NOTES:

- Please start the homework as early as possible. You can discuss the homework with your classmates, TAs and the instructor. But you need to hand in an independent copy by yourself. Anything related to plagiarism can lead to zero score in the grade.
- Submit hard copy of your homework before class:
  - Your answers to the problem set
  - Please also attach the code
- The abbreviation ISL refers to the book `http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf`. The abbreviation ESL refers to the book `https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf`.

1. Theoretical Problems

1. *Methods between Ridge and LASSO.* Consider the elstic-net optimization problem

$$\min_{\beta} \|y - X\beta\|^2 + \lambda\left(\alpha\|\beta\|_2^2 + (1-\alpha)\|\beta\|_1\right). \tag{1.1}$$

Show how one can turn this into a lasso problem, using an augmented version of X and y. (This is Ex. 3.30 in ESL.)

2. *Closed forms of LASSO, Ridge and Best Subset Selection.* Consider a special case of linear regression

$$y = X\beta + \epsilon,$$

where the columns of $\boldsymbol{X}$ are orthonormal, that is $X^T X = \mathrm{I}$. Let $\hat{\beta}$ denote the least square estimators and $\hat{\beta}_{(M)}$ denote the M-th largest absolute value of $\hat{\beta}$, that is $\left|\hat{\beta}_{(1)}\right| \geq \left|\hat{\beta}_{(2)}\right| \geq \cdots \geq \left|\hat{\beta}_{(p)}\right|$. Show the corresponding formulas for Best subset with size M, Ridge and Lasso estimator,

Estimator	Formula
Best subset (size M)	$\hat{\beta}_j \cdot I(\lvert\hat{\beta}_j\rvert \geq \lvert\hat{\beta}_{(M)}\rvert)$
Ridge	$\hat{\beta}_j/(1+\lambda)$
Lasso	$\mathrm{sign}(\hat{\beta}_j)(\lvert\hat{\beta}_j\rvert - \lambda)_+$

2. Applied Problems

3. *Bootstrap with Least squares, Ridge and Lasso.* Let $\beta = (\beta_1, \beta_2, \cdots, \beta_p)$ and let $\mathbf{x}, y$ be random variables such that the entries of $\mathbf{x}$ are i.i.d. standard normal random variables (i.e., with mean zero and variance one) and $y = \beta^T \mathbf{x} + \epsilon$ where $\epsilon \sim \mathcal{N}(0, 1)$.

(a) Simulate a dataset $(\mathbf{x}_1, y_1), ...(\mathbf{x}_n, y_n)$ as n i.i.d. copies of the random variables $\mathbf{x}, y$ defined above, with $n = 800, p = 200$ and $\beta_j = j^{-1}$.

(b) The goal of this problem is to construct confidence intervals for β_1 using Bootstrap method.

  (a) Construct confidence intervals for β_1 by bootstrapping the data and applying Least Squares to the bootstrapped data set.

  (b) Construct confidence intervals for β_1 by bootstrapping the data and applying Ridge to the bootstrapped data set.

  (c) Construct confidence intervals for β_1 by bootstrapping the data and applying Lasso to the bootstrapped data set.

(c) Comment on the obtained results.

4. *Shrinkage Method and Dimension Reduction.* Problem 9 at page 263 of ISL