Titel:
Conditional Feature Importance for Mixed Data

Abstract:
Despite the popularity of feature importance (FI) measures in interpretable machine learning, the statistical adequacy of these methods is rarely discussed. From a statistical perspective, a major distinction is between analyzing a variable's importance before and after adjusting for covariates - i.e., between marginal and conditional measures. Our work draws attention to this rarely acknowledged, yet crucial distinction and showcases its implications. Further, we reveal that for testing conditional FI, only few methods are available and practitioners have hitherto been severely restricted in method application due to mismatching data requirements. Most real-world data exhibits complex feature dependencies and incorporates both continuous and categorical data (mixed data). Both properties are oftentimes neglected by conditional FI measures. To fill this gap, we propose to combine the conditional predictive impact (CPI) framework with sequential knockoff sampling. The CPI enables conditional FI measurement that controls for any feature dependencies by sampling valid knockoffs - hence, generating synthetic data with similar statistical properties - for the data to be analyzed. Sequential knockoffs were deliberately designed to handle mixed data and thus allow us to extend the CPI approach to such datasets. We demonstrate through numerous simulations and a real-world example that our proposed workflow controls type I error, achieves high power and is in line with results given by other conditional FI measures, whereas marginal FI metrics result in misleading interpretations. Our findings highlight the necessity of developing statistically adequate, specialized methods for mixed data.