Uncertainty quantification in graph-based classification of high dimensional data

Andrea Bertozzi, UCLA

Classification of high dimensional data finds wide-ranging applications.
In many of these applications equipping the resulting classification with a measure of uncertainty may be as important as the classification itself. I discuss recent work that develops algorithms for, and investigate the properties of a variety of Bayesian models for the task of binary classification; via the posterior distribution on the classification labels, these methods automatically give measures of uncertainty. The methods are all based on the graph formulation of semisupervised learning. We provide a unified framework which brings together a variety of methods that have been introduced in different communities within the mathematical sciences. We introduce efficient numerical methods, suited to large datasets, for both MCMC-based sampling and gradient-based MAP estimation.
Through numerical experiments we study classification accuracy and uncertainty quantification for our models; these experiments showcase a suite of datasets commonly used to evaluate graph-based semisupervised learning algorithms. I conclude with an application involving a human-in-the-loop, such as a security analyst, who can interact with the algorithm by hand classifying a subset of data determined by the machine. I discuss recent results on these problems for classification of ego motion in body worn camera data.

This is joint work with Andrew Stuart, Xiyang Luo, Konstantinos Zygalakis and Hao Li.