Geometric Aspects of Statistical Learning Theory

Shahar Mendelson

The aim of the lectures is to survey the progress that has been made in Learning Theory over the last 15 years, mainly in the understanding of the way prediction bounds relate to the geometry of the underlying class.

In the lectures I hope to cover three main topics:

1) Aspects of the classical approach to prediction bounds, in which both class and target are assumed to be uniformly bounded in L_∞. This assumption allows one to analyze prediction problems using standard methods from Empirical Processes Theory, like contraction and concentration.

We will focus on a geometric viewpoint, highlighting the idea of comparing empirical and actual structures, and relating the error rate to the structure of certain random sets naturally associated with the problem.

2) The quadratic empirical process:
Analyzing the quadratic empirical process is an essential component in many natural problems, including prediction problems involving the squared loss.
Unfortunately, since powers (in this case, the square) of functions extenuate their peaky part, it is much harder to deal with the quadratic empirical process than with the standard empirical process.
I will explain why it is still possible to obtain a sharp estimate on this process even when the underlying class is unbounded. This requires accurate information on the fine structure of the same random sets mentioned in 1).
As applications, I will present sharp estimates on the singular values of random matrices and on Gelfand widths of convex bodies in various situations. I will also establish minimax bounds on the error rate for prediction problems involving subgaussian classes, (e.g. for compressed sensing and phase recovery problems).

3) All the methods that will be surveyed in 1) and 2) are based on various aspects of concentration - relatively straightforward in the bounded case, and considerably harder for unbounded classes.
However, by concentration results one means a two-sided estimate and I will show that the difficulty in obtaining concentration is due to the 'upper-tail', while the 'lower-tail' is almost universally true in some sense. Moreover, I will show that only this 'lower-tail' is truly needed for obtaining prediction bounds, resulting in a rather general theory of learning in situation where concentration is simply not possible.