Estimating Learnability

Gregory Valiant, Stanford University

We consider the problem of estimating how well a model class is capable of fitting a distribution of labeled data. We show that it is often possible to accurately estimate this ``learnability'' even when given an amount of data that is too small to reliably learn any accurate model. Our first result applies to the setting where the data is drawn from a d-dimensional distribution with isotropic covariance, and the label of each datapoint is an arbitrary noisy function of the datapoint. In this setting, we show that with O(sqrt(d)) samples, one can accurately estimate the fraction of the variance of the label that can be explained via the best linear function of the data. For comparison, even if the labels are noiseless linear functions of the data, a sample size linear in the dimension, d, is required to learn any function correlated with the underlying model. Our estimation approach also applies to the setting where the data distribution has an (unknown) arbitrary covariance matrix, allowing these techniques to be applied to settings where the model class consists of a linear function applied to a nonlinear embedding of the data. In this setting we give a consistent and minimax optimal estimator of the fraction of explainable variance that uses o(d) samples. Finally, our techniques also extend to the setting of binary classification, where we obtain analogous results under the logistic model, for estimating the classification accuracy of the best linear classifier.
We demonstrate the practical viability of our approaches on synthetic and real data.
This is based on joint work with Weihao Kong.