

STAT 452, HW 6

Due: Thursday, March 25

Reading: Chapter 6, pp. 203–228, from *An Introduction to Statistical Learning*

Exercise 1. Answer the following as TRUE or FALSE.

- (a) Lasso regression performs feature selection.
- (b) Ridge regression performs feature selection.
- (c) Ordinary least squares regression can overfit the data when the number of features p is large.
- (d) As the tuning parameter λ gets large, the coefficient estimates in ridge regression get shrunk towards zero.
- (e) Lasso will always perform better than ridge regression on withheld test data.

For the following exercises use the `College` data set from the ISLR library.

```
library(glmnet)
library(ISLR)
data("College")
```

You can read about this data set by typing `help(College)` in the console.

Exercise 2. Fit a linear regression model with `Apps`, the number of applications received, as the response and all the other variables in the data set as predictors. Then implement the following variable selection techniques on this data:

- (a) Backwards stepwise selection using the AIC.
- (b) Backwards stepwise selection using the BIC.

Which method selects the smaller set of variables? How do the selected models compare in terms of the R^2 on the training data?

Exercise 3. Fit a lasso model for `Apps`, using all the other variables as predictors.

- (a) Make a plot the coefficient paths for the lasso model.
- (b) Print out the table of coefficient estimates that corresponds to the “optimal” value of λ selected, internally by the software, using cross-validation. How many coefficients are set exactly equal to zero?

Bonus. [1 point] Repeat exercise 3(a), but this time exclude the predictor `Private` from the model.