응용통계학 과제1

(2020710815 박태익, 2021.4.16)

1 Ridge Regression

$X = U\Sigma V^T$, $U^TU = I_d$, $V^TV = I_d$ **일 때,**

1.1) 식 정리

주어진 식을 정리하면 다음과 같다.

$$
\begin{aligned}
&(\lambda I_d + X^TX)^{-1}X^T \\
&= (\lambda I_d + V\Sigma U^TU\Sigma V^T)^{-1}X^T \quad (\because X = U\Sigma V^T) \\
&= (\lambda I_d + V\Sigma I_d\Sigma V^T)^{-1}X^T \quad (\because U^TU = I_d) \\
&= (\lambda VV^T + V\Sigma\Sigma V^T)^{-1}X^T \quad (\because I_d = V^TV) \\
&= [V(\lambda + \Sigma\Sigma)V^T]^{-1}X^T \\
&= V^{T^{-1}}(\lambda I_d + \Sigma\Sigma)^{-1}V^{-1}X^T \\
&= V(\lambda I_d + \Sigma\Sigma)^{-1}V^TX^T \quad (\because V^TV = I_d \text{ 이므로 } V^T = V^{-1})
\end{aligned}
$$

1.2) λ의 변화에 따른 $\|\hat{\beta}\|_2$ 변화

위 식을 이용하여 $\hat{\beta}$를 나타내면,

$$
\begin{aligned}
\hat{\beta} &= (\lambda I_d + X^TX)^{-1}X^Ty \\
&= V(\lambda I_d + \Sigma\Sigma)^{-1}V^TX^Ty
\end{aligned}
$$

여기서 $(\lambda I_d + \Sigma\Sigma)^{-1}$의 행렬을 풀어서 살펴보면,

$$(\lambda I_d + \Sigma\Sigma)^{-1}$$

$$= \left[\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots \\ \dots & 0 & \lambda_{\dots} & 0 \\ 0 & \dots & 0 & \lambda_d \end{bmatrix} + \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots \\ \dots & 0 & \sigma_{\dots} & 0 \\ 0 & \dots & 0 & \sigma_d \end{bmatrix}\begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots \\ \dots & 0 & \sigma_{\dots} & 0 \\ 0 & \dots & 0 & \sigma_d \end{bmatrix}\right]^{-1}$$

$$= \left[\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots \\ \dots & 0 & \lambda_{\dots} & 0 \\ 0 & \dots & 0 & \lambda_d \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots \\ \dots & 0 & \sigma_{\dots}^2 & 0 \\ 0 & \dots & 0 & \sigma_d^2 \end{bmatrix}\right]^{-1}$$

$$= \begin{bmatrix} \lambda_1 + \sigma_1^2 & 0 & \dots & 0 \\ 0 & \lambda_2 + \sigma_2^2 & 0 & \dots \\ \dots & 0 & \lambda_{\dots} + \sigma_{\dots}^2 & 0 \\ 0 & \dots & 0 & \lambda_d + \sigma_d^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{\lambda_1 + \sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2 + \sigma_2^2} & 0 & \dots \\ \dots & 0 & \frac{1}{\lambda_{\dots} + \sigma_{\dots}^2} & 0 \\ 0 & \dots & 0 & \frac{1}{\lambda_d + \sigma_d^2} \end{bmatrix}$$

이 되므로,

따라서 다른 조건이 일정하다면,

$(\lambda I_d + \Sigma\Sigma)^{-1}$은 λ가 커질수록 작아지고 $\hat{\beta}$도 작아지게 되므로 $\| \hat{\beta} \|_2$도 작아진다.

2 Kaggle competition data

2.1) coefficients of ridge and lasso regression which corresponding penalty parameter

2.1.1) Ridge (mixture=0)

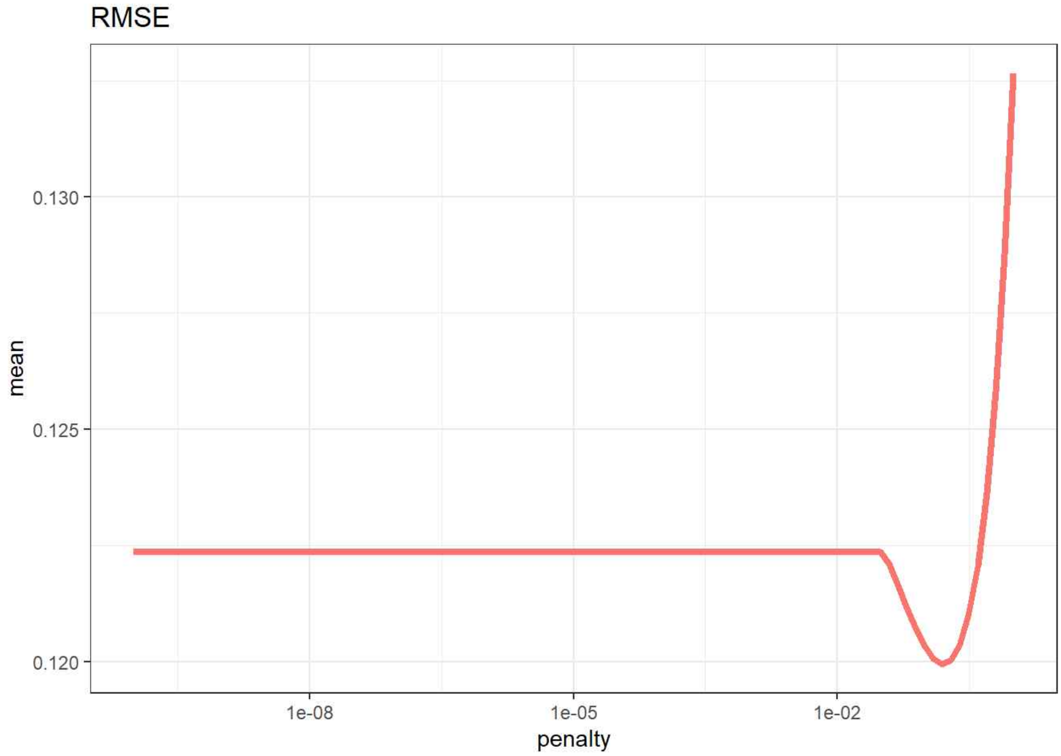
```
## # A tibble: 100 x 7
##     penalty .metric .estimator  mean     n std_err .config
##       <dbl> <chr>   <chr>      <dbl> <int>   <dbl> <fct>
##  1 1.00e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model001
##  2 1.26e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model002
##  3 1.59e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model003
##  4 2.01e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model004
##  5 2.54e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model005
##  6 3.20e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model006
##  7 4.04e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model007
##  8 5.09e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model008
##  9 6.43e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model009
## 10 8.11e-10 rmse    standard   0.122     5 0.00222 Preprocessor1_Model010
## # ... with 90 more rows
```

2.1.2) LASSO (mixture=1)

```
## # A tibble: 100 x 7
##     penalty .metric .estimator  mean     n std_err .config
##       <dbl> <chr>   <chr>      <dbl> <int>   <dbl> <fct>
##  1 1.00e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model001
##  2 1.26e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model002
##  3 1.59e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model003
##  4 2.01e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model004
##  5 2.54e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model005
##  6 3.20e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model006
##  7 4.04e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model007
##  8 5.09e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model008
##  9 6.43e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model009
## 10 8.11e-10 rmse    standard   0.125     5 0.00456 Preprocessor1_Model010
## # ... with 90 more rows
```

2.2) Plot the coefficients vs. the specified penalty values

2.2.1) Ridge (mixture=0)



2.2.2) LASSO (mixture=1)