Elements of Machine Learning
Assignment 3

William LaCroix (7038732)
wila00001@stud.uni-saarland.de

Problem 3 (T, 10 Points). Correlated Variables

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 will now explore this property in a very simple setting. Suppose that $n = 2, p = 2, x_{11} = x_{12}, x_{21} = x_{22}$. Furthermore, suppose that $y_1 + y_2 = 0, x_{11} + x_{21} = 0$ and $x_{12} + x_{22} = 0$, so that the estimated intercept in a least squares, ridge regression, or lasso model is zero, $\hat{\beta}_0 = 0$.

1. [2pts] Write out the ridge regression optimization problem explicitly in this setting.

ISLR 6.5 gives us:

$\Sigma_{i=1}^n \left(y_i - \beta_0 - \Sigma_{j=1}^p \beta_j x_{ij}\right)^2 + \lambda \Sigma_{j=1}^p \beta_j^2$, and since $\hat{\beta}_0 = 0$, we can reduce this somewhat to:
$\Sigma_{i=1}^n \left(y_i - \Sigma_{j=1}^p \beta_j x_{ij}\right)^2 + \lambda \Sigma_{j=1}^p \beta_j^2$, which we start by expanding the sums over p_j:
$\Sigma_{i=1}^n \left(y_i - (\beta_1 x_{i1} + \beta_2 x_{i2})\right)^2 + \lambda\left(\hat{\beta}_1^2 + \hat{\beta}_2^2\right)$, and finally expand the sums over n_i:
$\left(y_1 - \left(\hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12}\right)\right)^2 + \left(y_2 - \left(\hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22}\right)\right)^2 + \lambda\left(\hat{\beta}_1^2 + \hat{\beta}_2^2\right)$. Since $x_{11} = x_{12}, x_{21} = x_{22}$, we can reduce and rearrange the inner terms, such that:
$\left(y_1 - x_1\left(\hat{\beta}_1 + \hat{\beta}_2\right)\right)^2 + \left(y_2 - x_2\left(\hat{\beta}_1 + \hat{\beta}_2\right)\right)^2 + \lambda\left(\hat{\beta}_1^2 + \hat{\beta}_2^2\right)$.

2. [2pts] What does the solution space for ordinary least squares regression look like here?

OLS ($\lambda = 0$) seeks to minimize over the above, and since $y_1 + y_2 = 0$ and $x_1 + x_2 = 0$, we are trying to find $\hat{\beta}_1$ and $\hat{\beta}_2$ such that
$\left(y - x\left(\hat{\beta}_1 + \hat{\beta}_2\right)\right)^2 + \left(-y + x\left(\hat{\beta}_1 + \hat{\beta}_2\right)\right)^2 = 0$, solveable at $\hat{\beta}_1 + \hat{\beta}_2 = \frac{y}{x}$, we get unique solutions:
$\hat{\beta}_1 = \frac{y}{x} - \hat{\beta}_2$ and $\hat{\beta}_2 = \frac{y}{x} - \hat{\beta}_1$

3. [2pts] Show that in this example, the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$

Expanding the squares from part 1, we get:
$y_1^2 - 2y_1x_1\left(\hat{\beta}_1 + \hat{\beta}_2\right) + x_1^2\left(\hat{\beta}_1^2 + 2\hat{\beta}_1\hat{\beta}_2 + \hat{\beta}_2^2\right) + y_2^2 - 2y_2x_2\left(\hat{\beta}_1 + \hat{\beta}_2\right) + x_2^2\left(\hat{\beta}_1^2 + 2\hat{\beta}_1\hat{\beta}_2 + \hat{\beta}_2^2\right) + \lambda\left(\hat{\beta}_1^2 + \hat{\beta}_2^2\right)$,

followed by some truly unfortunate distribution:
$y_1^2 - 2y_1x_1\hat{\beta}_1 - 2y_1x_1\hat{\beta}_2 + x_1^2\hat{\beta}_1^2 + 2x_1^2\hat{\beta}_1\hat{\beta}_2 + x_1^2\hat{\beta}_2^2 +$
$y_2^2 - 2y_2x_2\hat{\beta}_1 - 2y_2x_2\hat{\beta}_2 + x_2^2\hat{\beta}_1^2 + 2x_2^2\hat{\beta}_1\hat{\beta}_2 + x_2^2\hat{\beta}_2^2 +$
$\lambda\hat{\beta}_1^2 + \lambda\hat{\beta}_2^2$
And since our goal is to minimize this function wrt. $\hat{\beta}_1$ and $\hat{\beta}_2$, we can take the derivative in each case and set it to zero. (continued on next page

First for $\hat{\beta}_1$, we get:
$-2y_1x_1 + 2x_1^2\hat{\beta}_1 + 2x_1^2\hat{\beta}_2 - 2y_2x_2 + 2x_2^2\hat{\beta}_1 + 2x_2^2\hat{\beta}_2 + 2\lambda\hat{\beta}_1 = 0$
Now, for $\hat{\beta}_2$:
$-2y_1x_1 + 2x_1^2\hat{\beta}_1 + 2x_1^2\hat{\beta}_2 - 2y_2x_2 + 2x_2^2\hat{\beta}_1 + 2x_2^2\hat{\beta}_2 + 2\lambda\hat{\beta}_2 = 0$
If we subtract these from each other, we end up with $2\lambda\hat{\beta}_1 = 2\lambda\hat{\beta}_2$,
or $\hat{\beta}_1 = \hat{\beta}_2$

4. [4pts] Explain how this example connects to the statement that ridge regression tends to give similar coefficient values to correlated variables.

The setup for this problem gave us $n = 2, p = 2, x_{11} = x_{12}, x_{21} = x_{22}$ and $y_1 + y_2 = 0, x_{11} + x_{21} = 0$ and $x_{12} + x_{22} = 0$. Being equal or opposites, these observations are perfectly correlated (1:1), and in part 3 we showed that the $\hat{\beta}_1$ and $\hat{\beta}_2$ were equal. In a scenario like ours with perfect correlation, we ended up with equal coefficients. In natural data sets, ie. those with high (but not 1:1) correlation, we would expect similar, but not equal regression coefficients.