Two geometric problems in optimal transport: Discrete and Gaussian measures

Yoav Zemel
University of Cambridge

September 8, 2019

This talk consists of two parts. The first part (joint with Max Sommerfeld, Jörn Schrieber & Axel Munk) proposes a simple subsampling scheme for fast randomized approximate computation of optimal transport distances on finite spaces. This scheme operates on a random subset of the full data and can use any algorithm as a black-box back-end, including state-of-the-art solvers and entropically penalized versions. It is based on averaging the exact distances between empirical measures generated from independent samples from the original measures and can easily be tuned towards higher accuracy or shorter computation times. To this end, we give non-asymptotic deviation bounds for its accuracy in the case of discrete optimal transport problems. In particular, we show that in many important instances, including images (2D-histograms), the approximation error is independent of the size of the full problem. We present numerical experiments that demonstrate that a very good approximation in typical applications can be obtained in a computation time that is several orders of magnitude smaller than what is required for exact computation of the full problem.

The second part of the talk (joint with Valentina Masarotto & Victor M. Panaretos) identifies the Procrustes metric on infinite-dimensional covariance operators with the optimal transport distance. The identification allows us to provide a detailed description of aspects of this distance that are of statistical importance, including: the manifold-like geometry of the space of covariance operators endowed with the Procrustes distance; key properties of the Fréchet mean of a random sample of covariance; and a data-generating mechanism associated with this distance. We take advantage of these properties for carrying out principal component analysis on covariance operators and for testing homogeneity of covariances of several functional populations using the optimal transport maps.