

X

<https://swayam.gov.in>https://swayam.gov.in/nc_details/NPTEL

2111cs010016@mallareddyuniversity.ac.in ✓

NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)Course
outlineAbout
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&lesso
n=60)
- ☐ Contours
Maps (unit?
unit=59&lesso
n=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) A team has a data set that contains 1000 samples for training a feed-forward neural **1 point** network. Suppose they decided to use stochastic gradient descent algorithm to update the weights. How many times do the weights get updated after training the network for 5 epochs?

- ☐ 1000
- ☐ 5000
- ☐ 100
- ☐ 5

No, the answer is incorrect.

Score: 0

Accepted Answers:

5000

2) 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).

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

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

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

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

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

☐ Line Search (unit? unit=59&lesso n=66)

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

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

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

☐ Week 4 Feedback Form: Deep Learning - IIT

- ☐ 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.

4) A team has a data set that contains 100 samples for training a feed-forward neural network. Suppose they decided to use the gradient descent algorithm to update the weights. Suppose further that they use line search algorithm for the learning rate as follows, $\eta = [0.01, 0.1, 1, 2, 10]$. How many times do the weights get updated after training the network for 10 epochs? (Note, for each weight update the loss has to decrease) **1 point**

- ☐ 100
- ☐ 5
- ☐ 500
- ☐ 10
- ☐ 50

No, the answer is incorrect.

Score: 0

Accepted Answers:

10

5) Select the behaviour of the Gradient descent algorithm that uses the following update rule, **1 point**

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

where w is a weight and η 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

6) In Nesterov accelerated gradient descent, what step is performed before determining the update size? **1 point**

- ☐ Increase the momentum
- ☐ Adjust the learning rate
- ☐ Decrease the step size
- ☐ Estimate the next position of the parameters

Ropar (unit?
unit=59&lesso
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 ()

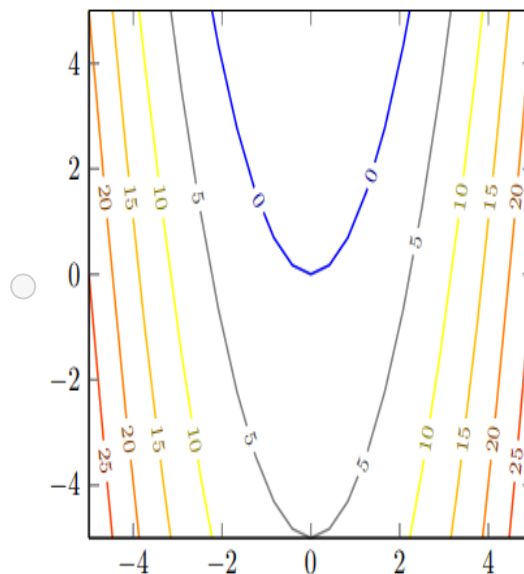
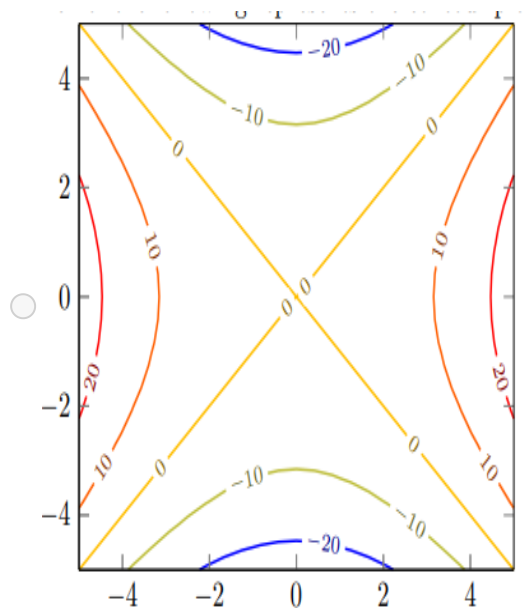
No, the answer is incorrect.

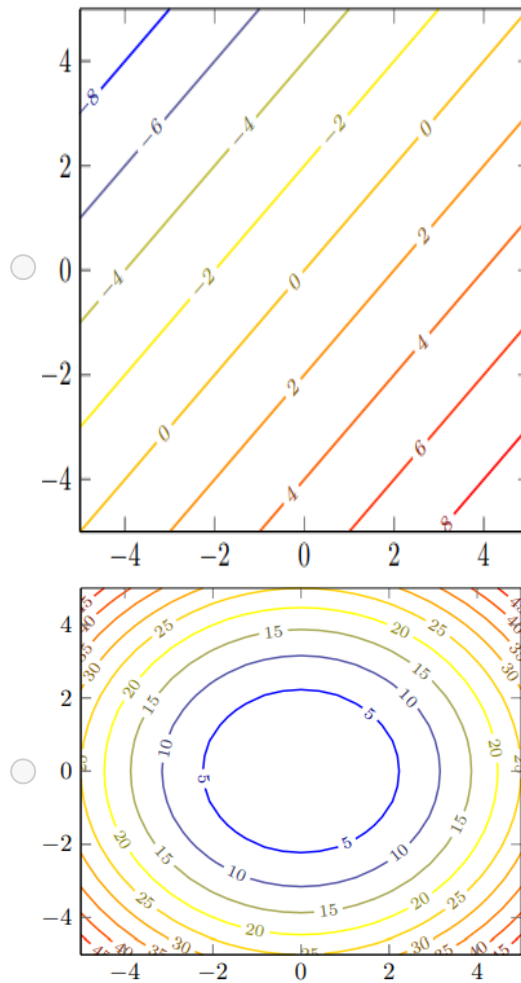
Score: 0

Accepted Answers:

Estimate the next position of the parameters

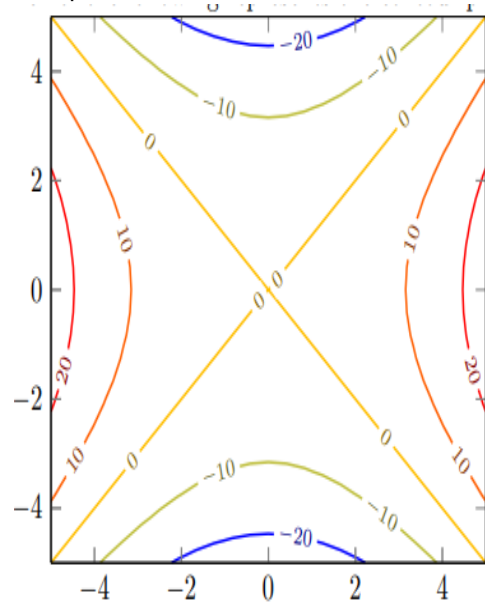
7) Which of the following represents the contour plot of the function $f(x,y) = x^2 - y^2$? **1 point**





No, the answer is incorrect.
Score: 0

Accepted Answers:



8) Which parameter in vanilla gradient descent determines the step size taken in the direction of the gradient? **1 point**

- ☐ Learning rate
- ☐ Momentum

- ☐ Gamma
- ☐ None of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

Learning rate

9) 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? **1 point**

- ☐ Momentum optimization
- ☐ Stochastic gradient descent
- ☐ Nesterov accelerated gradient descent
- ☐ Adagrad

No, the answer is incorrect.

Score: 0

Accepted Answers:

Nesterov accelerated gradient descent

10) 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.