X

![swayam logo] **(https://swayam.gov.in)** ![NPTEL logo] (https://swayam.gov.in/nc_details/NPTEL)

shweta_s@cs.iitr.ac.in ⌄

NPTEL (https://swayam.gov.in/explorer?ncCode=NPTEL) » Deep Learning - IIT Ropar (course)

≡

Click to register for Certification exam (https:// examform.nptel.ac.in/2025_10/ exam_form/ dashboard)

If already registered, click to check your payment status

Course outline

About NPTEL ()

How does an NPTEL online course work? ()

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 4 : Assignment 4

The due date for submitting this assignment has passed.
Due on 2025-02-19, 23:59 IST.

Assignment submitted on 2025-02-13, 16:38 IST

1) Using the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$, what would **1 point** be the bias-corrected first moment estimate after the first update if the initial gradient is 4?

○ 0.4
◉ 4.0
○ 3.6
○ 0.44

Yes, the answer is correct.
Score: 1
Accepted Answers:
4.0

2) In a mini-batch gradient descent algorithm, if the total number of training samples is **1 point** 50,000 and the batch size is 100, how many iterations are required to complete 10 epochs?

◉ 5,000
○ 50,000
○ 500
○ 5

Yes, the answer is correct.
Score: 1
Accepted Answers:
5,000

3) In a stochastic gradient descent algorithm, the learning rate starts at 0.1 and decays exponentially with a decay rate of 0.1 per epoch. What will be the learning rate after 5 epochs?

3.162

No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Range) 0.060,0.061

1 point

4) In the context of Adam optimizer, what is the purpose of bias correction?   **1 point**

◉ To prevent overfitting

◯ To speed up convergence

◯ To correct for the bias in the estimates of first and second moments

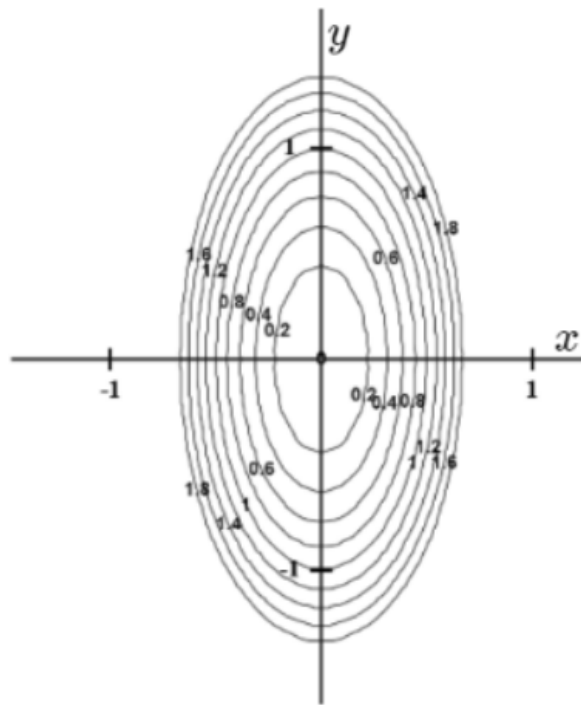◯ To adjust the learning rate

No, the answer is incorrect.
Score: 0
Accepted Answers:
To correct for the bias in the estimates of first and second moments

5) The figure below shows the contours of a surface.   **1 point**

Suppose that a man walks, from -1 to +1, on both the horizontal (x) axis and the vertical (y) axis. The statement that the man would have seen the slope change rapidly along the x-axis than the y-axis is,

◯ True

◉ False

◯ Cannot say

No, the answer is incorrect.
Score: 0
Accepted Answers:
True

6) What is the primary benefit of using Adagrad compared to other optimization algorithms? **1 point**

◉ It converges faster than other optimization algorithms.

○ It is more memory-efficient than other optimization algorithms.

○ It is less sensitive to the choice of hyperparameters(learning rate).

○ It is less likely to get stuck in local optima than other optimization algorithms.

No, the answer is incorrect.
Score: 0

Accepted Answers:
It is less sensitive to the choice of hyperparameters(learning rate).

7) What are the benefits of using stochastic gradient descent compared to vanilla gradient descent? **1 point**

☑ SGD converges more quickly than vanilla gradient descent.

☐ SGD is computationally efficient for large datasets.

☑ SGD theoretically guarantees that the descent direction is optimal.

☐ SGD experiences less oscillation compared to vanilla gradient descent.

No, the answer is incorrect.
Score: 0

Accepted Answers:
SGD converges more quickly than vanilla gradient descent.
SGD is computationally efficient for large datasets.

8) What is the role of activation functions in deep learning? **1 point**

◉ Activation functions transform the output of a neuron into a non-linear function, allowing the network to learn complex patterns.

○ Activation functions make the network faster by reducing the number of iterations needed for training.

○ Activation functions are used to normalize the input data.

○ Activation functions are used to compute the loss function.

Yes, the answer is correct.
Score: 1

Accepted Answers:
Activation functions transform the output of a neuron into a non-linear function, allowing the network to learn complex patterns.

9) What is the advantage of using mini-batch gradient descent over batch gradient descent? **1 point**

☐ Mini-batch gradient descent is more computationally efficient than batch gradient descent.

☐ Mini-batch gradient descent leads to a more accurate estimate of the gradient than batch gradient descent.

☑ Mini batch gradient descent gives us a better solution.

☑ Mini-batch gradient descent can converge faster than batch gradient descent.

No, the answer is incorrect.
Score: 0

Accepted Answers:

Jan 2025 ()

Mini-batch gradient descent is more computationally efficient than batch gradient descent.

Mini-batch gradient descent can converge faster than batch gradient descent.

10) In the Nesterov Accelerated Gradient (NAG) algorithm, the gradient is computed at: *1 point*

◯ The current position

◉ A "look-ahead" position

◯ The previous position

◯ The average of current and previous positions

Yes, the answer is correct.
Score: 1

Accepted Answers:
A "look-ahead" position