Part 2 Differential Privacy (25 points)

In this part, you will implement differential privacy for the federated learning task in part 1. To provide differential privacy of the local data for each client, random noises are injected to the dataset.

The Laplace mechanism perturbs the input data by injecting random noises drawn from the Laplace distribution. Specifically, for a 28×28 image X in the Federated EMNIST dataset, a Laplace noise of scale b is added to each pixel of the image to get the perturbed image X', i.e.,

$$X' = X + \varepsilon,$$

where ε is a 28×28 matrix with each entry ε_{ij} sampled from a 0-centered Laplace distribution with scale b. The Laplace distribution centered at 0 with scale b has the following probability density function.

$$\texttt{Lap}(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$$

The perturbed data X' are used for the local training at each client. Explore the effect of the injected noise scale b on the quality of the model trained by `FedAvg` in the following steps. Submit your answers along with any supporting plots in a .pdf file on Canvas. Submit your code on GitHub.

1. Use the perturbed dataset for the local training at each client. Use the `FedAvg` algorithm to train a 2-layer Neural Network with 128 hidden units and the RELU activation function for the image classification task in part 1 question 1. For each communication round, $C = 10\%$ of clients should be used. Plot the aggregated training and validation accuracy versus the number of communication rounds. Evaluate the trained model on the held-out test data and report the test accuracy.

2. Experiment with different noise scales and compare the training quality. Create a plot of the final training accuracy and test accuracy when different levels of perturbation are used. The y-axis should be the accuracy while the x-axis should be the noise scale b.

3. Based on your plot in 2, describe the relationship between the noise scale and the training and test accuracy. Which scale of the noise would you use to protect the data privacy while preserving the quality of the trained model? Briefly justify your choice.