Effects of outliers on Tikhonov regularization

Yan Zheng *

November 2020

Abstract

What is the impact of outliars to a regularized model when choosing different regularization form?

For ridge regression, the criterion in matrix form,

$$\mathrm{RSS}(\lambda) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^T\beta \tag{1}$$

And the solution is:

$$\hat{\beta}^{ridge} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} = (1 + N\lambda)^{-1}\hat{\beta}^{OLS} \tag{2}$$

where

$$\hat{\beta}^{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y \tag{3}$$

While for *LASSO*, the criterion is:

$$\arg\min_{\beta}\{\frac{1}{2}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p}x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\} \tag{4}$$

Its solution in matrix form when the covariates are orthonormal is:

$$\hat{\beta}^{lasso} = \hat{\beta}^{OLS}\max(0, 1 - \frac{N\lambda}{|\hat{\beta}^{OLS}|}) \tag{5}$$

When an outlier added into the formula, the effect will increase $\hat{\beta}^{OLS}$ the same amount in $\hat{\beta}^{ridge}$ and $\hat{\beta}^{lasso}$. Therefore we only compare $\partial\hat{\beta}^{ridge}/\partial\hat{\beta}^{OLS}$ and $\partial\hat{\beta}^{lasso}/\partial\hat{\beta}^{OLS}$ is enough.

*funded by Swiss Reinsurance

We can easily derive that:

$$
\begin{aligned}
\frac{\partial \hat{\beta}^{ridge}}{\partial \hat{\beta}^{OLS}} &= \frac{1}{1 + N\lambda} \\
\frac{\partial \hat{\beta}^{lasso}}{\partial \hat{\beta}^{OLS}} &= [0, 1]
\end{aligned}
\tag{6}
$$

Ridge regression will be less influenced by outliers compared with *Lasso*.