

Using Adaptive Basis Search Method To Interpret Black-Box Models

Ambrose Emmett-Iwaniw

University of Waterloo

arsemmet@uwaterloo.ca

Coauthor(s): Christiane Lemieux

Inference is an important part of machine learning and statistics. In the case of Black-Box models, Knowing which variables are important for prediction is an ongoing problem. The Quasi-Regression (QR) method is used to approximate a function of interest by a linear combination of orthonormal basis functions of $L^2[0, 1]^d$. The coefficients are integrals that do not have an analytical solution. Thus, a need arises for Monte Carlo methods to approximate them. QR is an approximate inference method that uses functional ANOVA and global sensitivity indices to explain which variables and interactions are important for the fit. The QR method performs poorly due to its high variance leading to slow convergence. The QR method was extended by using shrinkage parameters that act like a control variate by helping lower the variance. This was shown to perform better than the original QR method. Here we extend QR by using the RQMC method. Our results show, theoretically and empirically, that QR with Randomized quasi-Monte Carlo (RQMC) is superior to using the original QR method. In practice, the QR method can be time-consuming if the number of basis functions is large. If the function of interest is sparse many of these basis functions are irrelevant and can be removed. We address this problem by creating new adaptive basis search methods based on the RQMC method that adaptively selects important basis functions. These methods are shown to be much faster than the previous QR methods. Also, with properly selected cut-off parameters, the adaptive search methods are shown theoretically and empirically to perform better than QR with RQMC. Also, it is empirically shown that with properly selected cut-off parameters the adaptive basis search methods perform better than QR with shrinkage.