Data Complexity Analysis: Linkage between Context and Solution in Classification

Tin Kam Ho

Abstract. For a classification problem that is implicitly represented by a training data set, analysis of data complexity provides a linkage between context and solution. Instead of directly optimizing classification accuracy by tuning the learning algorithms, one may seek changes in the data sources and feature transformations to simplify the class geometry. Simplified class geometry benefits learning in a way common to many methods. We review some early results in data complexity analysis, compare these to recent advances in manifold learning, and suggest directions for further research.