

Manifold Regularization and Low-Density Techniques for Semi-supervised Learning

Vikas Sindhwani
University of Chicago
<http://www.cs.uchicago.edu/~vikass>

Tuesday, February 13, 2007
2:30-4:00pm
TECH Center 111

Abstract:

In many applications of machine learning, large amounts of data can be cheaply and automatically collected. However, the process of manually labeling this data for the purposes of training learning algorithms is often a slow, expensive, and/or error-prone process. Several semi-supervised inference algorithms have recently been designed that attempt to use both labeled and unlabeled examples for effective learning. In this talk, I will present two families of algorithms, Manifold Regularization and Low-Density Classification, for extending kernel methods (such as Support Vector Machines) for Semi-supervised learning. These algorithms are based on different assumptions on the structure and geometry of the probability distribution underlying the data.

Manifold Regularization is motivated by the observation that in many applications, data is very high-dimensional in its raw format, but points truly reside near a low-dimensional, non-linear manifold in the ambient space. This manifold is estimated by a nearest-neighbor graph over unlabeled examples and graph-Laplacian regularizers are combined with standard kernel methods, so that the problem of out-of-sample extension in a large class of graph-based techniques is naturally resolved.

Low-Density Classification techniques learn decision surfaces that respect data clusters revealed by unlabeled examples, and pass through low-density regions in the input space. The associated optimization problems are non-convex. I will present a deterministic annealing approach to alleviate local minima problems in this class of methods. Empirical results will be presented on large-scale text categorization problems.

Joint work with Partha Niyogi, Mikhail Belkin, Sathiya Keerthi, Chu Wei and Olivier Chapelle.