Predicting Regression Probability Distributions with Imperfect Data Using Optimal Transformations

Jerome H. Friedman

Stanford University

4:30 pm, 6/20/2018, 1195 Bordeaux Dr, Sunnyvale, CA 94089

ABSTRACT

The goal of regression analysis is to predict the value of a numeric outcome variable y given a set of joint values of other (predictor) variables x. Usually a particular set of x-values does not specify a repeatable value for y, but rather a probability distribution of possible y-values, p(y|x). This distribution has a location, scale and shape, all of which can depend on x, and are needed to infer likely values for y given x. Regression methods usually assume that training data y-values are numeric realizations from some p(y|x) measured with infinite precision. Often actual training data y-values are discrete, truncated and/or arbitrary censored versions of an underlying numeric y-value. Regression procedures based on an optimal transformation strategy are presented for estimating location, scale and shape of p(y|x) as general functions of x, in the possible presence of such imperfect training data.

Bio: Dr. Friedman is a Professor in Statistics at Stanford University and is a member of the US National Academy of Science. He is a co-author of the popular text book entitled "The Elements of Statistics Learning: Data Mining, Inference, and Prediction". Dr. Friedman is one of the world's leading researchers in statistics and data mining. He has published on a wide range of data mining topics including trees, nearest neighbor classification, logistical regressions, boosting, and high-dimensional data analysis.