

X


<https://swayam.gov.in>

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

2111cs010077@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&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.

Assignment submitted on 2024-08-20, 16:26 IST

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.

Yes, the answer is correct.

Score: 1

Accepted Answers:

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

2) 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

Yes, the answer is correct.

- ☐ 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

Score: 1

Accepted Answers:

10

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 η is a learning rate.

1 point

- ☐ 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

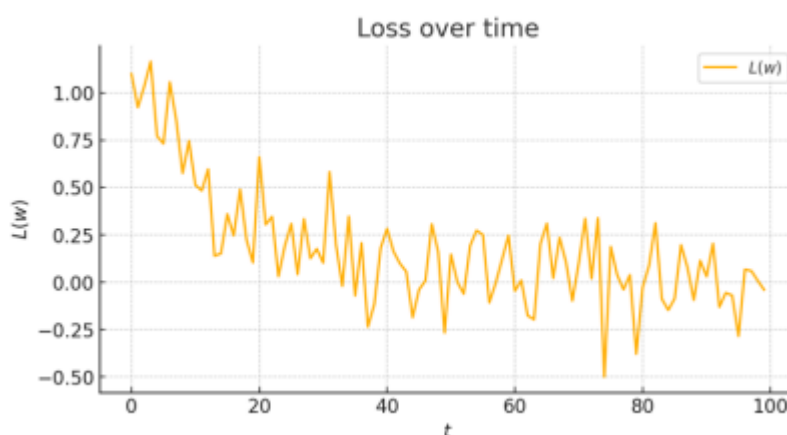
Yes, the answer is correct.

Score: 1

Accepted Answers:

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

4) The figure below shows the change in loss value over iterations

1 point

The oscillation in the loss value might be due to

- ☐ Mini-batch gradient descent algorithm used for parameter updates
- ☐ Batch gradient descent with constant learning rate algorithm used for parameter updates
- ☒ Stochastic gradient descent algorithm used for parameter updates
- ☐ Batch gradient descent with line search algorithm used for parameter updates

Partially Correct.

Score: 0.5

Accepted Answers:

*Mini-batch gradient descent algorithm used for parameter updates**Stochastic gradient descent algorithm used for parameter updates*

5) 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.

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 ()**

- ☐ Mini batch gradient descent gives us a better solution.
- ☒ Mini-batch gradient descent can converge faster than batch gradient descent.

Yes, the answer is correct.

Score: 1

Accepted Answers:

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

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

6) Given data where one column predominantly contains zero values, which algorithm **1 point** should be used to achieve faster convergence and optimize the loss function?

- ☒ Adam
- ☐ NAG
- ☐ Momentum-based gradient descent
- ☐ Stochastic gradient descent

Yes, the answer is correct.

Score: 1

Accepted Answers:

Adam

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

- ☒ Learning rate
- ☐ Momentum
- ☐ Gamma
- ☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

Learning rate

8) Which of the following algorithms will result in more oscillations of the parameter **1 point** during the training process of the neural network?

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

Yes, the answer is correct.

Score: 1

Accepted Answers:

Stochastic gradient descent

9) Which of the following is a variant of gradient descent that uses an estimate of the **1 point** next gradient to update the current position of the parameters?

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

☐ Adagrad

No, the answer is incorrect.

Score: 0

Accepted Answers:

Nesterov accelerated gradient descent

10) 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.

Yes, the answer is correct.

Score: 1

Accepted Answers:

None of the above.