## Module 12: Shrinkage Methods & Dimension Reduction

| Live Sessions | Office Hours | | Dr. Woo |
|---|---|---|---|
| Aug 7: 9:00 - 9:50 am EDT (Blue), 10:00 - 10:50 am EDT (Orange) | Tue & Thu: 9:00 - 9:50 am EDT (Blue), 10:00 - 10:50 am EDT (Orange). | | yjw4b@virginia.edu |

Module 12 is not covered in the Summer semester. The materials are up for your reference.

### MODULE WELCOME

In module 5, you learned about multicollinearity, which happens when some of the predictors are linearly dependent on each other. An impact of multicollinearity is that the estimated regression model is unstable; the removal or addition of an observation can severely influence the estimated regression model. The estimated coefficients also tend to be large numerically. In module 5, you also learned how to use variance inflation factors (VIFs) to identify the presence of multicollinearity, and how to remove some predictors to resolve multicollinearity.

In module 12, you will learn about two shrinkage methods: ridge regression and lasso. Shrinkage methods help us deal with the issue of stability in estimating the regression model. Shrinkage methods typically introduce a little bit of bias into our estimates while at the same time diminishing the variance of the estimates, sometimes by a lot, which leads to estimates that are much more stable.

Principal Component Analysis (PCA) is a method for dimension reduction when we have many predictors. PCA seeks to find a set of linearly uncorrelated variables that accounts for as much variability in the data as possible. The number of linearly uncorrelate variables is typically a lot smaller than the number of predictors.

In this module, you will complete and submit Project #2.

### ESSENTIAL QUESTIONS

- *What are the consequences of multicollinearity?*
- *How do we use the bias-variance trade-off in ridge regression and lasso?*
- *What are the pros and cons associated with principal component analysis (PCA)?*

### LEARNING OBJECTIVES

1. Describe the setup of ridge regression, lasso, and principal component regression.

2. Describe the bias-variance trade-off in model estimation.

### ASSIGNED RESOURCES

- *Introduction to Linear Regression Analysis*, Sections 9.5.3 to 9.5.4.
- Reading: ridge_lasso (PDF)
- Module 12 R tutorial, data set: meatspec from the faraway package in R, and R-code: tutorial_module 12.R

### OPTIONAL

Visit the Module 12 Discussion Forum to communicate with one another about the assigned readings and resources or to discuss any other topics of interest with your instructor or fellow students.

💬 Module 12 General Discussion

### MODULE OVERVIEW

📂 12.1: Introduction to the Lesson

📂 12.2: Ridge Regression

📂 12.3: The Lasso

📂 12.4: Principal Component Analysis

📂 12.5: R Tutorial for Module 12

📂 12.6: Module 12 Live Session

📂 12.7: Optional - Module 12 Assignments Due