

# Interpretable Machine Learning for the Social Sciences: Revisiting Penalized Regression in Large-Scale Analysis

Jin Eun Yoo

Korea National University of Education  
Department of Education  
250 Taeseongtabyeon-ro, Gangnae-myeon,  
Heungdeok-gu, Cheongju-si, Chungbuk 28173  
South Korea  
jeyoo@knue.ac.kr

**Abstract.** Penalized regression in large-scale data analysis is a promising area of continued growth. Compared to nonlinear ML methods such as deep learning, which produces prediction models notoriously difficult to interpret, penalized regression as a linear method yields interpretable prediction models; this brings great merit in social science research which particularly has valued explanation. Applying penalized regression to large-scale data, we can also find new important predictors which have been neglected in the literature. This can help researchers break free from the confines of traditional approaches and discover novel insights. In the similar context, it is notable that penalized regression rests on the sparsity assumption, and may be less suitable for small-scale studies. Starting with LASSO (Least Absolute Shrinkage and Selection Operator) for variable selection, variations of penalized regression have been developed including elastic net for multicollinearity issues and Mnet for consistent coefficient estimates. Recently, the scope of penalized regression has been expanded to significance testing and multilevel models, making it a versatile and powerful tool for a wide range of data analysis tasks. These developments have enriched the landscape of penalized regression and its capability in different research domains.

**Keywords.** penalized regression, large-scale data, interpretable predictive models

## References

- Agresti, A. (2002) *Categorical data analysis* (2nd ed.). John Wiley & Sons.  
<http://dx.doi.org/10.1002/0471249688>
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the 2nd International Symposium on Information Theory* (pp. 267-281). Akademiai Kiado.
- Bačák, V., & Kennedy, E. H. (2019). Principled machine learning using the super learner: An application to predicting prison violence. *Sociological Methods & Research*, 48(3), 698–721.  
<https://doi.org/10.1177/0049124117747301>
- Breheny, P., Zeng, Y., & Kurth, R. (2022). *Regularization paths for regression models with grouped covariates*. Retrieved from <https://cloud.rproject.org/web/packages/grpreg/grpreg.pdf>
- Bühlmann, P., & Mandozzi, J. (2014). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Computational Statistics*.  
<https://doi.org/10.1007/s00180-013-0436-3>
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348–1360.  
<https://doi.org/10.1198/016214501753382273>
- Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 849–911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- Goeman, J., Meijer, R., Chaturvedi, N. (2022). *L1 and L2 penalized regression models*. Retrieved from <https://cran.r-project.org/web/packages/penalized/vignettes/penalized.pdf>
- Groll, A., & Tutz, G. (2014). Variable selection for generalized linear mixed models by L1-penalized estimation. *Statistics and Computing*, 24(2), 137–154.  
<https://doi.org/10.1007/s11222-012-9359-z>

- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hastie, T., Qian, J., & Tay, K. (2023). *An introduction to glmnet*. Retrieved from <https://cloud.r-project.org/web/packages/glmnet/vignettes/glmnet.pdf>
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Huang, J., Breheny, P. Lee, S. Ma, S., & Zhang, C. H. (2016). The Mnet method for variable selection. *Statistica Sinica*, 26(3), 903–923. <http://dx.doi.org/10.5705/ss.202014.0011>
- Huang, J., Ma, S., & Zhang, C. H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, 18(4), 1603–1618.
- Immekus, J. C., Jeong, T., & Yoo, J. E. (2022). Machine learning procedures for predictor variable selection for schoolwork-related anxiety: evidence from PISA 2015 mathematics, reading, and science assessments. *Large-scale Assessments in Education*, 10(30). <https://doi.org/10.1186/s40536-022-00150-8>
- Javanmard, A., & Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1), 2869–2909. <https://www.jmlr.org/papers/volume15/javanmard14a.pdf>
- Kim, H. G., & Yoo, J. E. (2020). ICILS 2018 variable exploration to predict computer and information literacy: Variable selection in multilevel modeling via glmmLasso. *Journal of Education Science*, 22(4), 1–21. <https://doi.org/10.15564/jeju.2020.11.22.4.1>
- Koo, M., & Yoo, J. E. (2021) Intraclass correlation and the performance of penalized regression: exploration of predictors for TALIS 2018 teacher cooperation. *Asian Journal of Education*, 22(1), 31–59.
- Koo, M., & Yoo, J. E. (2025). Teachers' team innovativeness in TALIS 2018: An empirical and simulation study using glmmLasso for multilevel data. *Large-scale Assessments in Education*, 13, 19.
- Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research*, 33(2), 188–229.
- Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3), 907–927. <https://doi.org/10.1214/15-AOS1371>
- Lim, H. J., Yoo, J. E., Rho, M., & Ryu, J. J. (2022). Exploration of variables predicting sense of school belonging using the machine learning method—group Mnet. *Psychological Reports*. <https://doi.org/10.1177/00332941221133005>
- Meier, L. (2022). Package ‘grplasso’ (version 0.4-7). Retrieved from <https://cran.r-project.org/web/packages/grplasso/grplasso.pdf>
- Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, 34(3), 1436–1462.
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), 417–473.
- Rho, M., & Yoo, J. E. (2019). Exploration of variables relating to career decisions via adaptive LASSO. *The Journal of Yeolin Education*, 27(4), 133–155.
- Rho, M., & Yoo, J. E. (2021). Statistical inference after variable selection in penalized regression: Focusing on variables relating to adolescents' smartphone reliance. *Studies on Korean Youth*, 32(1), 147–174.
- Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461–464. <http://dx.doi.org/10.1214/aos/1176344136>
- Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, 7(2), 221–242.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310.
- Taylor, J., & Tibshirani, R. (2018). Post-selection inference for L1-penalized likelihood models. *The Canadian Journal of Statistics*, 46(1), 41–61. <https://doi.org/10.1002/cjs.11313>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58 (1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- van Zoonen, W., & van der Meer, T. G. L. A. (2016). Social media research: The application of supervised machine learning in organizational communication research. *Computers in Human Behavior*, 63, 132–141. <https://doi.org/10.1016/j.chb.2016.05.028>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*,

12(6), 1100–1122.

<https://doi.org/10.1177/1745691617693393>

<https://doi.org/10.1198/016214506000000735>

- Yoo, J. E. (2018). TIMSS 2011 Student and teacher predictors for mathematics achievement explored and identified via elastic net. *Frontiers in Psychology*, 9, 317. <https://doi.org/10.3389/fpsyg.2018.00317>
- Yoo, J. E. (2021). *AI, big data, and machine learning*. Hakjisa.
- Yoo, J. E., & Rho, M. (2020). Exploration of predictors for Korean teacher job satisfaction via a machine learning technique, group Mnet. *Frontiers in Psychology*, 11, 441. doi: 10.3389/fpsyg.2020.00441
- Yoo, J. E., & Rho, M. (2021, April 8–12). *Statistical inference after variable selection via penalized regression: Focusing on variables predicting belonging to school*. Paper presented at 2021 American Educational Research Association (AERA) Annual Meeting. Orlando, FL.
- Yoo, J. E., & Rho, M. (2022). Large-scale survey data analysis with penalized regression: A Monte Carlo simulation on missing categorical predictors. *Multivariate Behavioral Research*. <https://doi.org/10.1080/00273171.2021.1891856>
- Yoo, J. E., & Rho, M. (2023, May 4). *Penalized regression versus forward stepwise regression in variable selection and significance testing: Evidence from PISA 2015*. Paper presented at 2023 American Educational Research Association (AERA) Annual Meeting. Online.
- Yoo, J. E., Rho, M., & Lee, Y. (2022). Online students' learning behaviors and academic success: An analysis of LMS log data from flipped classrooms via regularization. *IEEE Access*, 10, 10740–10753. <https://doi.org/10.1109/ACCESS.2022.3144625>
- Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894–942. <https://doi.org/10.1214/09-AOS729>
- Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7, 2541–2563.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.

**Informaiton about prof.dr  
Jin Eun Yoo, Ph.D.**

**Professor (Measurement, Evaluation, and  
Research Methodology), Social Science  
Statistics**

**Frontiers in Psychology (SSCI) / Associat  
e Editor  
Innovation and Education / Associate Edit  
or**

**Measurement, Evaluation, and Statistics /  
Professor  
Department of Education  
Korea National University of Education**