Concentration estimates for learning with unbounded sampling

Zheng-Chu Guo · Ding-Xuan Zhou

Received: 17 May 2010 / Accepted: 12 September 2011 / Published online: 11 October 2011 © Springer Science+Business Media, LLC 2011

Abstract The least-square regression problem is considered by regularization schemes in reproducing kernel Hilbert spaces. The learning algorithm is implemented with samples drawn from unbounded sampling processes. The purpose of this paper is to present concentration estimates for the error based on $\ell^2$-empirical covering numbers, which improves learning rates in the literature.

Keywords Learning theory · Least-square regression · Regularization in reproducing kernel Hilbert spaces · Empirical covering number · Concentration estimates

Mathematics Subject Classifications (2010) 68T05 · 62J02

Communicated by Lixin Shen.

The work described in this paper is supported partially by the Research Grants Council of Hong Kong [Project No. CityU 103508] and National Science Fund for Distinguished Young Scholars of China [Project No. 10529101].

Z.-C. Guo
School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou 510275, People’s Republic of China
e-mail: gzhengchu2@cityu.edu.hk

Z.-C. Guo · D.-X. Zhou (✉)
Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong, People’s Republic of China
e-mail: mazhou@cityu.edu.hk
1 Introduction and main results

We consider the regression problem by learning with unbounded sampling processes. The target functions for learning are defined on a complete separable metric space $X$ (input space) and take values in $Y = \mathbb{R}$ (output space). Learning is implemented by algorithms and samples. In our setting it is assumed that a sample $z = \{z_i = (x_i, y_i)\}_{i=1}^m$ is drawn independently from a Borel probability measure $\rho$ on $Z := X \times Y$. Our target is the regression function defined by

$$f_\rho(x) = \int_Y y d\rho(y|x), \quad x \in X,$$

where $\rho(\cdot|x)$ is the conditional distribution of $\rho$ at $x \in X$.

The learning algorithm studied in this paper is based on a bounded Mercer kernel $K : X \times X \rightarrow \mathbb{R}$ which is a continuous, symmetric, and positive semi-definite function. The reproducing kernel Hilbert space (RKHS) $H_K$ associated with $K$ is the completion of span$\{Kx = K(\cdot, x) : x \in X\}$ with respect to the inner product $\langle \cdot, \cdot \rangle_K$ given by

$$\langle Kx, Kx' \rangle_K = K(x, x').$$

The learning algorithm considered here is a regularization scheme in $H_K$ given by

$$f_{z,\lambda} = \arg \min_{f \in H_K} \left\{ \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 + \lambda \| f \|_K^2 \right\}, \quad (1.1)$$

where $\lambda = \lambda(m) > 0$ is a regularization parameter.

There is a large literature on error analysis for algorithm (1.1). See e.g. [8, 10, 12, 15, 23, 28]. Most results are stated under the standard uniform boundedness assumption for the output that for some constant $M > 0$, $|y| \leq M$ almost surely. This standard assumption is abandoned in [5, 9, 22]. In [5, 9], it is assumed that the output satisfies the condition

$$\int_Y \left( \exp \left\{ -\frac{|y - f_{H_\lambda}(x)|^2}{M} \right\} - \frac{|y - f_{H_\lambda}(x)|}{M} - 1 \right) d\rho(y|x) \leq \frac{\sum_2}{2M^2}, \quad \text{a.e. } x \in X \quad (1.2)$$

for some constants $M, \Sigma > 0$, where $f_{H_\lambda}$ is the orthogonal projection of $f_\rho$ onto the closure of $H_K$ in $L_{\rho_X}^2$. Under the assumption $f_{H_\lambda} \in H_K$, bounds for the error $\| f_{z,\lambda} - f_\rho \|_{L_{\rho_X}^2}$ were given in [5, 9], where $\| f \|_{L_{\rho_X}^2} = (\int_X |f(x)|^2 d\rho_X)^{1/2}$ and $\rho_X$ is the marginal distribution of $\rho$ on $X$. How to relax the restriction $f_{H_\lambda} \in H_K$ in the error analysis with $|y| \leq M$ was studied in [12, 17]. Further to the previous study in [22], we consider the setting satisfying the following condition.

**Moment hypothesis**: there exist constants $M \geq 1$ and $C > 0$ such that

$$\int_Y |y|^{\ell} d\rho(y|x) \leq C\ell! M^\ell, \quad \forall \ell \in \mathbb{N}, x \in X. \quad (1.3)$$