

Burim Ramosaj
TU Dortmund
Faculty of Statistics

Interpretable Machines – Constructing valid Prediction Intervals with Random Forest

An important issue when using Machine Learning algorithms in recent research is the lack of interpretability. Although several algorithms provide accurate point predictions for various learning problems, uncertainty estimates connected with point predictions are rather sparse. A contribution to this gap for the Random Forest Regression Learner is presented here. Based on its Out-of-Bag procedure, several parametric and non-parametric prediction intervals are provided for Random Forest point predictions and theoretical guarantees for its correct coverage probability is delivered. Furthermore, a thorough investigation through Monte-Carlo simulation is conducted evaluating the performance of the proposed methods from three aspects: (i) Analyzing the correct coverage rate of the proposed prediction intervals, (ii) Inspecting interval width and (iii) Verifying the competitiveness of the proposed intervals with existing methods. The simulation yields that the proposed prediction intervals are robust towards non-normal residual distributions and are competitive by providing correct coverage rates and comparably narrow interval lengths, even for comparably small samples.