Solution method for Lasso

- Gradient descent method can be modified to obtain solution for the Lasso, although Lasso has a nondifferentiable objective function.
- Soft thresholding is a very immportant operator that is used in most solver for Lasso.
- Today: Resampling methods for regularization.

Resampling methods

- Resampling methods involve repeatedly drawing samples from a training set and refitting a model of interest on each example to gain additional information about the fitted model.
- Cross-validation: can be used to estimate the test error associated with a statistical learning method or to choose the appropriate level of flexibility.
- Cross validation can also be used to find the right amount of regularization in shrinkage methods.

Cross-validation

- Given a data set, a statistical learning method is considered "good" if it results in low test error. (prediction of the response on a new observation is good).
- Training error is easily obtained but it usually underestimates the test error.
- Cross validation methods estimate the test error by holding out a subset of training data from the fitting process and apply the learning method on those hold out observations.

Validation set approach

- Randomly vividing the available set of observation into two parts: training set and validation set (or hold -out set).
- Model is fit on the training set.
- Fitted model is used to predict the responses for the observations in the validation set.
- Resulting validation set error rate is an estimate of the test error rate.

Validation set approach

- Demonstration in Matlab
- Estimate using validation set can have high variance, depending on precisely which observations are included in the training set and validation set.
- With less training data, validation set error tend to overestimate the test error rate.

Leave one out cross validation(LOOCV)

- A single observation (x_1, y_1) is used for validation, $(x_2, y_2), \cdots, (x_n, y_n)$ are used for training. Fitted model is used on (x_1, y_1) to obtain:$MSE_1 = (y_1 - \hat{y_1})^2$.
- Repeat the process n times, each time with (x_i, y_i) as validation set, we get $MSE_1, MSE_2, \cdots, MSE_n$. Cross validation error for LOOCV is:

$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^{n} MSE_i.$$

- LOOCV has far less bias. It does not overestimate test error as much as validation set approach.
- LOOCV uses every observation as validation set so there is no randomness in the result.

k-fold cross validation

- LOOCV is expensive in terms of computation.
- For k fold cross validation, the set of observations is randomly divided into k groups of approximately equal size.
- The first group is held out as validation set, the rest k-1 groups are used as training set.MSE_1 is computed on the held out group.
- The process is repeated k times for k groups. The k fold CV estimate is computed by:

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^{k} MSE_k.$$

k fold cross validation

- Compared to LOOCV, k fold cross validation has to do much less computation.
- LOOCV and k fold cross validation have very similar estimates on MSE on testing data.
- K fold CV can be used to determined the location of the minimum point in the estimated test MSE curve. (how much flexibility our model should have to get the best test error MSE).
- K fold CV estimates of test MSE has more bias than LOOCV, but less variance.

Select tuning parameters using CV

- Choose a grid of λ values, and compute the cross validation error for each value of λ.
- Choose the value of tuning parameter λ that yields that smallest cross validation error.
- The model is refit using all of the available observations with the selected value of tuning parameter.

Ridge regression in Matlab

- Ridge regression with Matlab: Matlab command for ridge regression: ridge(X,y,k). X and y are matrix of predictors and responses. k is the tuning parameter.
- First generate a data set for training and testing using simulate_ridge.m calling them X and X_test.
- In this example we will implement 5 fold cross validation on the training set.
- Matlab command crossvalind will randomly split the data into 5 groups of approximately equal size. The syntax is: crossvalind('Kfold',N,k) where N is the number rows of the data and k is the number of groups. The output is a vector of numbers between 1 to 5.

Ridge and Lasso in Matlab

- Tuning parameter λ chosen over a grid with increasing values.
- A ridge regression model is fit with the training data and a tuning parameter lambda. MSE of the model with validation data is then calculated.
- This process is repeat with all 5 groups.
- The value of λ that yields smallest MSE is selected to fit a regression model with all original training data. Test error with unknown observation can then be evaluated.

Lasso regression in Matlab

- Lasso regression with Matlab: Matlab command for ridge regression: [beta,fitinfo]=lasso(X,y). X and y are matrix of predictors and responses.
- First generate a data set for training and testing using simulate_ridge.m calling them X and X_test.
- In this example we will implement 5 fold cross validation on the training set.
- Matlab command crossvalind will randomly split the data into 5 groups of approximately equal size. The syntax is: crossvalind('Kfold',N,k) where N is the number rows of the data and k is the number of groups. The output is a vector of numbers between 1 to 5.

Ridge and Lasso in Matlab

- Tuning parameter λ chosen over a grid with increasing values.
- A lasso regression model is fit with the training data and a tuning parameter lambda. MSE of the model with validation data is then calculated.
- This process is repeat with all 5 groups.
- The value of λ that yields smallest MSE is selected to fit a regression model with all original training data. Test error with unknown observation can then be evaluated.