Valid prediction intervals for regression problems

Nicolas Dewolf, Bernard De Baets, and Willem Waegeman

Ghent University, Belgium

Abstract. Over the last few decades, various methods have been proposed for estimating prediction intervals in regression settings, including Bayesian methods, ensemble methods, direct interval estimation methods and conformal prediction methods. An important issue is the validity and calibration of these methods: the generated prediction intervals should have a predefined coverage level, without being overly conservative. So far, no study has analysed this issue whilst simultaneously considering these four classes of methods. In this independent comparative study, we reviewed the above four classes of methods from a conceptual and experimental point of view in the i.i.d. setting. Results on benchmark data sets from various domains highlight large fluctuations in performance from one data set to another. These observations can be attributed to the violation of certain assumptions that are inherent to some classes of methods. We also illustrated how conformal prediction can be used as a general calibration procedure for methods that deliver poor results without a calibration step. The source code for the experiments can be found on GitHub: `https://github.com/nmdwolf/ValidPredictionIntervals`.

Keywords: Prediction interval - Regression - Calibration - Conformal prediction

1 Introduction

Machine learning methods, and in particular deep learning methods, often serve as work horses in artificial intelligence systems that have a strong impact on the daily life of humans, such as self-driving cars [2,11], machine translation [12,14] and medical diagnostics [3,5]. However, such systems are only accepted by humans if they exhibit a sufficient degree of reliability. As a result, analyzing what systems “know” and what they “don’t know” has become an important topic of recent deep learning research, using “uncertainty quantification” and “uncertainty estimation” as prominent buzz words [9,10,15,16].

Driven by popular application domains like computer vision and natural language processing, most of the recent literature heavily focuses on classification problems. Regression problems are often ignored, or at least less analyzed in such papers. Methods for uncertainty quantification in classification and regression problems usually differ substantially. Many traditional classification methods produce probability estimates, whereas in regressio, most traditional methods are so-called point predictors; they only predict one summary statistic of the conditional distribution. However, a point predictor cannot express how confident it is of a prediction, and typically a prediction interval is returned by more complicated methods to quantify uncertainty, i.e. the wider the interval, the larger the uncertainty. Such prediction intervals can be obtained by modelling the conditional distribution in an exact or approximate manner, using for example Bayesian methods [4,18] or ensemble methods [6,19]. Prediction intervals can also be estimated in a more direct manner, without modelling the conditional distribution entirely, using for example quantile regression [8] or conformal prediction methods [13,17].

The estimation of prediction intervals for regression has received little attention recently, and the last general review predates the ongoing deep learning wave [7] (at the time of writing another review appeared with a strong focus on fuzzy methods [1]). As a result, an up-to-date comparison of methods that generate prediction intervals was necessary. This paper bridged that gap, by focusing on four aspects:

i) Give an overview of the four general classes of methods that produce prediction intervals: Bayesian methods, ensemble methods, direct estimation methods and conformal prediction.

ii) Elaborate on the calibration (or validity) of prediction intervals and relate it to data and model properties.

iii) Show how conformal prediction can be applied as a general framework to obtain well-calibrated prediction intervals.

iv) Provide an in-depth experimental comparison of the four classes of methods, based on their performance across a wide range of data sets.

2 Paper information

Citation: Dewolf, N., Baets, B.D. & Waegeman, W. Valid prediction intervals for regression problems. Artif Intell Rev (2022). https://doi.org/10.1007/s10462-022-10178-5

URL: `https://link.springer.com/article/10.1007/s10462-022-10178-5`

Publication date: 18 April 2022

References

1. Cartagena, O., Parra, S., Muñoz-Carpintero, D., Marín, L.G., Sáez, D.: Review on fuzzy and neural prediction interval modelling for nonlinear dynamical systems. IEEE Access **9**, 23357–23384 (2021). https://doi.org/10.1109/ACCESS.2021.3056003
2. Chen, Z., Huang, X.: End-to-end learning for lane keeping of self-driving cars. In: 2017 IEEE Intelligent Vehicles Symposium (IV). pp. 1856–1860. IEEE (2017)

3. Fang, Y., Xu, P., Yang, J., Qin, Y.: A quantile regression forest based method to predict drug response and assess prediction reliability. PLOS ONE **13**(10) (2018)
4. Goan, E., Fookes, C.: Bayesian neural networks: An introduction and survey. In: Case Studies in Applied Bayesian Data Science, pp. 45–87. Springer (2020)
5. Jiang, X., Osl, M., Kim, J., Ohno-Machado, L.: Calibrating predictive model estimates to support personalized medicine. Journal of the American Medical Informatics Association **19**(2), 263–274 (2012)
6. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30, pp. 5580—-5590. Curran Associates, Inc. (2017)
7. Khosravi, A., Nahavandi, S., Creighton, D., Atiya, A.F.: Comprehensive review of neural network-based prediction intervals and new advances. IEEE Transactions on Neural Networks **22**(9), 1341–1356 (2011). `https://doi.org/10.1109/TNN.2011.2162110`
8. Koenker, R., Hallock, K.F.: Quantile regression. Journal of Economic Perspectives **15**(4), 143–156 (December 2001). `https://doi.org/10.1257/jep.15.4.143`
9. Liu, J., Paisley, J., Kioumourtzoglou, M.A., Coull, B.: Accurate uncertainty estimation and decomposition in ensemble learning. In: Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019)
10. Malinin, A., Gales, M.: Predictive uncertainty estimation via prior networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. pp. 7047—-7058. NIPS'18, Curran Associates Inc. (2018)
11. Michelmore, R., Wicker, M., Laurenti, L., Cardelli, L., Gal, Y., Kwiatkowska, M.: Uncertainty quantification with statistical guarantees in end-to-end autonomous driving control. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). pp. 7344–7350 (2020). `https://doi.org/10.1109/ICRA40945.2020.9196844`
12. Ott, M., Auli, M., Grangier, D., Ranzato, M.: Analyzing uncertainty in neural machine translation. In: International Conference on Machine Learning (ICML). pp. 3956–3965. PMLR (2018)
13. Romano, Y., Patterson, E., Candes, E.: Conformalized quantile regression. In: Advances in Neural Information Processing Systems. pp. 3543–3553 (2019)
14. Singh, S.P., Kumar, A., Darbari, H., Singh, L., Rastogi, A., Jain, S.: Machine translation using deep learning: An overview. In: 2017 International Conference on Computer, Communications and Electronics (Comptelix). pp. 162–167. IEEE (2017)
15. Teye, M., Azizpour, H., Smith, K.: Bayesian uncertainty estimation for batch normalized deep networks. In: International Conference on Machine Learning. pp. 4907–4916. PMLR (2018)
16. Van Amersfoort, J., Smith, L., Teh, Y.W., Gal, Y.: Uncertainty estimation using a single deep deterministic neural network. In: International Conference on Machine Learning. pp. 9690–9700. PMLR (2020)
17. Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World. Springer Science & Business Media (2005)
18. Williams, C., Rasmussen, C.: Gaussian processes for regression. In: Advances in Neural Information Processing Systems. pp. 514–520 (1996)
19. Zhang, H., Zimmerman, J., Nettleton, D., Nordman, D.J.: Random forest prediction intervals. The American Statistician **74**(4), 392–406 (2020)