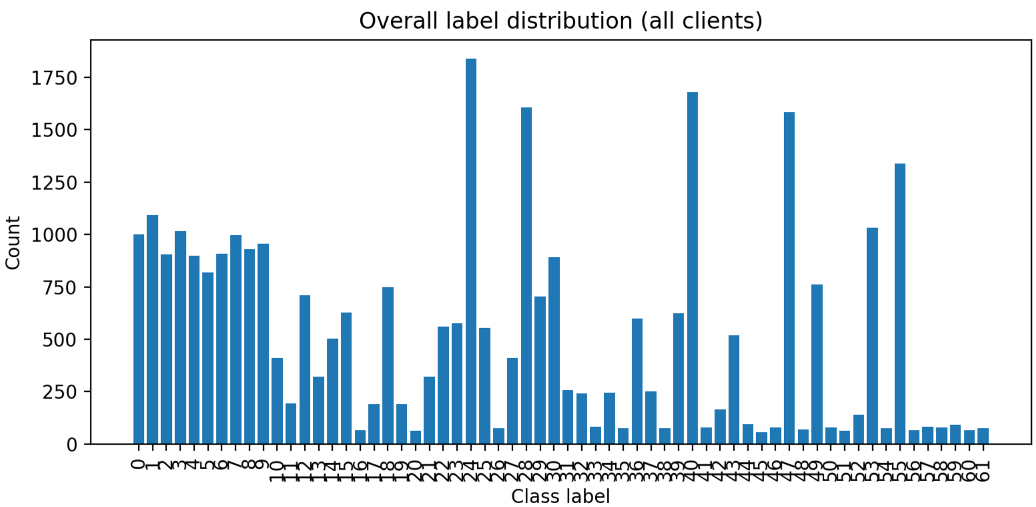
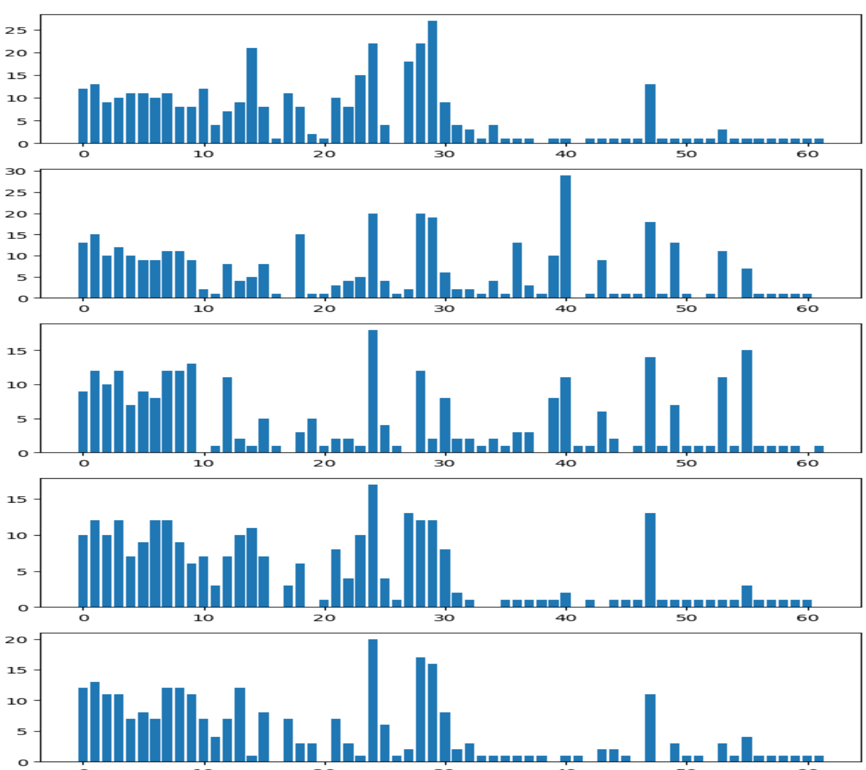
HW3 Report

hmsun0813
https://github.com/hmsun0813/IEMS469/hw3

Part 0 — Data Distribution (The 5 individual distribution generally mimic the overall distribution but with some variance, especially at the right tail.)



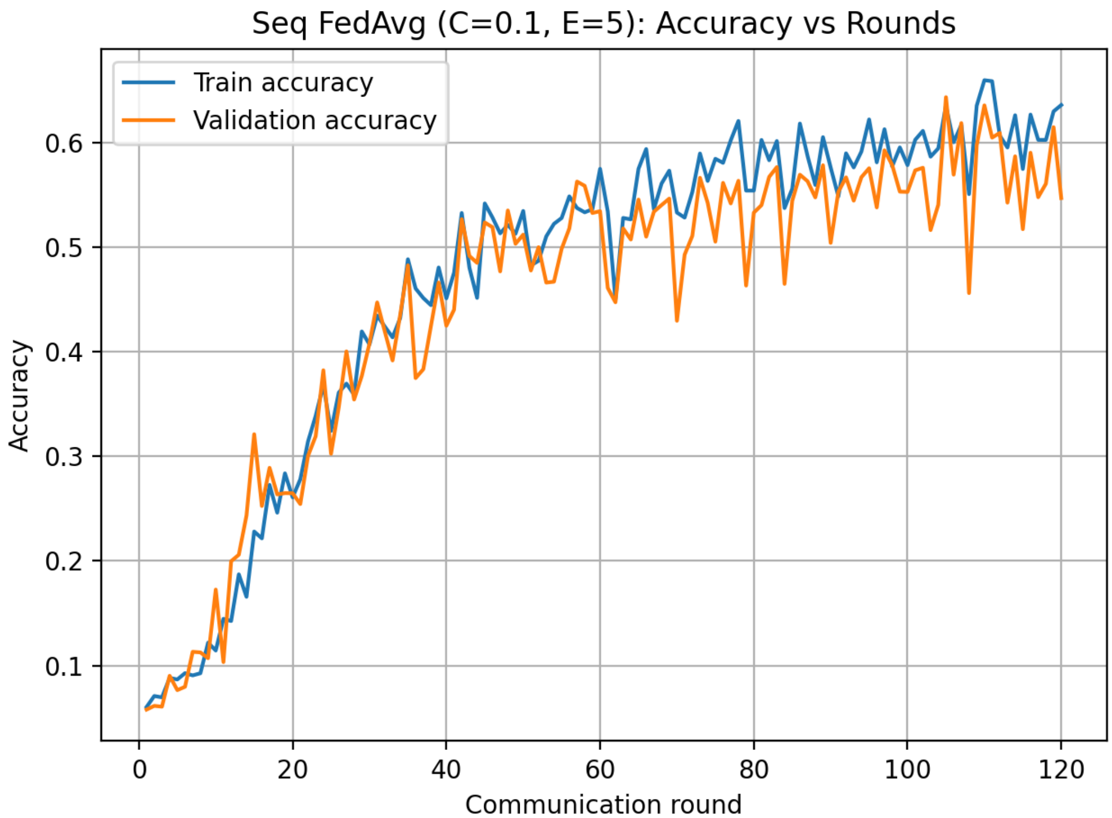


Part 1 — Sequential

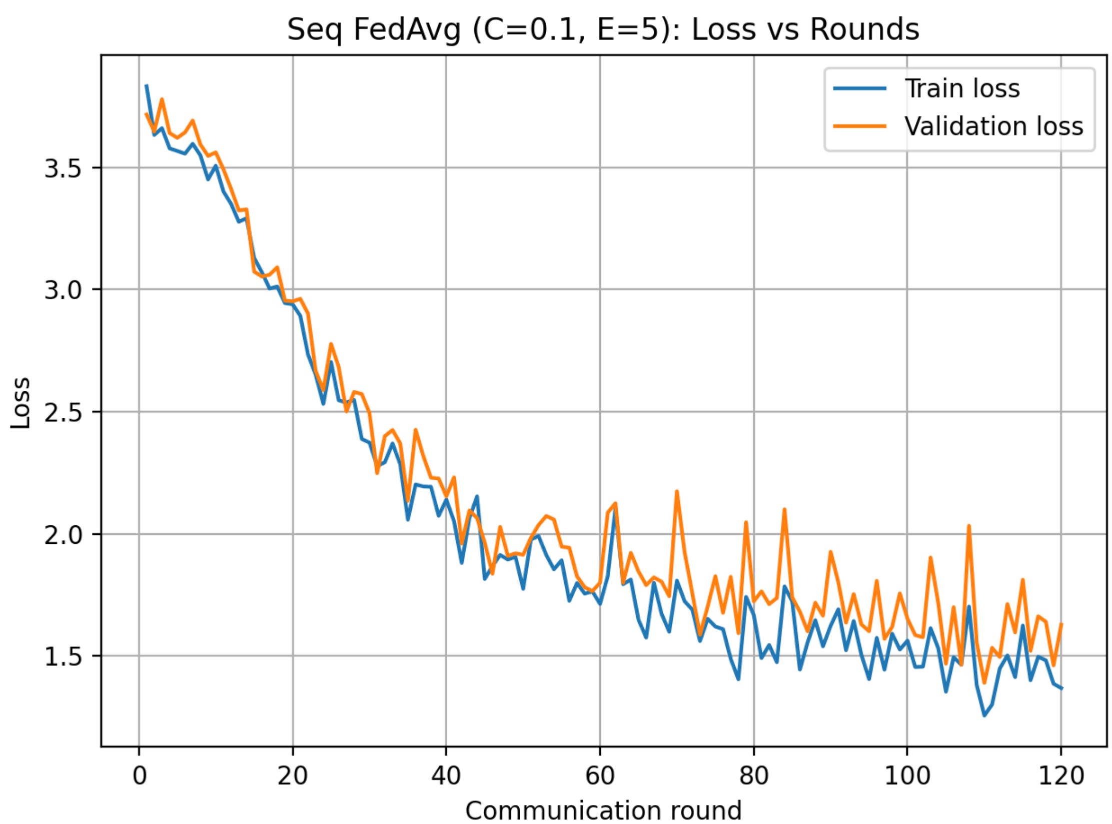
Comparing different C and E (plots for each setting can be found in GitHub repo), it indicates that a larger number of client per round and a larger number of training epoch per client local update effectively increases the accuracy:

C	E	training_acc	validation_acc	test_acc
0.02	1	0.1348	0.1527	0.1839
0.02	2	0.2960	0.2970	0.2853
0.02	3	0.2946	0.2910	0.2764
0.02	5	0.6114	0.7391	0.4985
0.05	1	0.1550	0.1226	0.1149
0.05	2	0.4205	0.4273	0.4140
0.05	3	0.4723	0.4342	0.4872
0.05	5	0.5909	0.5934	0.5449
0.1	1	0.1521	0.1542	0.1627
0.1	2	0.4196	0.3314	0.4198
0.1	3	0.4977	0.4992	0.4954
0.1	5	0.6359	0.5469	0.5846

The best performance (test_acc=0.5846) is achieved with C=10%, E=5, learning_rate=0.01, batch_size=64, rounds=120. Below is the plots for the aggregated training and validation loss and accuracy versus the number of communication rounds.

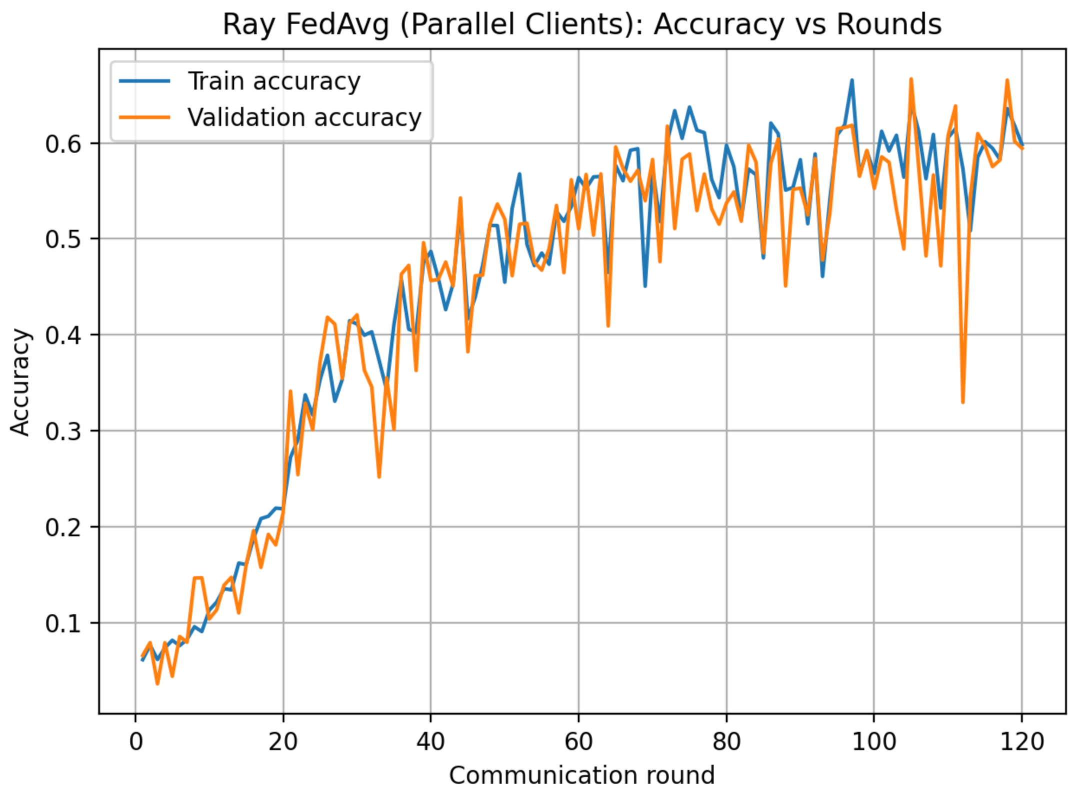


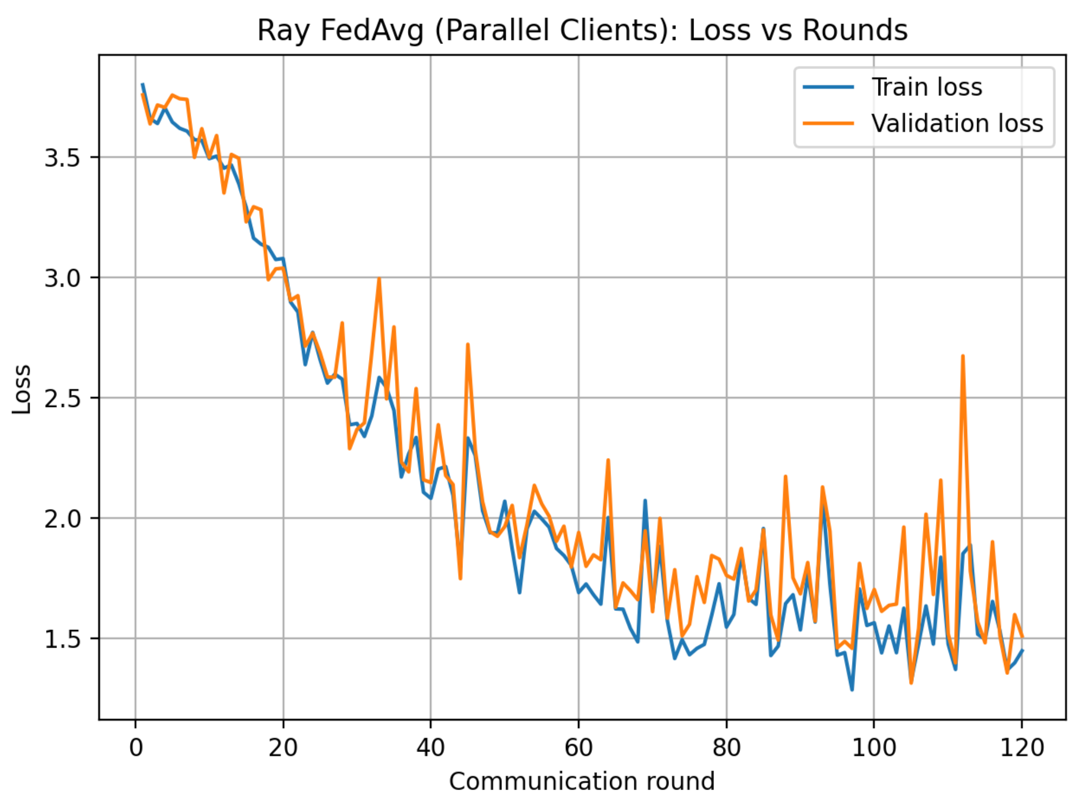
Seq FedAvg (C=0.1, E=5): Accuracy vs Rounds
Train accuracy
Validation accuracy
Accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
120
Communication round



Seq FedAvg (C=0.1, E=5): Loss vs Rounds
Train loss
Validation loss
Loss
1.5
2.0
2.5
3.0
3.5
0
20
40
60
80
100
120
Communication round

Part 1 — Parallel Client (test accuracy 0.5447; test loss 1.7998)





Part 2 — Differential Privacy

The results interestingly show a non-monotonic relationship between Laplace noise scale and model accuracy. Very small noise levels (b = 0.01 and b = 0.05) degrade both training and validation accuracy more sharply than expected, likely because light perturbations introduce instability in each client's local updates without providing meaningful regularization. Interestingly, the larger noise scale (b = 0.1) performs better than the smaller noise levels and nearly matches the no-noise baseline. This happens because high-magnitude Laplace noise is partly averaged out across clients in FedAvg, reducing its effective impact, while also providing a mild regularization effect that improves generalization.

Considering privacy and utility together, b = 0.1 provides the best trade-off: it injects the strongest privacy-preserving perturbation while still maintaining test accuracy close to the no-noise model. Thus, b = 0.1 is the preferred choice for protecting client data while preserving model quality.