Elements of Machine Learning

Assigment 3 - Problem 1

Sangeet Sagar(7009050), Philipp Schuhmacher(7010127)
{sasa00001,phsc00003}@stud.uni-saarland.de

January 7, 2022

Problem 1 (T, 8 Points) Ridge and Lasso Regression

1

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 estimate for the intercept in a least squares, ridge regression, or lasso model is zero: $\hat{\beta}_0 = 0$

(2 Points) Write out the ridge regression optimization problem in this setting.

given that :

$n = 2$ $\quad$ $x_{11} = x_{12} = x_{11}$ — (A)

$p = 2$ $\quad$ $x_{21} = x_{22} = x_{22}$ — (B)

and $y_1 + y_2 = 0$ — 1

$x_{11} + x_{21} = 0$ — 2

$x_{12} + x_{22} = 0$ — 3

from equation (2) and (3) we have:-

$x_{11} + x_{22} = 0$

$$\mathcal{L}_{Ridge}(\hat{\beta};\lambda) = \sum_{i=1}^{n=2}\left(y_i - \hat{\beta}_0 - \sum_{j=1}^{p=2}\hat{\beta}_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{2}\hat{\beta}_j^2$$

$$\mathcal{L}_{Ridge}(\hat{\beta};\lambda) = \left(y_1 - \cancel{\hat{\beta}_0}^{\,0} - (\hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12})\right)^2 + \left(y_2 - \cancel{\hat{\beta}_0}^{\,0} - (\hat{\beta}_2 x_{21} + \hat{\beta}_2 x_{22})\right)^2$$

$$+ \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)$$

$$\boxed{\mathcal{L}_{Ridge}(\hat{\beta};\lambda) = \left(y_1 - x_{11}(\hat{\beta}_1 + \hat{\beta}_2)\right)^2 + \left(y_2 - x_{22}(\hat{\beta}_1 + \hat{\beta}_2)\right)^2 + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)}$$

(4 Points) Proof that in this setting, the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$

Take derivative of $\mathcal{L}_{Ridge}$ wrt. $\hat{\beta}_1$, $\hat{\beta}_2$:-

$$\frac{\partial \mathcal{L}(\hat{\beta};\lambda)}{\partial \hat{\beta}_1} = 2\left(y_1 - x_{11}(\hat{\beta}_1+\hat{\beta}_2)\right)(-x_{11}) + 2\left(y_2 - x_{22}(\hat{\beta}_1+\hat{\beta}_2)\right)(-x_{22}) + 2\lambda\hat{\beta}_1 = 0$$

$$\frac{\partial \mathcal{L}(\hat{\beta};\lambda)}{\partial \hat{\beta}_2} = 2\left(y_1 - x_{11}(\hat{\beta}_1+\hat{\beta}_2)\right)(-x_{11}) + 2\left(y_2 - x_{22}(\hat{\beta}_1+\hat{\beta}_2)\right)(-x_{22}) + 2\lambda\hat{\beta}_2 = 0$$

Subtract both these equations :

$$\frac{\delta \mathcal{L}}{\delta \hat{\beta}_1} = \frac{\delta \mathcal{L}}{\delta \hat{\beta}_2}$$
$$2\lambda\hat{\beta}_1 = 2\lambda\hat{\beta}_2$$
$$\hat{\beta}_1 = \hat{\beta}_2$$

(2 Points) Explain how the explored setting connects to the statement that ridge regression tends to give similar coefficient values to correlated variables.

Correlated variables will tend to contribute equally to the response and hence ridge regression will penalize them similarly resulting in similar coefficient values.