# Influence plots for shrinkage estimators in high-dimensional data

Jae Eun Lee1, Dae-Heung Jang1

*1Department of Statistics, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Korea*

*E-mail: dhjang@pknu.ac.kr*

***Abstract:*** When estimating regression coefficients in high-dimensional data, we need to find influence points. This is because the impact of influential observations can be large. In this paper, we suggest the influence plots for identifying how the shrinkage estimates change for each explanatory variable when we delete observations. The influence plots using shrinkage estimates have been applied to an example of high-dimensional data.

***Keywords*:**High-dimensional data, Influential observations, LASSO influence plots, shrinkage estimators

# INTRODUCTION

High-dimensional data has the number of explanatory variable(p) much more bigger than the number of sample size. Many explanatory variables can make it possible to sparse in the high-dimensional data. In this case, we consider shrinkage estimators which can be regarded as a variable selection method. However, influential observations can greatly impact results. Therefore we need to consider detecting influential observations. This paper focuses on LASSO influence plots for finding influential observations. We also regards another shrinkage estimators such as SCAD, MCP and so on.

# METHODOLOGY

Jang and Anderson-Coook (2017, 2018) suggested LASSO influence plots that is useful to understand the impact of individual observations on the selected explanatory variables when using LASSO. We explain the influence plot through a dataset taken from R packages library(SIS). The dataset, ‘prostate.train’, consists of 102 observations with 12,600 explanatory variables. Figure 1 shows not only LASSO influence plot but also SCAD and MCP influence plots after SIS (Sure Independence Screening) for dataset ‘prostate.train’. In this plot, we see estimates change for the selected explanatory variables according to 42th observation deletion.

|  |  |  |
| --- | --- | --- |
| **C:\Users\jaeun\AppData\Local\Microsoft\Windows\INetCache\Content.Word\SIS influence plot_deletion_42.png** | **K:\1. 2017 학회논문발표\국외학회\2018\1. WMSRAI 2018\SIS_influence plot_prostate.train_SCAD\SIS-prostate.train-SCAD-influence plot_42.PNG** |  |

Figure 1 Influence plots (LASSO, SCAD, MCP) for prostate data

# CONCLUSIONS

Using influence plots, we are able to know that influence observation can make it possible to impact on shrinkage estimation.

# REFERENCES

Fan, J., Feng, Y., Saldana, D. F., Samworth, R. and Wu, Y. (2017). [http://www.stat.columbia.edu/ ~yangfeng/pubs/jss1375.pdf](http://www.stat.columbia.edu/%20~yangfeng/pubs/jss1375.pdf), Package ‘SIS’.

Jang, D-H, Anderson-Cook, C. M. (2017). Influence plots for LASSO. *Quality and Reliability Engineering International*, **33**, 1317-1326.

Jang, D-H, Anderson-Cook, C. M. (2018). Assessing robustness of factor ranking for supersaturated designs, *Quality and Reliability Engineering International* **, 34**, 417-426.