Question 2

October 7, 2019

It is well-known that ridge regression tends to give similar coefficient values to correlated variables, whereas lasso may give quite different coefficient values to correlated variables. We will now explore this property in a very simple setting. Assume n are the number of training samples, p the number of dimensions, x in the input and y the output.

0.0.1 (a) Write out the ridge regression optimization problem in this setting.

According to the optimization function in the slide:

$$\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

We can get the optimization problem in the provided setting:

$$(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2).$$

0.0.2 (b) Argue that in this setting, the ridge coefficient estimates satisfy $\beta_1 = \beta_2$.

For optimizing the problem, we need to take the derivatives of the above expression in respect with $\beta_1 and \beta_2$

We first take the derivative of the expression in respect with β_1:

$$x_1(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1) + x_2(y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2) + \lambda\hat{\beta}_1 = 0$$

And then, take the derivative of the expression in respect with β_2:

$$x_1(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1) + x_2(y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2) + \lambda\hat{\beta}_2 = 0$$

By substracting the above two equations, we can get:

$$\hat{\beta}_1 = \hat{\beta}_2$$

0.0.3 (c) Write out the lasso optimization problem in this setting.

According to the optimization function in the slide:

$$\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

We can get the optimization problem in the provided setting:

$$(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda(|\hat{\beta}_1| + |\hat{\beta}_2|).$$

0.0.4 (d) Argue that in this setting, the lasso coefficients $\beta_1 and \beta_2$ are not unique – in other words, there are many possible solutions to the optimization problem. Describe these solutions.

Here, we use the other form of the Lasso function:

$$(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 \text{ subject to } |\hat{\beta}_1| + |\hat{\beta}_2| \leq s$$

We then can use the given setting and make the equation be like this:

$$2[y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_1]^2 \geq 0.$$

We can get that the solutions of optimizing lasso is $|\hat{\beta}_1| + |\hat{\beta}_2| \leq s$, which is a bunch of solutions.

[ ]: