

Title:

On Two Sequential Methods in Classification

Abstract:

This talk covers two recent methods in statistical classification that lead to optimal risk bounds. The first method corresponds to binary loss classification, also known as the PAC learning setup. In this context, any method with a bounded permutation or leave-one-out error can be adapted to provide an optimal high-probability guarantee. By splitting the sample into two parts and sequentially adding examples from the second part, a technique based on a combination of forward and reverse martingale inequalities shows that a majority vote over these classifiers is statistically optimal, given a small leave-one-out error bound. The second method corresponds to logistic regression, where the algorithm is based on the conversion of the sequential algorithm into the statistical one. This algorithm uses a refined version of the exponential weights algorithm with data-dependent prior distributions and includes an additional step that aggregates solutions with large margins. This technique results in non-asymptotic statistical excess risk bounds that are completely distribution-free for the logarithmic loss, independent of the distribution of design vectors or the magnitude of optimal solutions.