09:00-12:00, Dec. 11th, 2015
Keynote Speech (1)

Information geometry and spontaneous data learning

Shinto Eguchi

Institute of Statistical Mathematics

Abstract

There are two parts in my talk, in which one focuses on generalized geodesic curves in a space of probability density functions; the other does a statistical application based on the geometry.

1. We introduce a class of paths or one-parameter models connecting arbitrary two probability density functions (pdf's). The path is derived by employing the Kolmogorov-Nagumo (K-N) average between the two pdf's defined by strictly increasing function φ, which we call φ-path. It gives another framework of the geometry for the space of all the pdfs. Such an information geometric insight provides understandings for probabilistic properties for statistical methods associated with the path connectedness. Here we overview a dualistic relation between statistical model and estimation, which is focused on a maximum entropy and minimum divergence. In particular, we show a close relation of φ-path, *U*-entropy and *U*-divergence. The φ-path is extended to a φ-surface, which is a maximal entropy model, on which the minimum divergence estimator is characterized by canonical statistics.

2. We discuss an approach called spontaneous data learning (SDL) to open novel explanatory paradigm connecting parametrics with nonparametrics. The statistical performance for SDL is explored from information geometric viewpoint, so that SDL gives a new perspective beyond the discussion for robustness or misspecification of parametric model. If the true distribution is exactly in the parametric model, the theory of statistical estimation has been well established, in which any minimum divergence estimator satisfies parametric consistency. We focus on a collapse of the parametric theory perturbing toward a nonparametric setting, where the true distribution may range from unimodality to multimodality; various estimators are targeted and investigated in a class of minimum divergence. In this context a selection of estimators is explored rather than model selection. Specifically we choose the power divergence class under a normal mean model, where the true distribution is, for example, a mixture of *K* distributions. Then we observe that the local minima of the empirical loss function for the power divergence properly suggest the *K*-means if they are mutually separated in the mixture distribution, and the order of power is appropriated selected. The resulting method for clustering analysis is shown to spontaneously detect the number *K* of clusters. Further, we observe that the normalized empirical loss function converges to the true density function if the power parameter goes to infinity. As a result the power parameter combines between the parametric and nonparametric consistency.