Available online at www.sciencedirect.com

ScienceDirect

Journal of Approximation Theory 236 (2018) 1–22

Journal of Approximation Theory

www.elsevier.com/locate/jat

Full Length Article

Coefficient-based l^q-regularized regression with indefinite kernels and unbounded sampling[☆]

Qin Guo[a], Cheng Wang[b,*], Peixin Ye[a]

[a] *School of Mathematical Sciences and LPMC, Nankai University, Tianjin, 300071, China*
[b] *Department of Mathematics, Huizhou University, Huizhou, Guangdong, 516007, China*

Received 9 October 2017; received in revised form 23 May 2018; accepted 30 July 2018
Available online 4 August 2018

Communicated by Bin Han

Abstract

We consider the kernel-based coefficient least squares learning algorithm for regression with l^q-regularizer, $1 < q \leq 2$. Our error analysis is carried out under more general conditions. The kernel function may be non-positive definite and the output sample values satisfy the moment hypothesis rather than the uniform boundedness. We derive the capacity dependent error bounds of the algorithm by constructing the stepping stone function for the indefinite kernels. When the output values are bounded, we obtain a learning rate that can be arbitrarily close to the best rate $O(m^{-1})$ under some mild conditions of the regression function and the hypothesis space.

MSC: 41A17; 68T05; 62J02

Keywords: Coefficient-based regularized regression; Indefinite kernel; Moment hypothesis; Stepping stone function; Learning rate

☆ The work was supported by the Natural Science Foundation of China [Grant Nos. 11271199, 11671213, 11401247] and the Natural Science Foundation of Guangdong Province [Grant No. 2015A030313674].
* Corresponding author.
E-mail address: wang_cheng20@163.com (C. Wang).

https://doi.org/10.1016/j.jat.2018.07.003

1. Introduction and main results

Supervised learning aims at finding the function relationship between inputs and outputs from the observed samples. In this paper we consider the least squares for regression which is a classical topic in learning theory. It can be formulated as follows. Let X be a compact metric space and $Y=\mathbb{R}$. Let ρ be a Borel probability measure on $Z=X\times Y$. We define the generalization error for a function $f:X\to Y$ as

$$\mathcal{E}(f)=\int_Z (f(x)-y)^2 d\rho.$$

Denote by $L^2_{\rho_X}(X)$ the Hilbert space of the square integrable functions defined on X with respect to the measure ρ_X, where ρ_X is the marginal measure of ρ on X and $\|f(\cdot)\|_{\rho_X}=(\int_X |f(\cdot)|^2 d\rho_X)^{\frac{1}{2}}$. The regression function f_ρ which minimizes $\mathcal{E}(f)$ over all $f\in L^2_{\rho_X}(X)$ is given by

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

where $\rho(\cdot|x)$ is the conditional distribution induced by ρ at $x\in X$. In the framework of supervised learning, ρ is unknown and what we have in hand is a set of random samples $\mathbf{z}=\{z_i\}_{i=1}^m=\{(x_i,\ y_i)\}_{i=1}^m\in Z^m$ drawn from the measure ρ independently and identically. The task of the regression problem is to find a good approximation $f_{\mathbf{z}}$ of f_ρ, which is derived from some learning algorithms by minimizing the empirical error

$$\mathcal{E}_{\mathbf{z}}(f)=\frac{1}{m}\sum_{i=1}^m (f(x_i)-y_i)^2.$$

The quality of approximation of $f_{\mathbf{z}}$ to f_ρ is measured by the excess generalization error

$$\|f_{\mathbf{z}}-f_\rho\|^2_{\rho_X}=\mathcal{E}(f_{\mathbf{z}})-\mathcal{E}(f_\rho).$$

A large class of learning algorithms for regression take the regularization scheme to prevent the ill-posedness in a reproducing kernel Hilbert space (RKHS) associated with a Mercer kernel. Such a kernel K is a function on $X\times X$ which is continuous, symmetric, and positive semi-definite. The RKHS $\mathcal{H}_K$ is defined to be the completion of the linear span of the set of functions $\{K_x:=K(x,\cdot):x\in X\}$ with the inner product

$$\left\langle \sum_{i=1}^n \alpha_i K_{x_i},\ \sum_{j=1}^m \beta_j K_{t_j}\right\rangle_K := \sum_{i=1}^n\sum_{j=1}^m \alpha_i\beta_j K(x_i,t_j).$$

The reproducing property of $\mathcal{H}_K$

$$f(x)=\langle f,\ K_x\rangle_K$$

holds for all $x\in X$ and $f\in\mathcal{H}_K$.

The regularized least squares algorithm for regression in $\mathcal{H}_K$ is given by

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

where $\lambda>0$ is the regularization parameter with $\lim_{m\to\infty}\lambda(m)=0$. The error analysis for the algorithm (1.1) has been well studied in the extensive literature, see [2,1,18,22].