



Nonparametric Convex Imputation and Application

Jian-Hui Ning

Department of Mathematics and Statistics, Central China Normal University, CHINA –
jhning@mail.ccnu.edu.cn

Michelle Liou

Institute of Statistical Science, Academia Sinica – mliou@stat.sinica.edu.tw

Philip E. Cheng*

Institute of Statistical Science, Academia Sinica – pcheng@stat.sinica.edu.tw

Methods of regression function imputation with incomplete data have exhibited various applications. Nearest neighbor regression and kernel regression have been used for imputing missing data in survey sampling and missing treatment assignments since 1970s. Asymptotic normality for estimating the mean of missing data was given for kernel imputation in 1986, and for k -nearest neighbor imputation in 2012. In this study, a novel convex mixture of these two regression imputation methods is constructed for extracting the advantage of both methods so as to offset irregular distribution conditions in the data. A naive mixture and its modified version are found to yield improved and stable performance for estimating the mean of an incomplete variable and for predicting discrete types with or without supervised learning under general distribution conditions. A simulation study of general missing data and an empirical study of two data sets in the UCI Machine Learning Repository are examined for useful applications.

Keywords: classification, convex mixture, kernel imputation, k -nearest neighbor imputation.