Biecek, Przemysław

DALEX: explainers for complex predictive models in R. (English) Zbl 07008340

J. Mach. Learn. Res. 19, Paper No. 84, 5 p. (2018)

Summary: Predictive modeling is invaded by elastic, yet complex methods such as neural networks or ensembles (model stacking, boosting or bagging). Such methods are usually described by a large number of parameters or hyper parameters – a price that one needs to pay for elasticity. The very number of parameters makes models hard to understand.

This paper describes a consistent collection of explainers for predictive models, a.k.a. black boxes. Each explainer is a technique for exploration of a black box model. Presented approaches are model-agnostic, what means that they extract useful information from any predictive method irrespective of its internal structure. Each explainer is linked with a specific aspect of a model. Some are useful in decomposing predictions, some serve better in understanding performance, while others are useful in understanding importance and conditional responses of a particular variable.

Every explainer presented here works for a single model or for a collection of models. In the latter case, models can be compared against each other. Such comparison helps to find strengths and weaknesses of different models and gives additional tools for model validation. Presented explainers are implemented in the `DALEX` package for `R`. They are based on a uniform standardized grammar of model exploration which may be easily extended.

MSC:

68T05 Learning and adaptive systems in artificial intelligence

Cited in **1** Document

Keywords:

interpretable machine learning; explainable artificial intelligence; predictive modelling; model visualization

Software:

R; ALEPlot; pdp; breakDown; shap; caret; live; archivist; auditor; DALEX

Full Text: Link

References:

[1] Dan Apley. ALEPlot: Accumulated Local Effects (ALE) Plots and Partial Dependence (PD) Plots, 2017. R package version 1.0.

[2] Przemyslaw Biecek and Marcin Kosinski. archivist: An R Package for Managing, Recording and Restoring Data Analysis Results. Journal of Statistical Software, 82(11):1–28, 2017. doi: 10.18637/jss.v082.i11.

[3] Bernd Bischl, Michel Lang, Lars Kotthoff, Julia Schiffner, Jakob Richter, Erich Studerus, Giuseppe Casalicchio, and Zachary M. Jones. mlr: Machine Learning in R. Journal of Machine Learning Research, 17(170):1–5, 2016. · Zbl 1392.68007

[4] Leo Breiman. Random forests. Machine Learning, 45(1):5–32, 2001. · Zbl 1007.68152

[5] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class, from the "Rashomon" Perspective, 2018. URL http://arxiv.org/abs/1801.01489.

[6] Alicja Gosiewska and Przemyslaw Biecek. auditor: an R Package for Model-Agnostic Visual Validation and Diagnostic, 2018. URL https://arxiv.org/abs/1809.07763.

[7] Brandon Greenwell. pdp: An R package for Constructing Partial Dependence Plots. The R Journal, 9(1):421–436, 2017.

[8] Jos Hernndez-Orallo. ROC curves for regression. Pattern Recognition, 46(12):33953411, 2013. doi: 10.1016/j.patcog.2013.06.014.

[9] Ulf Johansson, Cecilia Snstrd, Ulf Norinder, and Henrik Bostrm. Trade-off between accuracy and interpretability for predictive in silico modeling. Future Medicinal Chemistry, 3(6): 647–663, 2011. doi: 10.4155/fmc.11.23.

[10] Max Kuhn. Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 2008. doi: 10.18637/jss.v028.i05.

[11] Scott Lundberg and Su-In Lee.A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30, pages 4765–4774, 2017.

[12] Cathy O'Neil.Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, New York, NY, USA, 2016. · Zbl 1441.00001

[13] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[14] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?". ACM Press, 2016. doi: 10.1145/2939672.2939778. URL https://arxiv.org/abs/1602.04938.

[15] David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. In Proceedings of the 28th International Conference on Neural Information Processing Systems. MIT Press, 2015.

[16] Agnieszka Sitko and Przemyslaw Biecek. The Merging Path Plot: adaptive fusing of k-groups with likelihood-based model selection, 2017. URL https://arxiv.org/abs/1709.04412.

[17] Mateusz Staniak and Przemyslaw Biecek. Explanations of Model Predictions with live and breakDown Packages, 2018. URL https://arxiv.org/abs/1804.01955.