



Gradient boosting for extreme quantile regression

Jasper Velthoen¹; Clément Dombry²; Juan-Juan Cai³; Sebastian Engelke⁴

- 1 Department of Applied Mathematics, Delft University of Technology
- 2 Laboratoire de Mathématiques de Besançon, Univ. Bourgogne Franche-Comté,
- 3 Department of Econometrics and Data Science, Vrije Universiteit Amsterdam
- 4 Research Center for Statistics, University of Geneva

Abstract:

Extreme quantile regression provides estimates of conditional quantiles outside the range of the data. Classical methods such as quantile random forests perform poorly in such cases since data in the tail region are too scarce. Extreme value theory motivates to approximate the conditional distribution above a high threshold by a generalized Pareto distribution with covariate dependent parameters. This model allows for extrapolation beyond the range of observed values and estimation of conditional extreme quantiles. We propose a gradient boosting procedure to estimate a conditional generalized Pareto distribution by minimizing its deviance. Cross-validation is used for the choice of tuning parameters such as the number of trees and the tree depths. We discuss diagnostic plots such as variable importance and partial dependence plots, which help to interpret the fitted models. In simulation studies we show that our gradient boosting procedure outperforms classical methods from quantile regression and extreme value theory, especially for high-dimensional predictor spaces and complex parameter response surfaces. An application to statistical post-processing of weather forecasts with precipitation data in the Netherlands is proposed.

Keywords:

extreme quantile regression; gradient boosting; generalized Pareto distribution; extreme value theory; tree-based methods.