Bipartite ranking and extensions

Sylvain Robbiano

CIMFAV, Universidad de Valparaíso.

Abstract: The bipartite ranking problem consists in learning from a sample observations, while preserving the order of their associated labels. A standard approach in this context involves the introduction of a scoring function. The first part of the talk is dedicated to present the theoretical result and some algorithms to solve this supervised task.

In a second part, two extensions are presented. First, we consider this problem in the high dimensional situation, where the observations Xi lie in a space of dimension d, possibly much larger than the sample size n. We propose to estimate the optimal scoring function using the so-called Gibbs posterior distribution, which favors sparse additive estimators. Using elements from the PAC-Bayesian theory, we provide theoretical guarantees about our method.

Secondly, it is proved that, in the case where the data generating probability distribution has compact support, anomaly ranking is equivalent to (supervised) bipartite ranking, where the goal is to discriminate between the underlying probability distribution and the uniform distribution with same support. Exploiting this view, we then show how to use bipartite ranking algorithms, possibly combined with random sampling, to solve the anomaly ranking problem.