Manifold Learning: geometrical and statistical models of high-dimensional data

Alexander Bernstein, Skolkovo Institute of Science and Technology, Moscow, Russia
a.bernstein@skoltech.ru

Abstract

Many Data Analysis tasks deal with high-dimensional data, and curse of dimensionality is an obstacle to the use of many methods for their solutions. In many applications, real-world data occupy only a very small part of high-dimensional observation space whose intrinsic dimension is essentially lower than dimension of the ambient space. Manifold learning is the newly emerging direction in Data analysis whose main subject is the solution of various Data analysis tasks under Manifold assumption in accordance with which the data lie on or near an unknown manifold (Data manifold) of lower dimensionality embedded in an ambient high-dimensional space. General goal of Manifold learning is discovering a low-dimensional structure of high-dimensional manifold valued data from given dataset. If dataset points are sampled according to an unknown probability measure on an unknown DM, we face with statistical problems about manifold valued data.

Manifold learning uses various statistical and differential-geometrical methods for finding of geometrical and topological structure in data which, in turn, are grounded in solid results from various branches of mathematics, including combinatorial and discrete geometry, differential geometry, topology, functional analysis, and stochastic modelling.

The talk gives short review of statistical problems regarding unknown high-dimensional manifold valued data and the solutions to them, including own results obtained in recent years. Considered examples concern various statistical problems such as estimation of unknown Data manifold and related objects (manifold dimensionality, manifold tangent spaces, manifold Riemannian tensor, tangent vector fields on the manifold, etc.), multi-output regression problems on unknown Data manifold, estimation of density whose support in unknown Data manifold, etc.