

Inference for the Dimension of a Regression Relationship

Using Pseudo-covariates

Shih-Hao Huang

Department of Mathematics, National Central University

Abstract

In data analysis using dimension reduction methods, the main goal is to summarize how the response is related to the covariates through a few linear combinations. One key issue is to determine the minimal number of such linear combinations, which is the dimension of the sufficient dimension reduction (SDR) subspace. In this work, we propose an easily-applied approach to conduct inference for the dimension of the SDR subspace, based on augmentation of the covariate set with pseudo-covariates. Applying the partitioning principal to the possible dimensions, we use sequential testing to select the dimensionality, by comparing the strength of the signal arising from the actual covariates to that appearing to arise from the pseudo-covariates. Under a *uniform direction* condition, our test statistic asymptotically follows a beta distribution, and the family-wise type I error rate of our sequential testing is rigorously controlled. Simulation studies and an analysis of newborn anthropometric data demonstrate the robustness of the proposed approach, and indicate that the power is comparable to or greater than the alternatives. (Work done jointly with Kerby Shedden and Hsin-wen Chang).

Keywords: inference for dimension, nonparametric regression, sequential testing, sufficient dimension reduction, variable augmentation