/ Quizzes

A?

CS-E4740 - Federated Learning D, Lectures, 28.2.2024-29.5.2024

Course feedback

Mark 3.00 out of 3.00

Correct

Quizzes

Started on Sunday, 17 March 2024, 5:27 PM **State** Finished

Completed on Sunday, 17 March 2024, 7:27 PM **Time taken** 2 hours

Consider the objective function

Grade 9.00 out of 11.00 (**81.82**%)

Question 1

 $f(\mathbf{w}) = \mathbf{w}^T \mathbf{Q} \mathbf{w} + \mathbf{q}^T \mathbf{w}$

with some positive semi-definite matrix ${f Q}$ and some vector ${f q}$. This function is differentiable with gradient $abla f({f w})$. Which of the following options are correct expressions for $abla f({f w})$?

 \square a. $\mathbf{Q}\mathbf{w}$

lacksquare b. $2\mathbf{Q}\mathbf{w}+\mathbf{q}$

 \square c. $\mathbf{Q}\mathbf{q}$

Your answer is correct.

See the matrix cookbook: https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

The correct answer is:

 $2\mathbf{Q}\mathbf{w} + \mathbf{q}$

Question 2

Flag question Mark 3.00 out of 3.00

Flag question

Correct

Consider gradient descent using a fixed **positive** learning rate η to learn the parameters of a linear model by minimizing the average loss on a training set. In other words, we use gradient descent to solve ERM for linear regression. Which statements are correct?

 \square a. Gradient descent will always converge, for any training set, to a solution of ERM as long as $\eta < 1$.

 \square b. For a given training set, there is a threshold γ such that gradient descent converges for all learning rates $\eta < \gamma$. This threshold depends only on the label values of the data points in the training set but not on their features.

C. Gradient descent can only converge if there is a unique solution to ERM, i.e., only if there is a unique choice for the model parameters resulting in minimum training error.

 \square d. For a given training set, there is a threshold γ such that gradient descent converges for all learning rates $\eta < \gamma$. This threshold depends only on the features of the data points in the training set but not on their label values.

Your answer is correct.

See the discussion in Section 4.4. of the lecture notes.

The correct answer is: For a given training set, there is a threshold γ such that gradient descent converges for all learning rates $\eta < \gamma$. This threshold depends only on the features of the data points in the training set but not on their label values.

This task requires you to try out a list of different values for the learning rates used in Algorithm 2 (Gradient Descent). For each value in the list, you have to determine the number of iterations required

Question 3

Flag question Mark 3.00 out of 3.00 Correct

Flag question

Mark 0.00 out of 2.00

Incorrect

This question refers to **the student task #1** in the "Gradient Methods" assignment.

Consider the model parameters obtained after 1000 GD steps using learning rate 0.1. In what range is the resulting training error E_t and the resulting validation error E_v ?

lacksquare a. $E_t \in [32,35]$, $E_v \in [38,43]$ 🗸

lacksquare b. $E_t \in [27,30]$, $E_v \in [34,38]$

 $lacksymbol{\circ}$ c. $E_t \in [37,41]$, $E_v \in [44,47]$

lacksquare d. $E_t \in [20,25]$, $E_v \in [50,53]$

Your answer is correct. Please, join Slack channel (link) if you have any questions regarding the assignment.

This question refers to **the student task #2** in the "Gradient Methods" assignment.

The correct answer is:

Question 4

 $E_t \in [32,35]$, $\ E_v \in [38,43]$

to ensure that the objective function (training error) changes less than a prescribed tolerance. Select the correct statements.

 \square a. For each learning rate in the given list, the number of iterations is larger than 2000.

The list contains a learning rate for which the required number of iterations is smaller than 1420. ✓

The list contains a learning rate for which the required number of iterations is smaller than 200. The optimal learning rate is **between 0.24 and 0.26**.

The optimal learning rate is **between 0.29 and 0.30**.

The optimal learning rate is **between 0.31 and 0.32**.

Your answer is incorrect.

Please, join Slack channel (link) if you have any questions regarding the assignment.

The correct answers are:

The optimal learning rate is between 0.31 and 0.32., The list contains a learning rate for which the required number of iterations is smaller than 1420.

Finish review

Previous activity ■ "FL Design Principle"

Next activity "FL Algorithms" ►

Aalto-yliopisto Aalto-universitetet

Aalto University

Tuki / Support Opiskelijoille / Students

MyCourses instructions for

students • email: mycourses(at)aalto.fi

Opettajille / Teachers

MyCourses help

MyTeaching Support form

Palvelusta

 MyCourses rekisteriseloste Tietosuojailmoitus

Palvelukuvaus

Saavutettavuusseloste

About service

MyCourses protection of

privacy Privacy notice

 Service description Accessibility summary

Service

MyCourses registerbeskrivining

 Dataskyddsmeddelande Beskrivining av tjänsten

• Sammanfattning av tillgängligheten

Nguyen Binh (Log out)

