Nonparametric Regression for Manifold Data via Embedding Distance

Masaaki Imaizumi (Institute of Statistical Mathematics)
joint work with Keisuke Yano (The University of Tokyo)

1 Manifold Data Analysis

There are growing demands for analyzing data which have support on non-Euclidean spaces. For typical examples, directional data and periodic data are regarded to be contained in *circles*, and images represented by their landmarks compose Kendall's *shape spaces* that are related to *spheres*.

To characterize the geometry of non-Euclidean spaces containing data, an idea of a *manifold* are often introduced. A manifold is defined as a topological space equipped with a locally Euclidean structure. Using the geometry of manifolds, various statistical methods are applied to analyze data in manifolds. Patrangenaru and Ellingson (2015) provides an overview and recent developments of the manifold data analysis.

2 Nonparametric Regression for Manifold Data

Despite the possibility of the manifold data analysis, there are still statistical and computational difficulties for handling the geometry (Bhattacharya *et al.* (2012)), since geometric structures of manifolds are sometimes complex or unreliable from observations. To handle the difficulties, we propose a framework using an embedding approach which measures the geometry of manifolds in the Euclidean space into which the manifolds are embedded.

We provide an estimator for nonparametric regression for manifold data with the distance measured by an embedding. Let $\mathcal{M}_x$ and $\mathcal{M}_y$ be m_x- and m_y-dimensional compact Riemannian manifolds, $\{(X_i, Y_i)\}_{i=1}^n$ be a set of n i.i.d. random objects in $\mathcal{M}_x \times \mathcal{M}_y$, and $\|\cdot\|$ be the Euclidean norm. For the approach, we introduce embeddings $e_x : \mathcal{M}_x \to \mathbb{R}^{D_x}$ and $e_y : \mathcal{M}_y \to \mathbb{R}^{D_y}$, a projection operator $\Pi : \mathbb{R}^{D_y} \to e_y(\mathcal{M}_y)$, a kernel function $K : \mathbb{R}_+ \to \mathbb{R}_+$, and a bandwidth $h > 0$. Also, we define a matrix $\mathbf{X}_{x,i} = (1, e_x(X_i) - e_x(x))^T$ and a vector b with length $D_x + 1$ having its first element is 1 and other elements are 0. Then, we propose a local linear estimator for manifold data as

$$\widehat{f}(x) = e_y^{-1}(\Pi(b^T \widehat{\theta}_x)),$$

$$\text{where} \quad \widehat{\theta}_x \in \operatorname*{argmin}_{\theta \in \mathbb{R}^{(D_x+1)\times D_y}} \sum_{i=1}^{n} \left\| e_y(Y_i) - \theta^T \mathbf{X}_{x,i} \right\|^2 K(h^{-1} \| e_x(X_i) - e_x(x) \|).$$

In the presentation, we will report that the estimator with the embedding is computationally efficient while preserving statistical efficiency, and the embedding approach can address the statistical unreliability of manifolds.

Reference

- Patrangenaru, V., & Ellingson, L. (2015). *Nonparametric statistics on manifolds and their applications to object data analysis.* CRC Press.

- Bhattacharya, R. N. *et al.* (2012). Extrinsic analysis on manifolds is computationally faster than intrinsic analysis with applications to quality control by machine vision. *Applied Stochastic Models in Business and Industry*, 28(3), 222–235.