ON OPTIMAL ESTIMATORS IN LEARNING THEORY

V.N. Temlyakov

Abstract. This talk addresses some problems of supervised learning in the setting formulated by Cucker and Smale. Supervised learning, or learning-from-examples, refers to a process that builds on the base of available data of inputs x_i and outputs y_i, $i = 1, \dots, m$, a function that best represents the relation between the inputs $x \in X$ and the corresponding outputs $y \in Y$. The goal is to find an estimator $f_{\mathbf{z}}$ on the base of given data $\mathbf{z} := ((x_1, y_1), \dots, (x_m, y_m))$ that approximates well the regression function f_ρ (or its projection) of an unknown Borel probability measure ρ defined on $Z = X \times Y$. We assume that (x_i, y_i), $i = 1, \dots, m$, are indepent and distributed according to ρ.

There are several important ingredients in mathematical formulation of this problem. We follow the way that has become standard in approximation theory and has been used in recent papers. In this approach we first choose a function class W (a hypothesis space $\mathcal{H}$) to work with. After selecting a class W we have the following two ways to go. The first one is based on the idea of studying approximation of the $L_2(\rho_X)$ projection $f_W := (f_\rho)_W$ of f_ρ onto W. Here, ρ_X is the marginal prabability measure. This setting is known as the *improper function learning problem* or the *projection learning problem*. In this case we do not assume that the regression function f_ρ comes from a specific (say, smoothness) class of functions. The second way is based on the assumption $f_\rho \in W$. This setting is known as the *proper function learning problem*. For instance, we may assume that f_ρ has some smoothness. We will give some upper and lower estimates in both settings.