Ingrid Daubechies

Title: Convergence results and counterexamples for AdABoost and related algorithms.

Abstract:

It concerns a classification task, in which a database of examples (j, y_j) is given, where j ranges over an index set $\mathcal{D}$ of cardinality D, and where $y_j \in \{1, -1\}$ for each j; on the other hand, there is a collection $\mathcal{H}$ of cardinality H, of "hypotheses" h_i, which are maps from $\{1, 2, \ldots, D\}$ to $\{1, -1\}$. Hypothesis h_i is said to classify example j correctly if $h_i(j) = y_j$;. Typically none of the h_i is expected to get much more than 50% of the examples classified correctly: they are "weak" classifiers. We consider the case where a "strong" classifier can be constructed by convex combination of the weak classifiers. i.e. there exists some $\nu \in [0,1]^H$, with $\sum_{i=1}^H \nu_i = 1$, such that $\texttt{sign}\left(\sum_{i=1}^H \nu_i h_i(j)\right) = y_j$, for all $j \in \{1, 2, \ldots, D\}$. The goal is to construct the classifier with optimal margin, i.e. to determine $\nu \in [0,1]^H$ that maximizes $\min_{j \in \{1,2,\ldots,D\}} \left| y_j - \sum_{i=1}^H \nu_i h_i(j) \right|$; we assume that both $\mathcal{H}$ and $\mathcal{D}$ are much too large to adopt any type of steepest gradient algorithm.

AdaBoost is an iterative algorithm for classification, proposed by Freund and Schapire, that has been remarkably successful in a wide range of applications. It had been assumed to converge to the optimal margin classifier. We shall see that AdaBoost does always converge, but not necessarily to the optimal margin classifier, and define slight modifications of AdaBoost that have the desired limit.