Ch. 6.2 Lasso and Ridge Regression - Problem Bank Questions

October 9, 2019

1. Ridge Regression Conceptual Questions: Answer the following questions. For the true and false section, explain your answer.

(a) Explain in laymen's terms how and why the ridge regression penalty results in smaller coefficient estimates.
(b) Explain in laymen's terms what parameter shrinkage means and why we might want to shrink coefficients.
(c) Explain in laymen's terms why we do not want to shrink the intercept term with ridge regression.
(d) How might we go about selecting the λ tuning parameter for ridge regression?
(e) True or False: Ridge regression applies an L2 (quadratic) penalty to the coefficients.
(f) True or False: As $\lambda \to \infty$, ridge regression recovers the least squares estimate.

2. Ridge regression true and false. For each, explain your answer.
(a) T/F: ridge regression coefficients estimates are scale invariant.
(b) T/F: predictors should be standardized before applying ridge regression.
(c) T/F: As λ increases, the bias of the model increases.
(d) T/F: Increasing λ can decrease the variance of the predictor by a larger amount that the bias increases.

3. For the following scenarios, describe whether you would fit a least squares regression or ridge regression and why.
(a) The number of predictors p is larger than the number of training data observations n.
(b) The true model is very close to linear and the observation noise is very low.
(c) The true model is approximately linear but there is high observation noise.

4. True or False: Explain each of your answers.

(a) T/F: Ridge regression will set parameters equal to 0 for finite λ.
(b) T/F: The Lasso applies an L1 penalty to the coefficients.
(c) T/F: The Lasso results in sparse coefficient estimates.
(d) T/F: As $\lambda \to 0$, the lasso finds the null model solution.

5. Many machine learning models involve fitting models via optimization. A loss function is specified and then this quantity is minimized to find the optimal setting of the parameters. For example, for the lasso, we want to minimize:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right) + \lambda\sum_{j=1}^{p}|\beta_j|.$$

To find the optimal values of the parameters in our model, we often need to find the gradient of the loss function. Explain conceptually how optimizing the ridge regression objective is different from optimizing the lasso objective. Do you think it is easier to find the gradient of the ridge regression objective or the lasso objective? Why?

6. For the following scenarios, select whether you would choose to perform ridge regression or lasso and specify why.

(a) You think that the true number of predictors in the model is much smaller than the total possible number of predictors.
(b) You want to build a model that is easily interpretable.
(c) You think that the true model consists of many coefficients, all of approximately the same magnitude.
(d) You want to determine which features are important in your problem.

Problem Bank: Model Selection

November 6, 2019

1. Ridge regression minimizes the value of the loss function

$$RSS + \lambda \sum_{j=1}^{p} \beta_j^2$$

where p is the number of predictors and the β_j are the coefficients of the model.

(a) Explain how the parameter λ affects the ridge regression, making sure to reference the above formula. Why is this process sometimes referred to as "shrinkage"?

(b) Describe a process for "tuning" or "choosing the best value of" λ.

(c) Is the intercept term included in the summation on the right side of the loss function? Why or why not?

2. Ridge regression minimizes the value of the loss function

$$RSS + \lambda \sum_{j=1}^{p} \beta_j^2$$

(a) Is it important to scale your variables before running ridge regression? Why or why not? Please explain using the formula above.

(b) Is scaling your variables also important for the lasso? Why or why not?

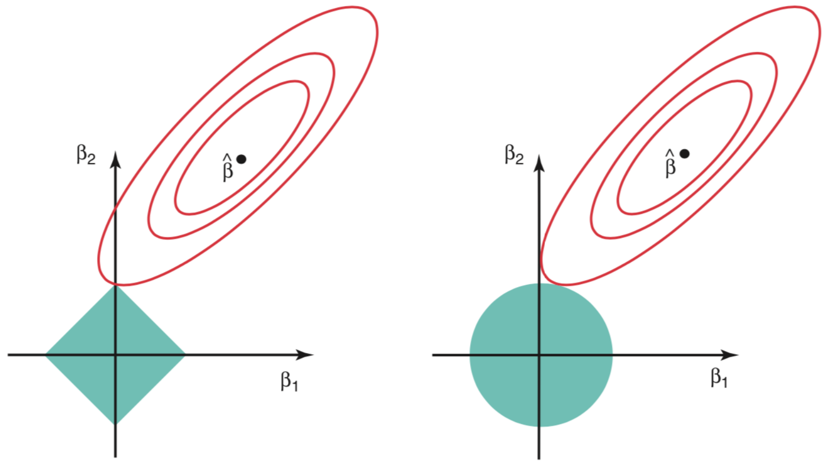
(c) Is scaling your variables also important for ordinary least squares? Why or why not?

3. Say whether each of the following statements about ridge regression is true or false:

   (a) As λ decreases, the flexibility of the model decreases.

   (b) As λ increases, the variance of the model decreases.

   (c) Ridge regression will automatically remove the least useful predictors from the model.



4. Contours of the error and constraint functions for the lasso (left) and ridge regression (right) are shown in the figure above. The solid blue areas are the constraint regions, $\mid \beta_1 \mid + \mid \beta_2 \mid \leq s$ and $\beta_1^2 + \beta_2^2 \leq s$ while the red ellipses are the contours of the RSS.

   (a) Explain why the constraint regions are representations of the L_1 vs. L_2 norms.

   (b) Does this give you some intuition behind why the lasso performs variable selection and ridge regression does not?

5. The lasso minimizes the loss function

$$RSS + \lambda \sum_{j=1}^{p} \mid \beta_j \mid$$

(a) Explain the difference between ridge regression and the lasso, based on their loss functions.

(b) In a situation where you have many predictors but a small amount of data points, do you expect ridge regression or the lasso to perform better?

6. Which one of the following statements is NOT true?

   (a) Tuning λ is essential if you want to use the lasso.

   (b) The lasso uses the L2 penalty, while ridge regression uses the L1 penalty.

   (c) Models produced by the lasso are generally more interpretable than those produced by ridge regression.

   (d) The lasso can be used in a similar way as best subset selection.