i + Ws_{t-1} + b)$$
$$\hat{y}_t = O(Vs_t + c)$$

Here, σ and O denote the sigmoid and softmax functions, respectively. Assume that $s_t \in \mathbb{R}^2$ and $y_t \in \mathbb{R}^8$

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 64 Question Id : 640653821482 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

With a total of 20 time steps (T = 20, implying prediction for a word of length 19), what is the total count of parameters (including bias) within the network?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

46

Question Number : 65 Question Id : 640653821483 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

If all the parameters (including bias) in the network are initialized to zero, what will be the predicted $\hat{y}_2$ at time step 2 (assume that indices start with 1) for the input $[0, 0, 0, 1, 0, 0, 0, 0]^T$?

Options :

6406532757015. ✖ $[0, 0, 0.5, 0.5, 0, 0, 0, 0]^T$

6406532757016. ✖ $[0.1, 0.1, 0.1, 0.3, 0.1, 0.1, 0.1, 0.1]^T$

6406532757017. ✖ $[0, 0, 0, 1, 0, 0, 0, 0]^T$

6406532757018. ✔ $[1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8]^T$

Question Number : 66 Question Id : 640653821484 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

If all the parameters (including bias) in the network are initialized to zero, what will be the total loss after 20 time steps (assume that indices start with 1) for the input $[0, 0, 0, 1, 0, 0, 0, 0]^T$? assume the loss to be cross-entropy at each time step. Write your answer correct to two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

41.5 to 41.66

Sub-Section Number :	8
Sub-Section Id :	640653120930

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 67 Question Id : 640653821485 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider training a Recurrent Neural Network (RNN) for a sequence prediction task. Which of the following statements accurately describe challenges in training RNNs?

Options :

6406532757020. ✔ RNNs suffer from the problem of ”short-term memory,” where they struggle to retain information from earlier time steps when processing long sequences, leading to limitations in capturing long-term dependencies.

6406532757021. ✖ RNNs do not suffer from the problem of ’short-term memory’ and can effectively retain information from earlier time steps even when processing long sequences.

6406532757022. ✔ Training RNNs requires careful initialization of weights and biases to ensure stable convergence during training, as poorly initialized weights can lead to gradient saturation and slow learning.

6406532757023. ✔ Architectural modifications such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been introduced to address the problem of ”short-term memory” in RNNs and improve their ability to capture long-term dependencies.

Sub-Section Number : 9

Sub-Section Id : 640653120931

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 68 Question Id : 640653821486 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

State True or False. GRUs are popular variant of LSTMs in which output and forget gates are combined.

Options :

6406532757024. ✖ TRUE

6406532757025. ✔ FALSE