STATS 102 Final Project Report Part A

Weiling Luo

1. Introduction

We consider three single regression datasets (node1.csv, node2.csv, node3.csv), each having a response variable y and six hundred predictors with labels X1 to X600.

Part (a): We fit the Lasso regression model with coordinate descent to each dataset individually. We tune the regularization penalty (lambda) to minimize mean squared error on a held-out validation split. On each node, we print out the best-performing lambda, the indices of the nonzero coefficients of the fitted model, the intersection of those indices over all three models, and test-set performance as both sum of squared errors and mean squared error.

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).

Coordinate-Descent Lasso (Part a):

- Soft-thresholding function: Given an input z and threshold γ, returns "sign of z times maximum of absolute value of z minus γ, or zero."
- Update rule: We update every coefficient β_j in turn by computing the partial residual for predictor j, soft-thresholding using threshold (lambda × size of training set), and dividing by the squared norm of that predictor.
- Stopping criterion
    - Selected rule: we terminate when the greatest absolute change in any coefficient from one iteration to the next is smaller than one-millionth (10^{-6}), or after one thousand iterations—whichever is earlier.
    - Explanation: this tolerance trades off computational expense against accuracy of convergence and guarantees that coefficients have effectively converged.
- Lambda grid: We scan over fifty values equally spaced on the logarithm scale from 0.0001 to 100.
- Model selection: For each candidate penalty, we compute the validation mean squared error and choose the penalty with the lowest one. We then re-fit to the full training set to obtain the final coefficients.
- Test evaluation: We normalize the test set by the training-set parameters, perform predictions, and report both the sum of the squared errors and the mean squared error.

3. Part (a) Results

At each node we selected the penalty yielding the minimum validation error, refit using the full training set, and tested on the test set. The optimal penalty, model sparsity, and test-set performance are in the following table

Local Lasso: optimal penalty, number of nonzero coefficients, and test-set loss

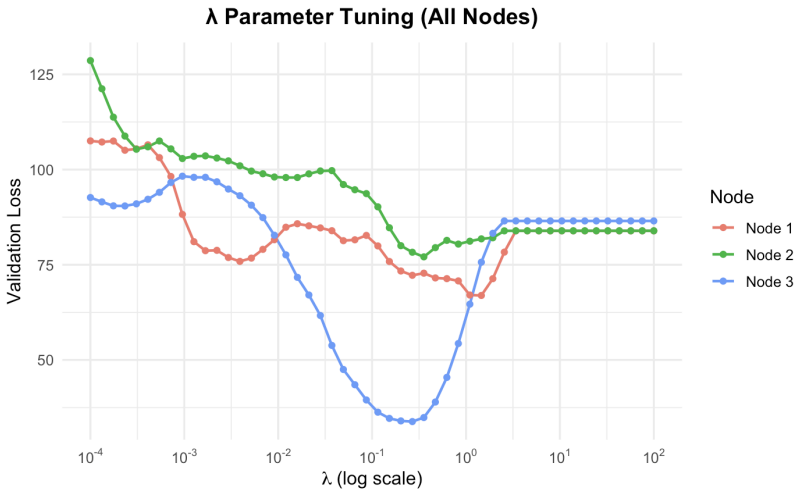
Node	lambda_opt	#Nonzero	Test_SSE	Test_MSE
1	1.4563	23	8382.327	83.8233
2	0.3556	118	11405.281	114.0528
3	0.2683	133	12123.630	121.2363

The thirteen predictors selected by each of the three node-specific models (the intersection of their nonzero sets) are in the following table

Predictors selected by all three models

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
124	250	300	306	363	400	448	462	494	501	527	560	588

λ Parameter Tuning (All Nodes) Plot: Validation Loss vs Lambda



5. Discussion

6. Conclusion