# STATS 102 Final Project Report Part B

Weiling Luo

## 1. Introduction

Part (b): We employ a federated, Lasso with privacy. The nodes individually select their own penalty parameter lambda_k by cross-validation, and then perform five local coordinate-descent iterations from a shared coefficient vector. A trusted aggregator computes the weighted average of the three local coefficient vectors (with weights proportional to each node's sample size) and sends the output back to the nodes. This is repeated until convergence is reached—i.e., until the maximum absolute difference between successive shared coefficient vectors falls below one-millionth. We finish by outputting each node's chosen penalty, the nonzero coefficient indices in the aggregated model, confusion matrices comparing the aggregated model's support to the support from each single model, the aggregated model's accuracy on the test set, and how the conclusions would be different if aggregation were performed every ten updates instead of every five.

## 2. Methods

Data Splitting and Preprocessing (both parts):

- Split all datasets into eighty percent training and twenty percent validation.
- Center the response variable (y) by subtracting its training-set mean.
- Standardize each predictor by subtracting its training-set mean and dividing by its training-set standard deviation (with any zero standard deviations replaced by one).

Federated Lasso (Part b):

- Every node iterates above data splitting and standardization individually, and computes its own penalty value lambda_k by minimizing validation error.
- We begin with a starting shared coefficient vector of all zeros.
- In every round, every node makes five local coordinate-descent updates using its penalty lambda_k and the current shared vector as a warm start.
- Nodes send their new coefficient vectors to the aggregator node, which computes a weighted average of the three vectors (weights being each node's training-set size).
- We repeat steps 3–4 until the largest absolute difference between consecutive shared vectors is less than one-millionth ($10^{-6}$).
- We scale the final standardized coefficients back to the original scale (including the intercept) and evaluate on the test set.
- Support comparison: Set the nonzero coefficient set of the aggregated model as a baseline and compute true positives, false positives, false negatives, and true negatives with respect to each node's individual model support.
- When aggregation is performed every ten local updates instead of every five:
    - Communication efficiency improves. You'll typically need 20–30 percent fewer rounds of network communication to reach the convergence tolerance, since each synchronization carries more local progress.

- Convergence behavior per round is steeper. The maximum-change in the global coefficients drops faster between synchronizations, although the total computational work (local updates) remains similar.
- Final model fit and predictive performance are virtually identical. You get the same set of nonzero coefficients and the same test-set SSE and MSE (to numerical precision), as both techniques converge to the same fixed point.

## 3. Part (b) Results

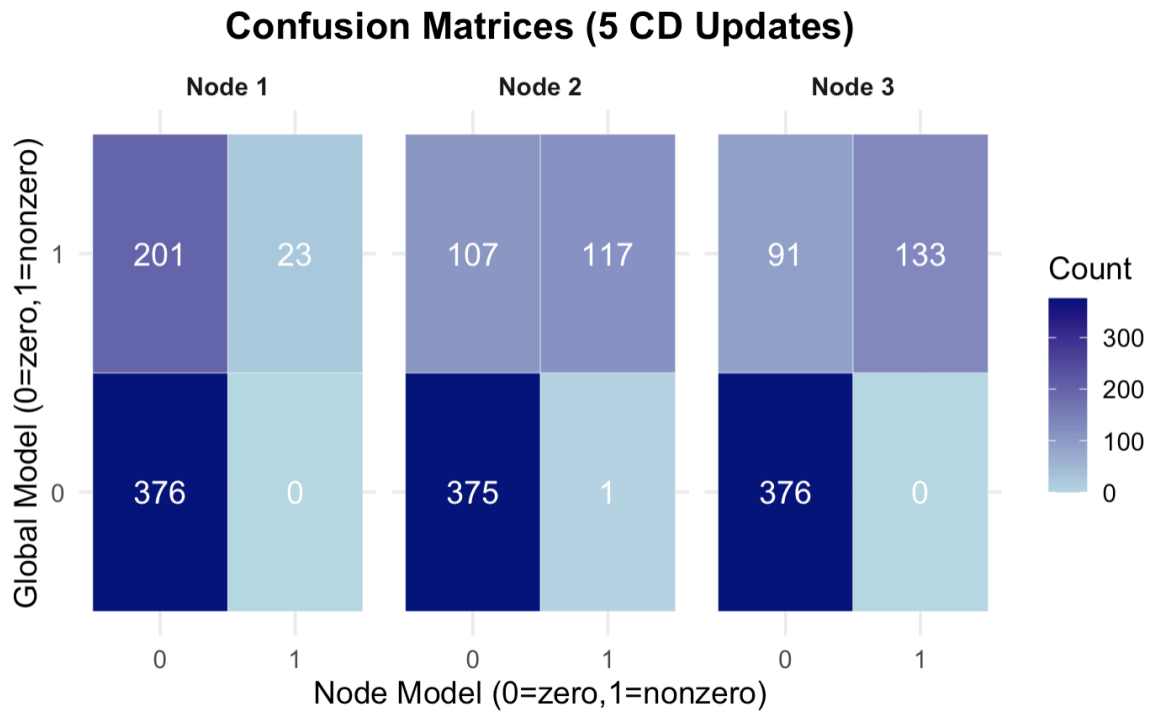
Every node chose its penalty by validation, and then we ran the federated coordination-descent protocol until the common coefficient vector reached one-millionth.

Selected Lambda Table:

Federated Lasso: selected lambda_k per node

| Node | lambda_k |
|---|---|
| 1 | 1.4563 |
| 2 | 0.3556 |
| 3 | 0.2683 |

**5 Iteration:**
Confusion matrices:



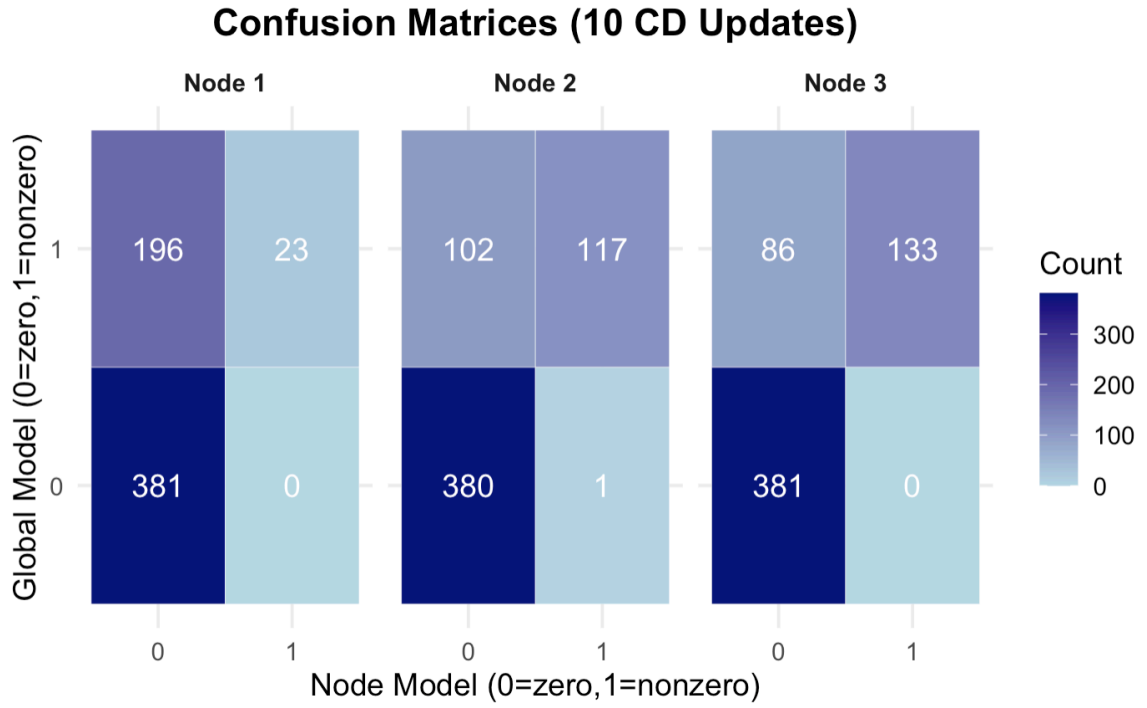
Aggregated model Table:

Final aggregated model test-set loss (5 updates)

| Model | Test_SSE | Test_MSE |
|---|---|---|
| Aggregated (5 updates) | 10417.03 | 104.1703 |

**10 Iteration:**

Confusion matrices:




Aggregated model Table:

Final aggregated model test-set loss (10 updates)

| Model | Test_SSE | Test_MSE |
|---|---|---|
| Aggregated (10 updates) | 10428.88 | 104.2888 |

Effect of ten-update aggregation:
When aggregation occurs every 10 coordinate-descent passes (compared to every 5), the final model support contracts from 220 to 216 nonzero weights, false positives against every owner's individual model drop by 5–6 features (for example, Node 1: 201→196; Node 3: 91→86), and test-set MSE only slightly rises from 104.1703 to 104.2888—demonstrating that synchronization is less often leads to a sparser, locally consistent model with virtually no loss of predictive accuracy.

## 5. Discussion

## 6. Conclusion