

X

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

2111cs010033@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.

Assignment submitted on 2024-08-21, 21:07 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) Select the true statements about the factor β 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 $\beta = 0.99$

Partially Correct.

● 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: 0.67

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, **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

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

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

Yes, the answer is correct.

Score: 1

Accepted Answers:

Adam

5) 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

Yes, the answer is correct.

Score: 1

Accepted Answers:

Estimate the next position of the parameters

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

- ☒ Learning rate
- ☐ Momentum

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 ()**

- ☐ Gamma
- ☐ None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

Learning rate

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

1 point

- ☐ 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.

Yes, the answer is correct.

Score: 1

Accepted Answers:

It gets stuck before reaching the minima.

8) 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

Yes, the answer is correct.

Score: 1

Accepted Answers:

Nesterov accelerated gradient descent

9) Consider a gradient profile $\nabla W = [1, 0.9, 0.6, 0.01, 0.1, 0.2, 0.5, 0.55, 0.56]$. 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 $\eta_6 = \eta / \sqrt{v_t + \epsilon}$? **1 point**

- ☐ 0.03
- ☒ 0.06
- ☐ 0.08
- ☐ 0.006

Yes, the answer is correct.

Score: 1

Accepted Answers:

0.06

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.

Yes, the answer is correct.

Score: 1

Accepted Answers:

Exponential weighted moving average and gradient variance.