Density estimation through flows in feature space

Esteban G. Tabak
Courant Institute
New York University

Berlin, November 29, 2010

Abstract

A central problem in the analysis of data is that of density estimation: given a set of independent observations $x_j \in \mathbf{R}^n,\ j = 1, \ldots, m$, estimate its underlying probability distribution $\rho(x)$. Density estimation is used in classification, clustering and dimensional reduction, as well as in more field-specific applications such as medical diagnosis, option pricing and weather prediction.

This talk will describe a non-parametric methodology for density estimation, based on normalizing flows in feature-space. Here the observations x_j act as active Lagrangian markers of a flow $\phi_t(x)$ in $\mathbf{R}^n$, converging to a variable $y = \phi_\infty(x)$ which is normally distributed. The knowledge of a change of variables $y(x)$ which normalizes the flow implies a knowledge of the original probability density $\rho(x)$.

After describing the basic algorithm and illustrating it in applications, extensions will be discussed to classification, conditional density estimation and density estimation with constraints.