"Best Choices for Regularization Parameters in Learning Theory: On the Bias-Variance Problem"

Felipe Cucker[a] and Steve Smale[b]

[a]Department of Mathematics
City University of Hong Kong
83 Tat Chee Avenue, Kowloon
HONG KONG
e-mail: macucker@math.cityu.edu.hk

[b]Department of Mathematics
University of California, Berkeley
Berkely, California
e-mail: smale@math.berkeley.edu

The goal of learning theory is to find an approximation of a function $f_\rho : X \to Y$ known only through a set of pairs $\mathbf{z} = (x_i, y_i)_{i=1}^m$ drawn from an unknown probability measure ρ on $X \times Y$ (f_ρ is the "regression function" of ρ).

An approach championed by Poggio with ideas going back to Ivanov and Tikhonov is to minimize

$$\frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2 + \gamma\|Af\|_{L^2}^2$$

where A is an differential operator and L^2 is the Hilbert space of square integrable functions on X with measure ρ_X on X defined via ρ.

This minimization is well-conditioned and solved by straightforward finite dimensional least squares linear algebra to yield $f_{\gamma,\mathbf{z}} : X \to Y$. The problem is posed: *How good an approximation is $f_{\gamma,\mathbf{z}}$ to f_ρ, or measure the error $\int_X (f_{\gamma,\mathbf{z}} - f_\rho)^2$?* and *What is the best choice of γ to minimize this error?*

Our goal in this talk is to give some answers to these questions.