Effective dimension in sample-complexity bounds for RKHS interpolation

Andrew D. McRae

We consider the problem of interpolating a function from randomly-sampled points, using kernel interpolation in a reproducing kernel Hilbert space (RKHS) framework. We show that the number of samples required to estimate any function in the kernel's RKHS (or native space) to within a specified amount of L_2 error is proportional to the *effective dimension* of the RKHS in L_2, which is determined by the eigenvalues of the kernel's associated integral operator.

More precisely, suppose k is a positive definite kernel (with associated RKHS $\mathscr{H}$) on a space X and μ is a probability measure on X. Let $T \colon L_2 \to L_2$ be the integral operator with kernel k and measure μ, and let $t_1 \geq t_2 \geq \cdots$ be its eigenvalues. Under some mild assumptions, if p is a positive integer, and we take $n \gtrsim p \log p$ samples of $f^* \in \mathscr{H}$ at points chosen independently according to μ, then we show that the kernel interpolant $\hat{f}$ satisfies

$$\|\hat{f} - f^*\|_{L_2}^2 \lesssim t_p \|f^*\|_{\mathscr{H}}^2$$

with high probability, uniformly in f^*.

The value of p at which the eigenvalue t_p becomes negligible (say, smaller than a chosen tolerance ϵ) is what we are calling the *effective dimension*. In many cases, the eigenvalues decay rapidly (often exponentially) beyond a particular value of p, so this dimension does not strongly depend on the threshold ϵ. Some examples that have fast eigenvalue decay include a Gaussian radial basis function kernel on a domain in Euclidean space (a common setup in machine learning), a sinc function kernel on a finite interval (important for the estimation of bandlimited functions), and the heat kernel on a Riemannian manifold.

Our results also extend to the RKHS regression problem, in which our function samples are corrupted by noise, and we must use Tikhonov regularization in the RKHS to ensure a stable solution. In this case, our work recovers state-of-the-art existing results for regularized RKHS regression.