X

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

bhandarishivay2001@gmail.com

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

Week 4: Assignment 4

The due date for submitting this assignment has passed.

Due on 2022-08-24, 23:59 IST.

Assignment submitted on 2022-08-24, 21:59 IST

1) Consider the movement on the 3D error surface for Vannila Gradient Descent Algorithm. Select all the options that are TRUE. ***1 point***

- [x] Smaller the gradient, slower the movement
- [x] Larger the gradient, faster the movement
- [x] Gentle the slope, smaller the gradient
- [ ] Steeper the slope, smaller the gradient

Yes, the answer is correct.
Score: 1

Accepted Answers:

Smaller the gradient, slower the movement

Larger the gradient, faster the movement

Gentle the slope, smaller the gradient

2) Pick out the drawback in Vannila gradient descent algorithm. ***1 point***

- ( ) Very slow movement on gentle slopes
- (•) Increased oscillations before converging
- ( ) escapes minima because of long strides
- ( ) Very slow movement on steep slopes

No, the answer is incorrect.
Score: 0

Accepted Answers:

Week 11 ()

Week 12 ()

Download Videos ()

Books ()

Text Transcripts ()

Live Sessions ()

Problem Solving Session ()

Very slow movement on gentle slopes

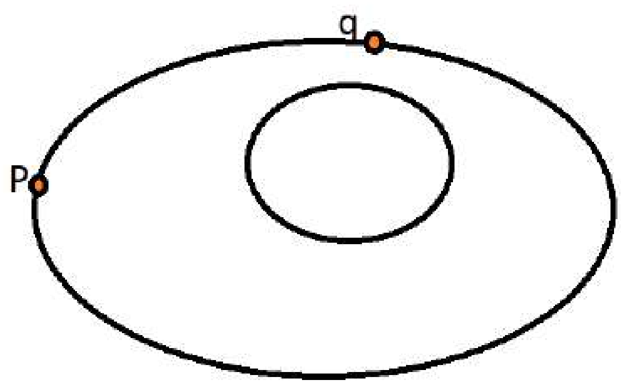
3) Comment on the update at the t^{th} update in the Momentum-based Gradient Descent. ***1 point***

- weighted average of gradient
- Polynomial weighted average
- Exponential weighted average of gradient
- Average of recent three gradients

Yes, the answer is correct.
Score: 1

Accepted Answers:
Exponential weighted average of gradient

4) Given a horizontal slice of the error surface as shown in the figure below, if the error at the position p is 0.49 then what is the error at point q? ***1 point***



- 0.70
- 0.69
- 0.49
- 0

Yes, the answer is correct.
Score: 1

Accepted Answers:
0.49

5) Identify the update rule for Nesterov Accelerated Gradient Descent. ***1 point***

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

- [ ]

$$
\begin{aligned}
update_t &= \gamma \cdot update_{t-1} + \eta \nabla w_t \\
w_{t+1} &= w_t - update_t
\end{aligned}
$$

- [x]

$$
\begin{aligned}
w_{look\_ahead} &= w_t - \gamma \cdot update_{t-1} \\
update_t &= \gamma \cdot update_{t-1} + \eta \nabla w_{look\_ahead} \\
w_{t+1} &= w_t - update_t
\end{aligned}
$$

- [ ]

$$
\begin{aligned}
v_t &= v_{t-1} + (\nabla w_t)^2 \\
w_{t+1} &= w_t - \frac{\eta}{\sqrt{v_t + \epsilon}} * \nabla w_t
\end{aligned}
$$

Yes, the answer is correct.
Score: 1

Accepted Answers:

$$
\begin{aligned}
w_{look\_ahead} &= w_t - \gamma \cdot update_{t-1} \\
update_t &= \gamma \cdot update_{t-1} + \eta \nabla w_{look\_ahead} \\
w_{t+1} &= w_t - update_t
\end{aligned}
$$

6) Select all the options that are TRUE for Line search. ***1 point***

- [x] w is updated using different learning rates
- [x] updated value of w always gives the minimum loss
- [ ] Involves minimum calculation
- [x] Best value of Learning rate is used at every step

Yes, the answer is correct.
Score: 1

Accepted Answers:
w is updated using different learning rates
updated value of w always gives the minimum loss
Best value of Learning rate is used at every step

7) Assume you have 1,50,000 data points, Mini batch size being 25,000, one epoch implies one pass over the data, and one step means one update of the parameters, What is the number of steps in one epoch for Mini-Batch Gradient Descent? ***1 point***

- [ ] 1

- ○ 1,50,000
- ◉ 6
- ○ 60

Yes, the answer is correct.
Score: 1

Accepted Answers:
6

8) Which of the following learning rate methods need to tune two hyperparameters? ***1 point***

I. step decay
II. exponential decay
III. 1/t decay

- ○ I and II
- ◉ II and III
- ○ I and III
- ○ I, II and III

Yes, the answer is correct.
Score: 1

Accepted Answers:
II and III

9) How can you reduce the oscillations and improve the stochastic estimates of the gradient that is estimated from one data point at a time? ***1 point***

- ◉ Mini-Batch
- ○ Adam
- ○ RMSprop
- ○ Adagrad

Yes, the answer is correct.
Score: 1

Accepted Answers:
Mini-Batch

10) Select all the statements that are TRUE. ***1 point***

- [ ] RMSprop is very aggressive when decaying the learning rate
- [x] Adagrad decays the learning rate in proportion to the update history
- [ ] In Adagrad, frequent parameters will receive very large updates because of the decayed learning rate
- [x] RMSprop has overcome the problem of Adagrad getting stuck when close to convergence

Yes, the answer is correct.
Score: 1

Accepted Answers:
Adagrad decays the learning rate in proportion to the update history
RMSprop has overcome the problem of Adagrad getting stuck when close to convergence