

X


<https://swayam.gov.in>

[https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL)

2111cs010024@mallareddyuniversity.ac.in ✓

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

## Course outline

About NPTEL ()

How does an NPTEL online course work? ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

- ☐ Recap: Learning Parameters: Guess Work, Gradient Descent (unit? unit=59&lesson=60)
- ☐ Contours Maps (unit? unit=59&lesson=61)

## Week 4 : Assignment 4

The due date for submitting this assignment has passed.

**Due on 2024-08-21, 23:59 IST.**

As per our records you have not submitted this assignment.

1) 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).*

2) Select the true statements about the factor  $\beta$  used in the momentum based gradient descent algorithm. **1 point**

- ☐ Setting  $\beta = 0.1$  allows the algorithm to move faster than the vanilla gradient descent algorithm
- ☐ Setting  $\beta = 0$  makes it equivalent to the vanilla gradient descent algorithm
- ☐ Setting  $\beta = 1$  makes it equivalent to the vanilla gradient descent algorithm
- ☐ Oscillation around the minimum will be less if we set  $\beta = 0.1$  than setting  $\beta = 0.99$

No, the answer is incorrect.



☐ Momentum based Gradient Descent (unit? unit=59&lesson=62)

☐ Nesterov Accelerated Gradient Descent (unit? unit=59&lesson=63)

☐ Stochastic And Mini-Batch Gradient Descent (unit? unit=59&lesson=64)

☐ Tips for Adjusting Learning Rate and Momentum (unit? unit=59&lesson=65)

☐ Line Search (unit? unit=59&lesson=66)

☐ Gradient Descent with Adaptive Learning Rate (unit? unit=59&lesson=67)

☐ Bias Correction in Adam (unit? unit=59&lesson=68)

☐ Lecture Material for Week 4 (unit? unit=59&lesson=69)

☐ Week 4 Feedback Form: Deep Learning - IIT

Score: 0

Accepted Answers:

*Setting  $\beta = 0.1$  allows the algorithm to move faster than the vanilla gradient descent algorithm*

*Setting  $\beta = 0$  makes it equivalent to the vanilla gradient descent algorithm*

*Oscillation around the minimum will be less if we set  $\beta = 0.1$  than setting  $\beta = 0.99$*

3) Select the behaviour of the Gradient descent algorithm that uses the following update rule,

$$w_{t+1} = w_t - \eta \nabla w_t$$

where  $w$  is a weight and  $\eta$  is a learning rate.

- ☐ The weight update is tiny at a steep loss surface
- ☐ The weight update is tiny at a gentle loss surface
- ☐ The weight update is large at a steep loss surface
- ☐ The weight update is large at a gentle loss surface

No, the answer is incorrect.

Score: 0

Accepted Answers:

*The weight update is tiny at a gentle loss surface*

*The weight update is large at a steep loss surface*

4) Which of the following algorithms will result in more oscillations of the parameter during the training process of the neural network?

- ☐ Stochastic gradient descent
- ☐ Mini batch gradient descent
- ☐ Batch gradient descent
- ☐ Batch NAG

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Stochastic gradient descent*

5) Which of the following are among the disadvantages of Adagrad?

- ☐ It doesn't work well for the Sparse matrix.
- ☐ It usually goes past the minima.
- ☐ It gets stuck before reaching the minima.
- ☐ Weight updates are very small at the initial stages of the algorithm.

No, the answer is incorrect.

Score: 0

Accepted Answers:

*It gets stuck before reaching the minima.*

6) Which of the following is a variant of gradient descent that uses an estimate of the next gradient to update the current position of the parameters?

- ☐ Momentum optimization
- ☐ Stochastic gradient descent

1 point

1 point

1 point

1 point



Ropar (unit?  
unit=59&less  
n=187)

Quiz: Week 4  
: Assignment  
4  
(assessment?  
name=288)

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

Week 11 ()

Week 12 ()

Download  
Videos ()

Books ()

Text  
Transcripts  
()

Problem  
Solving  
Session -  
July 2024 ()

- ☐ Nesterov accelerated gradient descent
- ☐ Adagrad

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Nesterov accelerated gradient descent*

7) Consider a gradient profile  $\nabla W = [1, 0.9, 0.6, 0.01, 0.1, 0.2, 0.5, 0.55, 0.56]$ . **1 point**  
Assume  $v_{-1} = 0$ ,  $\epsilon = 0$ ,  $\beta = 0.9$  and the learning rate is  $\eta_{-1} = 0.1$ . Suppose that we use the Adagrad algorithm then what is the value of  $\eta_6 = \eta / \sqrt{v_t + \epsilon}$ ?

- ☐ 0.03
- ☐ 0.06
- ☐ 0.08
- ☐ 0.006

No, the answer is incorrect.

Score: 0

Accepted Answers:

*0.06*

8) Which of the following can help avoid getting stuck in a poor local minimum while training a deep neural network? **1 point**

- ☐ Using a smaller learning rate.
- ☐ Using a smaller batch size.
- ☐ Using a shallow neural network instead.
- ☐ None of the above.

No, the answer is incorrect.

Score: 0

Accepted Answers:

*None of the above.*

9) What are the two main components of the ADAM optimizer? **1 point**

- ☐ Momentum and learning rate.
- ☐ Gradient magnitude and previous gradient.
- ☐ Exponential weighted moving average and gradient variance.
- ☐ Learning rate and a regularization term.

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Exponential weighted moving average and gradient variance.*

10) 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 required for training.
- ☐ Activation functions are used to normalize the input data.

☐ Activation functions are used to compute the loss function.

No, the answer is incorrect.

Score: 0

Accepted Answers:

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

