Assignment 3

February 24, 2016

1 RIDGE REGRESSION, LASSO, AND MODEL SELECTION

Problem 1 It is well-known that ridge regression tends to give similar coefficient values to correlated variables, whereas the lasso may give quite different coefficient values to correlated variables. We suppose that n=2, p=2, $x_{11} = x_{12}, x_{21} = x_{22}$. Also suppose that $y_1 + y_2 = 0$, $x_{11} + x_{21} = 0$, and $x_{12} + x_{22} = 0$, so that the estimate for intercept in a least square, ridge regression, or lasso model is 0, in other words, $\hat{\beta}_0 = 0$.

- Write the ridge regression optimization problem in this setting.(3pts)
- Show that the ridge coefficient estimates staisfy: $\hat{\beta}_1 = \hat{\beta}_2$ (3pts).
- Write the lasso optimization problem in this setting. (3pts)
- Show that in this setting, the lasso coefficients $\hat{\beta}_1, \hat{\beta}_2$ are not unique, in other words, there are many possible solutions to the optimization problem. Describe these solutions. (5pts).

Problem 2 Suppose we estimate the regression coefficients in a linear regression model by minimizing:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \text{ such that: } \sum_{i=1}^{n} |\beta_j| \leq s$$

for $s \geq 0$. This is the constrained formulation of the Lasso.

- As s increases from 0 to ∞, what can you say about the training residual sum of squares (RSS). Support your choice.(4pts)

1. Increase intially, then start dereasing in an inverted U shape.
2. Decrease intially and, then eventually start increasing in a U shape.
3. Steadily in crease.
4. Steadily decrease.
5. Remain constant.

- As s increases from 0 to ∞, what can you say about the testing RSS. Support your choice. (4pts)
- As s increases from 0 to ∞, what can you say about the variance of the estimator. Support your choice. (4pts)
- As s increases from 0 to ∞, what can you say about the bias of the estimator. Support your choice. (4pts)

Problem 3 Suppose you have a data set with $p \gg n$ (the number of predictors is much bigger than the number of observations). You have a design matrix X and a quantitative response vector y. You plan to fit a linear regression model.

- Is the ordinary least square solution unique? Why? (3pts)
- Is the ridge regression solution unique? Why? (3pts)
- Suppose you compute the ridge regressio solution $\hat{\beta}(\lambda)$ for each value of λ and data X,y. Let λ decreases from ∞ to 0. What can you say about the ridge solution when $\lambda \downarrow 0$. (3pts).

Problem 4 You have the standard linear regression setting, matrix of predictors $X \in \mathscr{R}^{n\times p}$, the columns of this matrix is: $x_1, x_2, \cdots, x_p$. Let $\hat{\beta} \in \mathscr{R}^p$ be least square solution:

$$\hat{\beta} = \min_{\beta} \frac{1}{2} \| y - X\beta \|_2^2.$$

- Show that, if a vector $v \in \mathscr{R}^p$ such that: $Xv = 0$ then $\hat{\beta} + c.v$ is also a least square solution, for any value of $c \in \mathscr{R}$. (4pts)
- If the columns of X: $x_1, x_2, \cdots, x_p$ are linearly independent, which vectors $v \in \mathscr{R}^p$ satisfy $Xv = 0$. (4pts)

2 MATLAB PROGRAMMING

Problem 1 Write your own Matlab function to perform best subset selection: best_set.m.

- Inputs are: Training data with predictors matrix: $X\_train \in \mathscr{R}^{n\times p}$,y_train, and validating data: $X\_val \in \mathscr{R}^{n\times p}$, y_val. (2pts).

- For $k = 1, 2, \cdots, p$, perform linear regression on all models that contain exactly k predictors. (5pts)
- Among these models, pick the one with largest R^2. (3pts)
- Among p best models, choose the one that has the best validating RSS on the validation set. (3pts)
- Output the best linear model. (3pts)
- You can use Matlab function fitlm.m to fit linear models. You can also use function combnk(1:p,k) that gives all possible combinations of subset of size k of numbers from 1 to p.

Problem 2 Using the Boston house price data set in *housing.mat*, we are going to compare the performance of linear regression, ridge, and lasso. The description of the data is provided in the file *housing.name.txt*. Notice that in the description, the last attributes (house price) is seperated to vector y in the data file. The data sets have 506 observations and 13 predictors. Submit your code in matlab file: boston_housing.m.

- Divide the data set randomly into 3 equal parts for training, validating, and testing. You can use crossvalind or any other options you want. Each part will have roughly 130 observations. (3pts)
- Build a linear regression model using the training data using fitlm in Matlab. (3pts)
- Report estimated coefficients, their standard error, and statistical significance of each predictors. Is the model significance? (Perform F-test). (8 pts)
- Perform ridge regression on the training data on a grid of 40 values of tuning parameters λ. Choose the values of λ so they are equidistant on the log-scale, ranging from a small value to a large one. Something like: lambda=-6:12/40:6. lambda=exp(lambda), would work. This gives you a grid of values of λ ranging from e^{-6} to e^{6}. For each value of λ, compute the RSS on the validating set. Plot RSS on validating set vs. value of λ. Report the value of λ that gives the smallest value of RSS on validating set. (10pts)
- Perform lasso regression on the training data on a grid of 40 values of tuning parameters λ. Choose the values of λ so they are equidistant on the log-scale, ranging from a small value to a large one similar to the previous part. For each value of λ, compute the RSS on the validating set. Plot RSS on validating set vs. value of λ. Report the value of λ that gives the smallest value of RSS on validating set. (10pts)
- Compare the RSS of three methods linear regression, ridge regression, and lasso regression with optimal choice of tuning parameters λ that you obtain from previous part. Draw conclusion on which predictors are chosen by each method. (8pts).