Local linear regression on manifolds and its geometric interpretation

Ming-Yen Cheng[1]* (鄭明燕) and Hau-Tieng Wu[2]
[1]Department of Mathematics, National Taiwan University, Taipei, Taiwan
[2]Department of Statistics, University of California, Berkeley, USA

Abstract

High-dimensional data analysis has been an active area, and the main focuses have been variable selection and dimension reduction. In practice, it occurs often that the variables are located on an unknown, lower-dimensional nonlinear manifold. Under this manifold assumption, one purpose of this paper is regression and gradient estimation on the manifold, and another is developing a new tool for manifold learning. To the first aim, we suggest directly reducing the dimensionality to the intrinsic dimension of the manifold, and performing the popular local linear regression (LLR) on a tangent plane estimate. An immediate consequence is a dramatic reduction in the computational time when the intrinsic dimension is much smaller than the ambient space dimension. We provide rigorous theoretical justification of the convergence of the proposed regression and gradient estimators by carefully analyzing the curvature, boundary, and non-uniform sampling effects. A bandwidth selector that can handle heteroscedastic errors is proposed. To the second aim, we analyze carefully the behavior of our regression estimator both in the interior and near the boundary of the manifold, and make explicit its relationship with manifold learning, in particular estimating the Laplace-Beltrami operator of the manifold. In this context, we also make clear that it is important to use a smaller bandwidth in the tangent plane estimation than in the LLR. A simulation study and applications to the Isomap face data and a clinical computed tomography scan dataset are used to illustrate the computational speed and estimation accuracy of our methods.

Keywords: diffusion map, dimension reduction, high-dimensional data, manifold learning, nonparametric regression