

4.1.1: R - Penalized Regression Methods (Ridge Regression, LASSO, and Elastic Net)

Stat 5100: Dr. Bean

Example: (Ridge Regression; recall Handout 2.6.1 example) A study seeks to relate (in females) amount of body fat (Y) to triceps skinfold thickness (X_1), thigh circumference (X_2), and midarm circumference (X_3). Amount of body fat is expensive to measure, requiring immersion of person in water. This expense motivates the desire for a predictive model based on these inexpensive predictors.

```
# Load the data
library(stat5100)
data(bodyfat)

# Look at the original fit along with VIF:
bodyfat_lm <- lm(body ~ triceps + thigh + midarm, data = bodyfat)
olsrr::ols_vif_tol(bodyfat_lm)

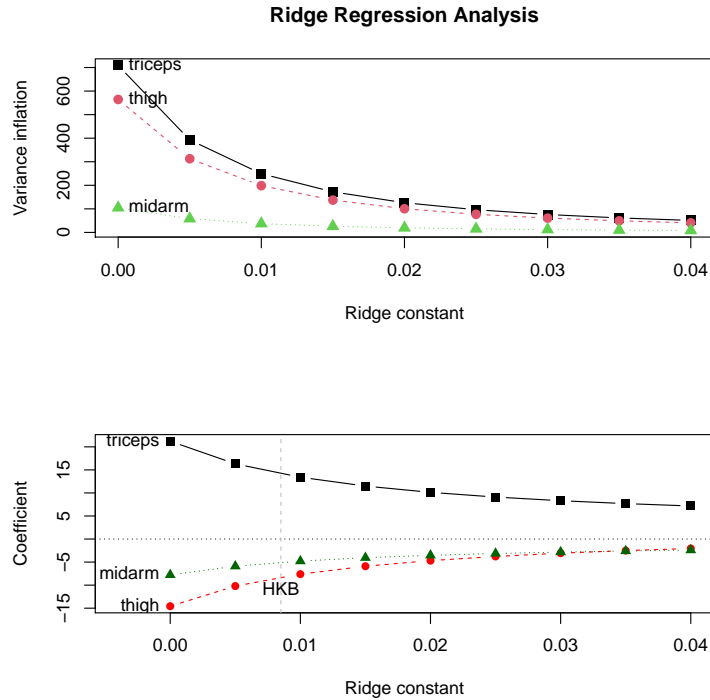
##   Variables   Tolerance    VIF
## 1   triceps 0.001410750 708.8429
## 2    thigh 0.001771971 564.3434
## 3   midarm 0.009559681 104.6060

# Try ridge regression as a remedial measure
# -----
# We use the ridge() function inside the genridge package to do this. Note that
# instead of specifying our model using a formula (formulas in R are of the
# form  $Y \sim X_1 + X_2 + X_3$ ), we create a dataframe of just our predictor variables
# and a vector of our response variable. There are many great packages to
# perform ridge regression with, but this one was selected for this course
# due to the greater plotting features available in this particular package.
y <- bodyfat$body

# Our X must come in the form of a matrix. First we take out the "body" column
# from the dataframe, and then we convert it to a matrix.
X <- as.matrix(subset(bodyfat, select = -body))

# Range of lambda values to try
lambda_seq <- seq(0, 0.04, by = 0.005)

# Create the model and plot to check for an optimal value of lambda
bodyfat_test_ridge_lm <- genridge::ridge(y, X, lambda = lambda_seq)
stat5100::ridge_vif_trace_plot(bodyfat_test_ridge_lm)
```



```
a <- glmnet::glmnet(X, y, alpha = 0, lambda = lambda_seq)
```

The plot on top is the VIF plot, and the plot on the bottom is the ridge trace plot. We look for the smallest ridge constant such that the VIF is “small enough” (roughly around 1 if possible) and the coefficient estimates have become “stable” (usually implying that the coefficient estimate curve has leveled off)

Here, we will pick $\lambda = 0.02$. Let’s fit a model with this ridge constant and then look at the coefficient estimates.

```
bodyfat_ride_lm <- genridge::ridge(y, X, lambda = 0.02)
bodyfat_ride_lm$coef
```

```
##      triceps      thigh      midarm
## 0.02 10.12698 -4.682273 -3.52701
```