A precise high-dimensional asymptotic theory for boosting

This talk will introduce a precise high-dimensional asymptotic theory for Boosting on separable data, taking both statistical and computational perspectives. We will consider the common modern setting where the number of features (weak learners) and the sample size are both large and comparable, and in particular, look at scenarios where the data is separable in an asymptotic sense. On the statistical front, we will provide an exact analysis of the generalization error of Boosting, when the algorithm interpolates the training data and maximizes an empirical L1 margin. The angle between the Boosting solution and the ground truth can be explicitly characterized. On the computational front, we will provide a sharp analysis of the stopping time when Boosting approximately maximizes the empirical L1 margin. Our theory provides several insights into properties of Boosting, for instance, we discover that the larger the margin, the smaller the proportion of active features (with zero initialization). This is based on joint weork with Tengyuan Liang.