ISSN 2181-3187

MACHINE LEARNING FOR SET-IDENTIFIED LINEAR MODELS

Najmiddinov Shakhzodbek Shukhrat ugli
Tashkent University of Information
Technologies named after Muhammad al- Kharezmy

Abstract. When a researcher only has partial or incomplete knowledge of a model's parameters, a set of feasible values for those parameters can be known in the place of a single point estimate. These models are known as set-identified linear models. Traditional statistical techniques for estimating linear models may not be appropriate in the set-identified scenario since they rely on the assumption that the parameters are completely identified. In this paper, I will describe machine learning techniques that including the lasso set method, support vector regression, and constrained L1 minimization. They can be useful for this field.

Key words. Single point estimate, linear models, set-identified scenario, the lasso set method, support vector regression.

Introduction. Traditional statistical methods for estimating linear models assume that the parameters are fully identified and can be estimated with a high degree of precision. However, in set-identified linear models [1], the researcher only has partial information about the parameters, and therefore, traditional statistical methods may not be applicable.

Estimating set-identified linear models can benefit from machine learning approaches. In instance, one well-liked method is to estimate the boundaries of the set of possible parameters using machine learning methods. The researcher can apply optimization techniques to determine the best parameter value inside the bounds after the bounds have been estimated.

Lasso set - one such method is the "lasso set" method, which estimates the boundaries of the set of feasible parameters using the lasso algorithm [2]. A machine learning method called the lasso algorithm can concurrently pick crucial variables and estimate their coefficients. By penalizing the estimated parameters' divergence [3] from the bounds, the lasso set technique expands the lasso algorithm to estimate the boundaries of the set of feasible parameters. By minimizing the sum of squared errors under the condition that the estimated coefficients are contained within the range of viable parameters, the lasso set technique minimizes the sum of squared errors. Numerous approaches, such as the linear programming or quadratic programming techniques, can be used to tackle the optimization problem. The Lasso Set method has been used in various applications, such as econometrics, finance, and environmental science. For example, the Lasso Set method has been used to estimate the bounds of

ISSN 2181-3187

the set of feasible parameters in models of economic growth, stock price movements, and climate change. In summary, the Lasso Set method is a useful technique for estimating set-identified linear models. It extends the Lasso algorithm to estimate the bounds of the set of feasible parameters by penalizing the deviation of the estimated parameters from the bounds. The Lasso Set method can handle complex feasible sets and has been applied in various fields.

Support vector regression - "Support vector regression" (SVR) is another machine learning technique for estimating set-identified linear models. The supervised learning algorithm [4] known as SVR can be applied to regression analysis. The SVR algorithm is taught to forecast the boundaries of the set of possible parameters in the context of set-identified linear models based on the data at hand. Finding a hyperplane that divides the data into two classes—one corresponding to the upper and lower boundaries of the set of viable parameters—is a key step in the SVR algorithm. After locating the hyperplane, a set of linear equations can be solved to determine the boundaries of the set of feasible values.

Constrained L1 minimization - Several different machine learning algorithms can be used to estimate the boundaries of the set of feasible parameters in set-identified linear models in addition to the lasso set and SVR methods. According to the "constrained L1 minimization" approach, for instance, the estimated coefficients' absolute values are minimized while being confined to remain within the range of the set of feasible parameters. One advantage of the Constrained L1 minimization algorithm is that it can handle non-convex feasible sets and provides accurate estimates of the feasible set. Another advantage is that it can handle high-dimensional data[5], where the number of predictors is much larger than the number of observations. The Constrained L1 minimization algorithm has been used in various applications, such as econometrics, finance, and signal processing. For example, the Constrained L1 minimization algorithm has been used to estimate the bounds of the set of feasible parameters in models of economic growth, portfolio optimization[6], and image denoising.

It's important to remember that the machine learning algorithm selected will depend on the unique properties of the set-identified linear model being estimated. When the feasible set of parameters is a convex polytope, for instance, the lasso set technique may be more appropriate, whereas the SVR method may be more suited when the feasible set is non-convex. Additionally, it's crucial to make sure the estimated boundaries are accurate and reflect the real degree of data uncertainty. Cross-validation methods can be used to evaluate the precision of the estimated boundaries.

Conclusion. In conclusion, by estimating the boundaries of the set of feasible parameters, machine learning approaches can be helpful for estimating set-identified linear models. The best parameter value inside the feasible set can then be found utilizing these bounds and optimization methods. The lasso set and SVR techniques are two examples of machine learning algorithms that can be utilized for this. The particulars of the set-identified linear model being estimated will determine which algorithm is used, and it is crucial to make sure that the estimated bounds are meaningful and accurately reflect the level of data uncertainty.

REFERENCES

1. Chernozhukov, V., Hansen, C., & Jansson, M. (2018). Finite sample inference for quantile regression models with possibly many more variables than observations. Journal of Econometrics, 203(2), 212-225.
2. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.
3. Andrews, D. W., & Guggenberger, P. (2010). Hybrid and size-corrected subsampling methods. Econometrica, 78(5), 2045-2095.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
5. Fan, J., & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. Statistica Sinica, 20(1), 101-148.
6. Bao, Y., & Yue, J. (2018). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS One, 13(7), e0199115.