MATTHEW THORPE, University of Warwick

How Many Labels Do You Need in Semi-Supervised Learning?

Semi-supervised learning (SSL) is the problem of finding missing labels from a partially labelled data set. The heuristic one uses is that "similar feature vectors should have similar labels". The notion of similarity between feature vectors explored in this talk comes from a graph-based geometry where an edge is placed between feature vectors that are closer than some connectivity radius. A natural variational solution to the SSL is to minimise a Dirichlet energy built from the graph topology. And a natural question is to ask what happens as the number of feature vectors goes to infinity? In this talk I will give results on the asymptotics of graph-based SSL using an optimal transport topology. The results will include a lower bound on the number of labels needed for consistency.