ELG 5142 Ubiquitous Sensing for Smart Cities

Assignment 5

Announcement Date: 2022-07-28

Submission Deadline: 2022-08-10

Teaching Assistant: Arda Onsu <monsu022@uottawa.ca>

In this assignment you are given a federated learning environment with *20 clients* and *FashionMnist* dataset.

In code each client trains its local model with its own data and send modal parameters to the server. After server receives all data, it starts federated aggregation and update global model to new global model and distribute it to the clients. This is called a communication round and it continues until the round count.

However, there are malicious clients that is used to interfere with the server aggregations. These malicious clients shuffle their label and updates their weights wrongly. There are two kinds of malicious clients

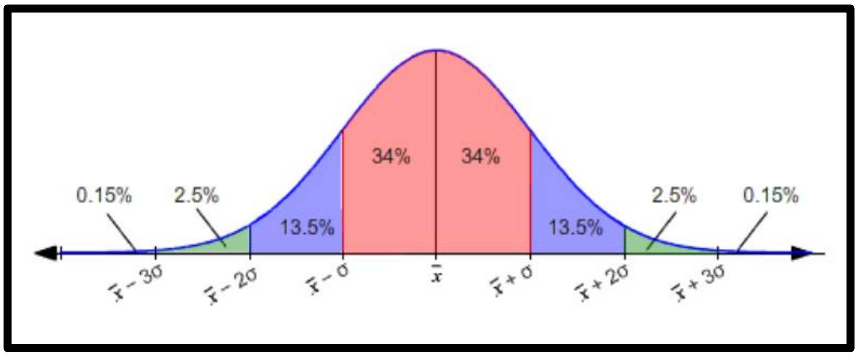
- **Partially Malicious Client**: it shuffles *half of its labels*. *15%* of clients are partially malicious. It is coded by number *1*.
- **Fully Malicious Client**: it shuffles *all of its labels*. *5%* of clients are Fully Malicious. It is coded by number *2*.

All clients send their parameters to the server. You are supposed to:

1. Detect if clients are partially malicious, fully malicious or benign. Then print the results.
2. Update federated aggregation. If client is benign, it can participate in the aggregation. It the client is partially malicious, it can participate in the aggregation with a probability of 50%. However, if client is fully malicious, it cannot participate the aggregation

For Task 1 you are going to use *loss* values and *gaussian distributions* over data.

- If the loss value of a client is between mean+standard_deviation and mean-standard_deviation [i.e., pink region] this client is considered as *benign (coded by 0)*. *(Case 1)*
- If case 1 does not hold for the client, and loss value of the client is between mean+2*standard_deviation and mean-2*standard_deviation [i.e., purple region], this client is considered as *partially malicious (coded by 1)*. *(Case 2)*
- Otherwise, it is considered as a fully malicious client [i.e., green region] *(coded by 2)*. *(Case 3)*



Deliverables and grading scheme:

1. Correctly find if clients are partially malicious, fully malicious, or *benign* in aggregation parts of server. (40 pts)
2. Print malicious situation of clients before aggregation *(1: partially malicious, 2: fully malicious, 0 benign)*. (10 pts)
3. Perform the federated averaging where fully malicious clients do not participate and partially clients 50% participate. (50 pts)

Tips

- ➢ You can combine Task 1 and Task 2 in *federated averaging part*. You need to change *federated_averaging(server, clients)* function and write your own federated averaging function.
- ➢ In the aggregation part, you can multiply model parameter with some *coefficient*. If the model participates in the aggregation, the coefficient can be 1, otherwise, it will be 0. You can put all coefficients into a list.
- ➢ You can use list of lost values for each client (*client_loss_split*) as a parameter of federated averaging function.