# Math 680, Fall 2020
# Homework 4 Due: Friday, 11/20/2020

**General Instructions:**

- Turn in all your HW through Canvas.
- All the HW files (except the R code) should be saved as a single PDF, and named in the form "Last-name_hw4.pdf".
- The code should be saved as "Last-name_hw4_code.r".
- Test your R code before submission to make sure it can be executed successfully by the "source()" function.

1. Consider the following LASSO problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2, \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \le t,$$

where $t \ge 0$ is a constant.

(a) If $t = 0$, compute $\hat{\beta}_j^{lasso}$, for $j = 1, \dots, p$.

(b) Define $t_0 = \sum_{j=1}^{p} |\hat{\beta}_j^{\text{ols}}|$. Prove that, if $t \ge t_0$, then

$$\hat{\beta}_j^{\text{lasso}} = \hat{\beta}_j^{\text{ols}}.$$

2. Assume $X^T X = I$. Prove the following problems are equivalent:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}(y_i - \mathbf{x}_i^T \beta)^2 + \sum_{j=1}^{p} J(\beta_j)$$

$$\min_{\beta_j} (\hat{\beta}_j^{\text{ols}} - \beta_j)^2 + J(\beta_j), \quad \text{for } j = 1, \dots, p,$$

where $J$ is a penalty function.

3. Consider the linear regression model
$$Y = \sum_{j=1}^{4} X_j \beta_j + \epsilon.$$
Assume that $X$ is orthonormal matrix. Suppose we fit the OLS and obtain $\widehat{\boldsymbol{\beta}}^{ols} = (1.1, -0.8, 0.3, -0.1)$.

   (a) Compute $\widehat{\boldsymbol{\beta}}^{ridge}$ for $\lambda = 1$ and $\lambda = 0.4$.

   (b) Compute $\widehat{\boldsymbol{\beta}}^{lasso}$ for $\lambda = 1$ and $\lambda = 0.4$.

   (c) Compare (1) and (2) and make comments.

4. Download the *prostate* cancer data set from the website http://statweb.stanford.edu/∼tibs/ElemStatLearn/. The data set contains eight predictors (columns 1-8), $\mathbf{X} \in R^8$. The outcome variable $Y$ is given by column 9. The last column (column 10) is the train/test indicator, indicating 67 "training" data observations and 30 "testing" observations. Let $n = 67$ and $\tilde{n} = 30$. Denote the training set by $\{(\mathbf{x}_i, y_i), i = 1, \cdots, n\}$ and the test set by $\{(\tilde{\mathbf{x}}_i, \tilde{y}_i), i = 1, \cdots, \tilde{n}\}$.

   **Analysis**: Consider the linear regression of $Y$ on $\mathbf{X}$. Use the training set to fit a regression model, $\hat{f}(\mathbf{x}) = \hat{\beta}_0 + \hat{\beta}^T \mathbf{x}$. For any fitted model $\hat{f}(\mathbf{x})$, calculate its "training error" by $TrainErr = \frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{f}(\mathbf{x}_i)]^2$, and its "test error" by $TestErr = \frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}}(\tilde{y}_i - \hat{f}(\tilde{\mathbf{x}}_i)]^2$. Please complete the following.

   (a) Fit the standard linear regression model using the ordinary least squares. Report its $R^2$, p-values of regression coefficients, the set of significant predictors (at the level $\alpha = 0.05$), $TrainErr$, and

$TestErr$. You can use R functions "lm()" and "summary()" to do the analysis.

(b) Apply forward selection to select variables use R function "regsubsets()" in the package "leap". You should get a sequence of eight models, $\widehat{M}_1, \cdots, \widehat{M}_8$, in the increasing order of model size. For each model $\widehat{M}_j, j = 1, \cdots, 8$, report its regression coefficients, calculate its $TrainErr$, and calculate its BIC using the formula

$$BIC(\widehat{M}_j) = n \log(TrainErr) + \log(n)|\widehat{M}_j|,$$

where $|\widehat{M}_j|$ is the number of variables in the model (including the intercept). Choose the best model by minimizing BIC, and report the set of important variables selected by BIC. Furthermore, use the selected variables to refit the OLS and report TestErr.

(c) In part (b), replace BIC by AIC,

$$AIC(\widehat{M}_j) = n \log(TrainErr) + 2|\widehat{M}_j|.$$

Choose the best model by minimizing AIC, and report the set of selected variables. Furthermore, use the selected variables to refit the OLS and report TestErr.

5. Fit the LASSO regression for the prostate cancer data set. You can use R functions "lars()" and "cv.lars()"

(a) Select the parameter with 5-fold CV, using the minimum CV rule. Report the best $\lambda$, the selected model, the estimated regression coefficients, and the TestErr.

(b) Select the parameter with 5-fold CV, using the one-standard deviation rule for CV. Report the best $\lambda$, the selected model, the estimated regression coefficients, and the TestErr.