

A statistical learning approach to ranking

Empirical risk minimization revisited

Nicolas VAYATIS

Abstract:

The statistical ranking problem consists of ordering instances from a high-dimensional space. This question arises in a large variety of applications, ranging from the design of search engines in information retrieval to medical diagnosis or credit-risk screening. A natural approach consists of "projecting" these instances onto the real line through some real-valued scoring function. We focus here on the setup where binary label information is available. This problem is known as the bipartite ranking problem. In scoring applications, the majority of ranking methods is developed in the spirit of logistic regression and relies on the statistical modelling of the regression function using additive models. The statistical learning approach is different insofar as it avoids the difficult problem of estimating the distribution in a high dimensional setup and focuses on prediction. Statistical learning strategies can be seen as implementations of the optimization of performance measures based on data. In this talk, we describe the nature of the ranking problem, and discuss various inference principles in this context.

Keywords:

Bipartite ranking, ROC curve, AUC, higher order statistics