Information Sciences 268 (2014) 370–380

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Statistical analysis of the moving least-squares method with unbounded sampling

Fangchao He [a,b], Hong Chen [c], Luoqing Li [a,*]

[a] School of Mathematics and Statistics, Hubei University, Wuhan 430062, China
[b] School of Science, Hubei University of Technology, Wuhan 430068, China
[c] College of Science, Huazhong Agricultural University, Wuhan 430070, China

ARTICLE INFO

Article history:
Received 26 November 2012
Received in revised form 1 December 2013
Accepted 4 January 2014
Available online 10 January 2014

Keywords:
Learning theory
Moving least-squares method
Unbounded sampling
Sample error
Concentration inequality

ABSTRACT

Moving least-squares method is investigated with samples drawn from unbounded sampling processes. Convergence analysis is established by imposing incremental conditions on moments of sample output and window width. Satisfied convergence rates are derived by means of projection operator and some concentration inequalities.

1. Introduction

In the traditional setting of machine learning, the least-squares (LS) method is usually used to derive a suitable approximation function with good generalization performance [2,6,15,17,23]. It is noteworthy that the LS method is a kind of global approximate method, i.e., the learning samples commonly are regular or concentrated. In many practical machine learning and engineering applications, a large number of regular or concentrated samples need to be learned, a few irregular or scattered samples also need to be investigated because of their special usefulness [3,9,12,14]. For example, in geographical contour drawing, it is important to derive a set of contours but the height is available only for some scattered data sample points [14]. Therefore, it is vital to seek a suitable local approximation method to deal with scattered data.

It is widely recognized that the moving least-squares (MLS) method is an important local approximation method. After McLain introduced the MLS method to approximate a function based on scattered data [14], motivated by various real applications including problems related to approximation theory, data smoothing, statistics and numerical analysis, this method has been studied intensively in the literature of machine learning [9,12,13,19,20,22,25]. The MLS method is described briefly as follows: given a compact set $\Omega \subseteq \mathbb{R}^n$, a continuous function $u \in C(\Omega)$ is required to be reconstructed from its values $u(x_1), \ldots, u(x_N)$ on scattered, distinct data points $V = \{x_1, \ldots, x_N\}$. Then for any $x \in \Omega$, the approximate value $f^*(x)$ of $u(x)$ is defined as follows:

$$f^*(x) = \arg\min_{f \in \mathcal{P}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \omega(x, x_i)(f(x_i) - u(x_i))^2 \right\},$$

* Corresponding author. Tel.: +86 18971655636; fax: +86 2788663346.
E-mail addresses: lilq@hubu.edu.cn, lilq@live.cn (L. Li).

0020-0255/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2014.01.001

where $\mathcal{P} \subseteq C(\Omega)$ is a finite dimensional subspace, usually spanned by polynomials, and ω is a continuous weight function. In short, the major advantage of MLS over LS is flexibility with weights decaying smoothly with the distance from the scattered data points, resulting in a smooth approximation.

Currently, most error analysis results are obtained mainly under the standard boundedness assumption, which requires that sample output labels should be bounded almost surely by some constant. At the same time, research effort has been made to abandon this standard assumption by some relaxed conditions [2,7,16] or moment hypotheses [1,10,11,21]. Many developments have been achieved in the research on regression concerning unbounded sampling [10,21]. However, it is noticeable that previous research mainly focuses on LS regression. To the best of our knowledge, there is not an approximation performance analysis result about using the MLS method. In this paper, we discuss the generalization performance of the MLS method with unbounded sampling.

We first recall the preliminaries about the regression problem. The framework of the regression problem is based on a compact metric space X (input space) and label set $Y = \mathbb{R}$ (output space). It is assumed that training samples $\mathbf{z} = \{z_i = (x_i, y_i)\}_{i=1}^m$ are drawn independently from an underlying probability distribution ρ on $Z := X \times Y$. In regression setting, the aim of learning algorithms is to predict the output of future samples drawn according to the unknown probability distribution ρ. The target function for learning is defined (e.g., [5,17,28]) by

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

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

In this paper, we consider the learning of the regression function f_ρ by the MLS method with unbounded sampling. We study the case when the hypothesis space $\mathcal{H}$ for learning is a finite dimensional subspace of $C(X)$, the space of all continuous functions on X. The most common and important example of hypothesis space $\mathcal{H}$ is the space Π_l of polynomials of degree at most l.

The main purpose of this paper is to conduct sample error analysis relating the MLS method where sample outputs satisfy the following moment hypothesis.

Moment hypothesis: There exist constants $M \geqslant 1$ and $C > 0$ such that

$$\int_Y |y|^\ell d\rho(y|x) \leqslant C\ell! M^\ell, \quad \forall \ell \in \mathbb{N},\ x \in X. \tag{1}$$

The MLS method involves a hypothesis space $\mathcal{H}$ and an MLS weight function [19,20].

Definition 1. Let the hypothesis space $\mathcal{H}$ is a $\tilde{d}$-dimensional subspace of $C(X)$ consisting of Lipschitz functions on X. The MLS weight function $\Phi: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ satisfies:

(a) $\int_{\mathbb{R}^n} \Phi(x,t)\, dt = 1$ for each $x \in \mathbb{R}^n$,
(b) there exist some $q > n+1,\ c_q,\ \tilde{c}_q > 0$ such that

$$\Phi(x,t) \geqslant c_q, \quad \forall |x-t| \leqslant 1,$$

and

$$|\Phi(x,t)| \leqslant \frac{\tilde{c}_q}{(1+|x-t|)^q}, \quad \forall x,t \in \mathbb{R}^n. \tag{2}$$

Given a sample $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m \in Z^m$, a hypothesis space $\mathcal{H}$ and an MLS weight function Φ, we define the estimator $f_\mathbf{z}$ of f_ρ in a pointwise way as

$$f_\mathbf{z}(x) = f_{\mathbf{z},\sigma,x}(x), \quad x \in X,$$

by the MLS method as follows:

$$f_{\mathbf{z},\sigma,x} = \arg\min_{f \in \mathcal{H}} \left\{ \frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (f(x_i) - y_i)^2 \right\}, \tag{3}$$

where $\sigma > 0$ is a window width.

More information has been accumulated on the local approximation of f_ρ by $f_\mathbf{z}$ in the literature of statistics [8,18] and machine learning [13,19,20,22], where the sample $\mathbf{z}$ is deterministic and well distributed. In learning theory, we are interested in bounding the error $\|f_\mathbf{z} - f_\rho\|^2_{L^2_{\rho_X}}$ and its convergence rates as $m \to \infty$ which are used to measure the performance of the learning algorithm. At this time, it is difficult for us to establish the convergence estimate of $\|f_\mathbf{z} - f_\rho\|_{L^2_{\rho_X}}$ directly with unbounded sampling in terms of the previous analysis techniques and results (see [19,20]). To overcome this difficulty,