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

Started on	Sunday, 17 March 2024, 8:02 PM
State	Finished
Completed on	Sunday, 17 March 2024, 10:02 PM
Time taken	2 hours
Grade	**9.00** out of 11.00 (**81.82**%)

Question 1

Flag question Mark 0.00 out of 2.00 Incorrect

GTVMin-based methods learn local model parameters by minimizing the sum of local loss and the scaled total variation of model parameters. What is the effect of using a large value for this scaling factor α (see Equation 5.9 from the lecture notes (link))?

- a. Using a large value for α results in local model parameters with small values for the local loss (training error) at the expense of relatively large variations of model parameters across edges of the empirical graph.
- b. Using a large value for α puts more emphasis on learning local model parameters that do not vary too much across edges of the empirical graph. ✔
- c. Using a large value for α results in local model parameters with small total variation at the expense of higher values for the local loss (training error). ✔
- d. Using a large value for α puts more emphasis on learning local model parameters with small local loss.
- e. The convergence is slower with larger α-value. ✖

Your answer is incorrect.

Choosing a large value for α puts more emphasis on enforcing similar local model parameters at well-connected nodes of the empirical graph. Using a smaller α puts more emphasis on having accurate predictions on the local dataset.

The correct answers are: Using a large value for α puts more emphasis on learning local model parameters that do not vary too much across edges of the empirical graph.,
Using a large value for α results in local model parameters with small total variation at the expense of higher values for the local loss (training error).

Question 2

Flag question Mark 2.00 out of 2.00 Correct

What is the motivation for using stochastic gradient descent (SGD) instead of gradient descent for solving GTVMin?

- a. SGD is more accurate for smaller datasets.
- b. SGD requires less computation (time) than GD methods. ✔
- c. For the same number of iterations, SGD always achieves better performance than GD methods.

Your answer is correct.

Computing the gradient of local loss functions typically requires to sum over all data points in the local datasets. However, local datasets might consist of a large number of data points which cannot be accessed quickly enough (e.g., because they are stored in the "cloud"). To speed up the computation of gradients, SGD approximates the sum over all data points by a sum of small set of randomly selected data points.

The correct answer is: SGD requires less computation (time) than GD methods.

Question 3

Flag question Mark 3.00 out of 3.00 Correct

*This question refers to **the student task #1** in the "FL Algorithms" assignment.*

What are the average training and validation errors obtained by FedGD algorithm with the hyperparameters specified in the task?

- a. Both average training and validation errors are **in the [20, 22] interval**. ✔
- b. Both average training and validation errors are **in the [24, 26] interval**.
- c. Both average training and validation errors are **in the [22, 24] interval**.
- d. The average training error is **in the [19, 25] interval** and the average validation error is **in the [22, 26] interval**.
- e. The average training error is **in the [24, 26] interval** and the average validation error is **in the [22, 26] interval**.

Your answer is correct.

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

The correct answer is:
Both average training and validation errors are **in the [20, 22] interval**.

Question 4

Flag question Mark 4.00 out of 4.00 Correct

*This question refers to **the student task #2** in the "FL Algorithms" assignment.*

What are the average training and validation errors obtained by FedSGD algorithm with the hyperparameters specified in the task?

- a. The average training error is **in the [19, 25] interval** and the average validation error is **in the [25, 28] interval**.
- b. Both average training and validation errors are **in the [24, 26] interval**.
- c. Both average training and validation errors are **in the [22, 24] interval**. ✔
- d. The average training error is **in the [24, 26] interval** and the average validation error is **in the [26, 28] interval**.
- e. Both average training and validation errors are **in the [20, 22] interval**.

Your answer is correct.

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

The correct answer is:
Both average training and validation errors are **in the [22, 24] interval**.

Finish review

Previous activity
◄ "Gradient Methods"

Next activity
"FL Flavors" ►