MA477: Data Science

Lesson 9 Board Sheet — 29 January 2026

United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

1 Shrinkage Lesson Objectives

- Understand the mathematical formulation of shrinkage methods and be able to compare/contrast them.
- Employ shrinkage methods in regression settings

Discussion Questions

We will organize this lecture around the following five questions:

1. Why might shrinking regression coefficients improve prediction, even if it makes the model less flexible?
2. What do the regression penalties do to the coefficients as the tuning parameter λ increases?
3. Why must predictors be standardized before applying shrinkage methods?
4. How do shrinkage methods' penalties differ, and why does that difference lead to variable selection?
5. Any thoughts on "smart" ways to select λ? Why might these fail?
6. How is the tuning parameter λ chosen in practice, and what is being optimized?

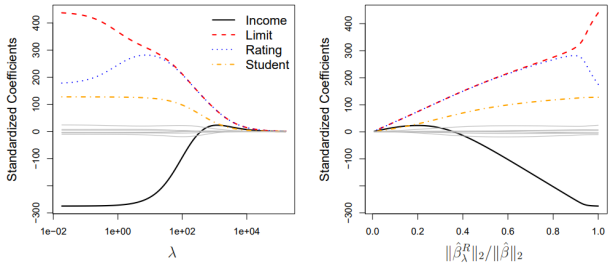

FIGURE 6.4. *The standardized ridge regression coefficients are displayed for the* `Credit` *data set, as a function of* λ *and* $\|\hat{\beta}^R_\lambda\|_2/\|\hat{\beta}\|_2$.

Figure 1

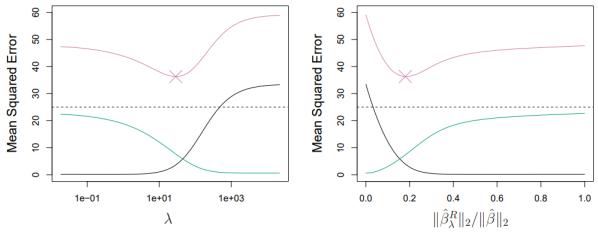

FIGURE 6.5. *Squared bias (black), variance (green), and test mean squared error (purple) for the ridge regression predictions on a simulated data set, as a function of* λ *and* $\|\hat{\beta}^R_\lambda\|_2/\|\hat{\beta}\|_2$. *The horizontal dashed lines indicate the minimum possible MSE. The purple crosses indicate the ridge regression models for which the MSE is smallest.*

Figure 2

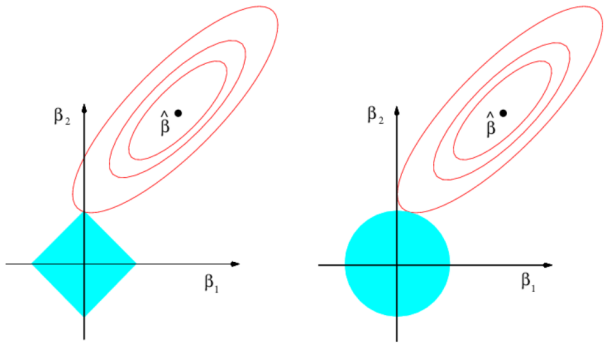

FIGURE 6.7. *Contours of the error and constraint functions for the lasso* (left) *and ridge regression* (right). *The solid blue areas are the constraint regions,* $|\beta_1|+|\beta_2| \leq s$ *and* $\beta_1^2+\beta_2^2 \leq s$, *while the red ellipses are the contours of the RSS.*

Figure 3

250 6. Linear Model Selection and Regularization

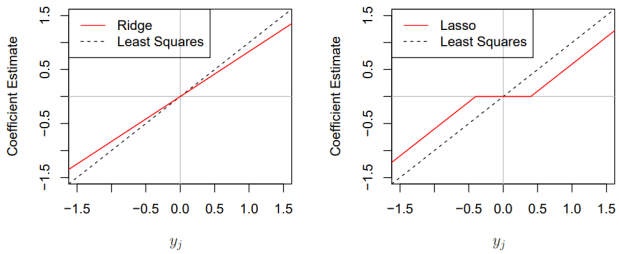

FIGURE 6.10. *The ridge regression and lasso coefficient estimates for a simple setting with* $n = p$ *and* $\mathbf{X}$ *a diagonal matrix with* 1*'s on the diagonal.* Left: *The ridge regression coefficient estimates are shrunken proportionally towards zero, relative to the least squares estimates.* Right: *The lasso coefficient estimates are soft-thresholded towards zero.*

Figure 4