Research Working Papers

Explaining Machine Learning by Bootstrapping Partial Marginal Effects and Shapley Values

by: Thomas R. Cook, Zach Modig and Nathan M. Palmer

November 15, 2021

Partial marginal effects and Shapley values can help with interpreting machine learning models. We use these tools to interpret hedonic preferences for real estate.

RWP 21-12, November 2021; updated August 2024

Machine learning and artificial intelligence are often described as "black boxes." Traditional linear regression is interpreted through its marginal relationships as captured by regression coefficients. We show that the same marginal relationship can be described rigorously for any machine learning model by calculating the slope of the partial dependence functions, which we call the partial marginal effect (PME). We prove that the PME of OLS is analytically equivalent to the OLS regression coefficient. Boot- strapping provides standard errors and confidence intervals around the point estimates of the PMEs. We apply the PME to a hedonic house pricing example and demonstrate that the PMEs of neural networks, support vector machines, random forests, and gradient boosting models reveal the non-linear relationships discovered by the machine learning models and allow direct comparison between those models and a traditional linear regression. Finally we extend PME to a Shapley value decomposition and explore how it can be used to further explain model outputs.

JEL Classifications: C14, C18, C15, C45, C52

View the original 2021 paper here.

Article Citations

- Cook, Thomas R., Zach Modig, and Nathan M. Palmer. 2024. "Explaining Machine Learning by Bootstrapping Partial Marginal Effects and Shapley Values." Federal Reserve Bank of Kansas City, Research Working Paper no. 21-12, August. Available at https://doi.org/10.18651/RWP2021-12

Author

Thomas R. Cook

Data Scientist

Tom Cook is a Data Scientist in the Economic Research Department of the Federal Reserve Bank of Kansas City. He joined the bank in August 2016 after completing his PhD in Political Science at the University of Colorado. He also holds an MPA from DePaul University, and a BA in Philosophy and Political Science from the University of Iowa. His substantive research interests are the roles of time and information transmission in political and economic strategic behavior. Methodologically, his research at the bank focuses on the development of machine learning, neural networks, and advanced statistical models for use in economic research.