Name: Hamed Vaheb

High Dimensional Statistics: Exercise Sheet 3

Exercise 1. Consider the Ridge-regression estimator, obtained replacing the L^1 penalty in the definition of the Lasso estimator with an L^2 penalty:

$$\hat{b}_{\text{Ridge}}(\lambda) := arg \min_{b \in \mathbb{R}^k} \left(\frac{||Y - Xb||_2^2}{n} + \lambda ||b||_2^2 \right)$$

Determine the explicit formula for $\hat{b}_{\text{Ridge}}(\lambda)$.

Solution. We can express the right hand side of the minimization problem as the following:

$$\begin{aligned}
\frac{||Y - Xb||_2^2}{n} + \lambda ||b||_2^2 &= \frac{1}{n} < y - Xb, y - Xb > + \lambda < b, b > \\
&= \frac{1}{n} \big((y - Xb)^T (y - Xb) \big) + \lambda b^T b \\
&= \frac{1}{n} \big(y^T y - 2b^T X^T y + b^T X^T X b \big) + \lambda b^T b
\end{aligned}$$

In order to find the minimum, we first find the critical points of the last expression, which is obtained by setting derivative of the last equality (with respect to b) to zero. Taking the derivative with respect to b yields the following:

$$\frac{\partial \frac{1}{n} \big(y^T y - 2b^T X^T y + b^T X^T X b \big) + \lambda b^T b}{\partial b} = \frac{1}{n} \big(-2X^T y + 2X^T X b \big) + 2\lambda b$$

Setting the derivative to zero yields the following:

$$\begin{aligned}
0 &= \frac{1}{n} \big(-2X^T y + 2X^T X b \big) + 2\lambda b \\
0 &= \frac{1}{n} \big(-X^T y + X^T X b \big) + \lambda b \\
\frac{1}{n} (X^T y) &= \frac{1}{n} (X^T X b) + \lambda b \\
\frac{1}{n} (X^T y) &= (\frac{1}{n} X^T X + \lambda I) b \\
(\frac{1}{n} X^T X + \lambda I)^{-1} \frac{1}{n} (X^T y) &= b^*
\end{aligned}$$

Now we have to check whether this point is minimum or maximum. For this purpose, we calculate the second derivative.

$$\frac{\partial^2 \frac{1}{n} \big(y^T y - 2b^T X^T y + b^T X^T X b \big) + \lambda b^T b}{\partial b^2} = \frac{\partial \frac{1}{n} \big(-2X^T y + 2X^T X b \big) + 2\lambda b}{\partial b}$$

Which is equal to

$$\frac{2}{n} X^T X + 2\lambda I$$

What we obtained is positive definite, therefore the critical point b^* is a minimum.

We can conclude that

$$b^* = \hat{b}_{\text{Ridge}}(\lambda) = (\frac{1}{n} X^T X + \lambda I)^{-1} \frac{1}{n} (X^T y)$$