Covering numbers in Gaussian RKHS

Thomas Kühn (Universität Leipzig, Germany)

Abstract. The topic of the talk is motivated by a problem in mathematical learning theory. In the pioneering paper by Cucker and Smale (*Bull. AMS* 2001) it was shown that functional analytic concepts could be very fruitful in statistical learning theory. Modern machine learning methods, like support vector machines, often use Gaussian kernels and their corresponding reproducing kernel Hilbert spaces (RKHS). Then, in estimating the probabilistic error and the number of required samples, bounds for the covering numbers of the unit ball of the RKHS are needed.

Gaussian kernels are functions of the form

$$K(x, y) = \exp(-\sigma^2 |x - y|^2), \quad x, y \in X,$$

where X is a subset of $\mathbb{R}^d$ and $\sigma > 0$ a parameter. Since these kernels are continuous and positive definite, they generate a RKHS which consists of continuous functions only.

Given $\varepsilon > 0$, the covering number $N(\varepsilon, A; M)$ of a subset A in a metric space M is defined as the minimal number of balls in M of radius ε that are necessary to cover A. The function $H(\varepsilon, A; M) := \log N(\varepsilon, A; M)$, which is often sufficient in applications, is Kolmogorov's famous $\varepsilon-$entropy.

In the context of learning theory, Ding-Xuan Zhou (*J. Complexity* 2002) proved that for $X = [0, 1]^d$ the covering numbers $N(\varepsilon)$ of the unit ball of a Gaussian RKHS on X, considered as a subset of $C(X)$, satisfy the upper estimate

$$\log N(\varepsilon) = \mathcal{O}\left(\left(\log \tfrac{1}{\varepsilon}\right)^{d+1}\right) \quad \text{as } \varepsilon \to 0$$

and conjectured that the correct bound is $\left(\log \frac{1}{\varepsilon}\right)^{d/2+1}$.

The aim of this talk is to prove that the exact asymptotic behaviour is

$$\log N(\varepsilon) = \mathcal{O}\left(\frac{\left(\log \frac{1}{\varepsilon}\right)^{d+1}}{\left(\log \log \frac{1}{\varepsilon}\right)^{d}}\right) \quad \text{as } \varepsilon \to 0\,.$$

This shows that Zhou's upper bound was almost sharp, up to a double logarithmic factor, but his conjecture was far too optimistic. While Zhou derived his result from smoothness properties of the Fourier transform of the kernel, our proof is based on the explicit description of Gaussian RKHS, which was established recently by Steinwart, Hush and Scovel (*IEEE Trans. Information Theory* 2006).