Learning sparse gradients for variable selection and dimension reduction

Gui-Bo Ye · Xiaohui Xie

Abstract Variable selection and dimension reduction are two commonly adopted approaches for high-dimensional data analysis, but have traditionally been treated separately. Here we propose an integrated approach, called sparse gradient learning (SGL), for variable selection and dimension reduction via learning the gradients of the prediction function directly from samples. By imposing a sparsity constraint on the gradients, variable selection is achieved by selecting variables corresponding to non-zero partial derivatives, and effective dimensions are extracted based on the eigenvectors of the derived sparse empirical gradient covariance matrix. An error analysis is given for the convergence of the estimated gradients to the true ones in both the Euclidean and the manifold setting. We also develop an efficient forward-backward splitting algorithm to solve the SGL problem, making the framework practically scalable for medium or large datasets. The utility of SGL for variable selection and feature extraction is explicitly given and illustrated on artificial data as well as real-world examples. The main advantages of our method include variable selection for both linear and nonlinear predictions, effective dimension reduction with sparse loadings, and an efficient algorithm for large p, small n problems.

Keywords Gradient learning · Variable selection · Effective dimension reduction · Forward-backward splitting

1 Introduction

Datasets with many variables have become increasingly common in biological and physical sciences. In biology, it is nowadays a common practice to measure the expression values of

Editor: P. Long.

G. Ye · X. Xie (✉)
School of Information and Computer Science, University of California, Irvine, CA 92697, USA
e-mail: xhx@ics.uci.edu

G. Ye
e-mail: yeg@uci.edu