

Learning from Data of Variable Quality

Michael Kearns

University of Pennsylvania

We initiate the study of learning from multiple sources of limited data, each of which may be corrupted at a different rate. We develop a complete theory of which data sources should be used for two fundamental problems: estimating the bias of a coin, and learning a classifier in the presence of label noise. In both cases, efficient algorithms are provided for computing the optimal subset of data.