Pattern Recognition Letters 40 (2014) 88–95

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Confidence and prediction intervals for semiparametric mixed-effect least squares support vector machine ☆

Qiang Cheng [a,*], Jale Tezcan [b], Jie Cheng [c]

[a] Department of Computer Science, Southern Illinois University, Carbondale, IL 62901, United States
[b] Department of Civil and Environmental Engineering, Southern Illinois University, Carbondale, IL 62901, United States
[c] Department of Computer Science and Engineering, The University of Hawaii, Hilo, HI 96720, United States

ARTICLE INFO

Article history:
Received 29 June 2013
Available online 24 December 2013

Keywords:
Semiparametric function estimation
Mixed effect modeling
Least squares support vector machine
Confidence interval
Prediction interval

ABSTRACT

We consider estimating the confidence and prediction intervals for semiparametric mixed-effect least squares support vector machine (LS-SVM). Explicit formulas are derived for confidence and prediction intervals. The accuracy of the derived analytical equations is assessed by comparing with wild cluster bootstrap-*t* method on simulated and real-world data with different levels of random-effect and residual variances, and different numbers of clusters. Close match between the derived expressions and the bootstrap results is observed.

1. Introduction

Significant correlations often exist between observations from the same case, or subject, in datasets from scientific research. To account for these correlations mixed-effect models are often used to analyze such data including longitudinal data. Recently a semiparametric mixed effect model using least squares support vector machine (LS-SVM) has been proposed [1,2]. Its performance has been shown to be superior to NONMEM, a commonly used, physics-based mixed-effect model [3]. It is expected that this semiparametric mixed-effect model will find important applications in various domains. For many real applications, while it is desirable to know the uncertainty associated with modeling and prediction in terms of confidence and prediction intervals, an effective analytical method of estimating such intervals for the semiparametric mixed effect model for LS-SVM is yet to be developed. To this end, this paper provides analytical expressions for the confidence and prediction intervals for the LS-SVM based semiparametric mixed-effect model. Our analytical expressions are expected to facilitate the use of the semiparametric model in diverse fields. The accuracy of the analytical equations is demonstrated on simulated and real-world data by using the wild cluster bootstrap-*t* method.

In deriving the analytical expressions we first show the semiparametric mixed-effect LS-SVM estimate can be explicitly formulated as a linear smoother. Based on the closed-form expression of the smoothing vector, we then derive closed-form expressions of the bias and variance at any point of interest e.g. design points and prediction points. For variance estimation at any point, we use the approach of first estimating the variance at the design points with residual maximal likelihood estimation (REML) which is achieved by converting the semiparametric mixed-effect model to an equivalent linear mixed-effect model, and then combining the variances at the design points with the corresponding linear smoothing vector. The closed-form expressions for the uncertainty, including the bias and variances, as well as the approach to variance estimation are novel.

This paper adopts the following notation. Scalars and vectors are denoted by lower letters, and matrices by capital letters. A vector or a matrix is explicitly defined as it first appears. Script letters denote spaces, in particular, $\mathcal{R}^k$ denotes k-dimensional space of real numbers. The notation of $(\cdot)_i$ denotes the ith element of a vector. We use $\mathrm{diag}(x)$ to denote a $k \times k$ diagonal matrix whose diagonal vector is $x \in \mathcal{R}^k$, or $\mathrm{diag}(X)$ to denote a $k \times 1$ diagonal vector when $X \in \mathcal{R}^{k\times k}$. We use $0_{k\times l}$ ($1_{k\times l}$) to denote a k by l matrix of zeros (ones). The symbol $\otimes$ denotes the outer product and $\langle\cdot,\cdot\rangle$ inner product. The notation $N_k(m,\Sigma)$ denotes a multivariate Gaussian distribution with mean vector $m \in \mathcal{R}^k$ and variance matrix $\Sigma \in \mathcal{R}^{k\times k}$.

The rest of the paper is organized as follows. Relevant studies are briefly reviewed in Section 2. Section 3 reviews the semiparametric mixed-effect LS-SVM model. The expressions for confidence and prediction intervals are derived in Section 4. Numerical results showing the accuracy of derived expressions are presented in Section 5. Section 6 concludes this paper.

☆ This paper has been recommended for acceptance by G. Moser.

* Corresponding author. Tel.: +1 (618) 453 6056.
E-mail address: qcheng@cs.siu.edu (Q. Cheng).

0167-8655/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.12.010

2. Related work

Linear mixed-effect models are standard for analyzing longitudinal data, where the random effects are often estimated based on the theory of the best linear unbiased prediction (BLUP) [4]. For particular applications such as pharmaceutical experiments [3], nonlinear mixed effect models have been widely in use based on domain-specific models. Bootstrap based resampling methods for estimating the confidence and prediction intervals have been developed for statistical applications. The bootstrap-t procedure, also known as a percentile-t procedure, was first proposed by Efron [5]. The wild bootstrap was introduced for regression on non-clustered data [6], and later extended to a clustered setting [7]. The extended version is known as the wild cluster bootstrap-t procedure. Bootstrap procedures are simple yet effective; however, they provide little insight into how uncertainty depends on the variance components.

LS-SVM was first introduced as a kernel machine learning method alternative to SVM [8]. The confidence and prediction intervals for LS-SVM without random effects have been estimated analytically in [9]. Based on LS-SVM, a non-parametric mixed-effect LS-SVM model has recently been introduced in [1,2,10]. To our best knowledge, analytical expressions for confidence and prediction intervals, specific to the mixed effect LS-SVM model, have not been derived. This is the focus of the current paper.

3. Review of semiparametric mixed effect LS-SVM model

Assume there are $i = 1,\dots,N$ subjects or cases, and for each subject i there are $j = 1,\dots,n_i$ observations with response variables y_{ij}, fixed-effect covariate vectors $x_{ij} \in \mathcal{R}^p$ and random-effect covariate vectors $z_{ij} \in \mathcal{R}^q$. For a semiparametric model, let $x_{ij} = (x_{1ij}^t, x_{2ij}^t)^t$ be partitioned into two sub-vectors with $x_{1ij} \in \mathcal{R}^{p_1}$ corresponding to linear fixed effects and $x_{2ij} \in \mathcal{R}^{p_2}$ corresponding to nonlinear fixed effects. The following semiparametric mixed-effect LS-SVM model is considered [1,2]:

$$y_{ij} = b_0 + \beta^t x_{1ij} + w^t \phi(x_{2ij}) + b_i^t z_{ij} + \epsilon_{ij}, \quad j = 1,\dots,n_i; \quad i = 1,\dots,N, \tag{1}$$

where b_0 is the intercept term, $\beta \in \mathcal{R}^{p_1}$ is the regression parameter vector, $\phi(x_{2ij})$ is a nonlinear feature mapping function which corresponds to the nonparametric part of model, $b_i \sim N_q(0, B_i)$ is the random-effect parameter vector, and $\epsilon_i = (\epsilon_{i1},\dots,\epsilon_{in_i})^t \sim N_{n_i}(0, R_i)$ are error vectors. In this semiparametric mixed-effect model, covariance matrices B_i and R_i are assumed to be known or can be estimated (e.g., using REML).

To estimate the model (1), the following optimization problem is defined,

$$\min_{w,b_0,\beta,b_i,\epsilon_{ij}} \quad \frac{1}{2} w^t w + \frac{\lambda_1}{2} \sum_{i=1}^{N} b_i^t B_i^{-1} b_i + \frac{\lambda_2}{2} \sum_{i=1}^{N} \sum_{j,k=1}^{n_i} \epsilon_{ij}^t R_{ijk}^{-1} \epsilon_{ik}, \tag{2}$$

subject to the equality constraints given in Eq. (1) for all $j = 1,\dots,n_i$ and $i = 1,\dots,N$.

The Lagrangian function for (2) is given as by

$$L = \frac{1}{2} w^t w + \frac{\lambda_1}{2} \sum_{i=1}^{N} b_i^t B_i^{-1} b_i + \frac{\lambda_2}{2} \sum_{i=1}^{N} \sum_{j,k=1}^{n_i} \epsilon_{ij}^t R_{ijk}^{-1} \epsilon_{ik} + \sum_{i=1}^{N} \sum_{j=1}^{n_i} \alpha_{ij} (y_{ij} - b_0 - \beta^t x_{1ij} - w^t \phi(x_{2ij}) - b_i^t z_{ij} - \epsilon_{ij}), \tag{3}$$

where α_{ij} are Lagrange multipliers. By using the optimality conditions and the so-called "kernel trick" of $\langle \phi(x_{2ij}), \phi(x_{2kl}) \rangle = K(x_{2ij}, x_{2kl})$, the following system of linear equations can be obtained:

$$\begin{pmatrix} 0_{(p_1+1)\times(p_1+1)} & \bar{X}_1^t \\ \bar{X}_1 & \bar{W} \end{pmatrix} \begin{pmatrix} \bar{\beta} \\ \alpha \end{pmatrix} = \begin{pmatrix} 0_{(p_1+1)} \\ \bar{y} \end{pmatrix}, \tag{4}$$

here $\bar{\beta} = (b_0, \beta^t)^t, \alpha = (\alpha_1^t, \cdots, \alpha_N^t)^t$ with $\alpha_i = (\alpha_{i1}, \cdots, \alpha_{in_i})^t$, $\bar{y} = (y_1^t, \cdots, y_N^t)^t \in \mathcal{R}^{N_n}$ with $y_i = (y_{i1}, \cdots, y_{in_i})^t$ and $N_n = \sum_{k=1}^{N} n_k$, $\bar{X}_1 = (1_{N_n}, X_1)$

where $X_1 \in \mathcal{R}^{N_n \times p_1}$ represents the matrix obtained by stacking x_{1ij}^t's, and

$$W = K + \frac{1}{\lambda_1} \tilde{Z} \tilde{B} \tilde{Z}^t, \quad \bar{W} = W + \frac{1}{\lambda_2} \tilde{R}, \tag{5}$$

where $K \in \mathcal{R}^{N_n \times N_n}$ is the kernel matrix consisting of $K(x_{2ik}, x_{2il})$ with $k, l = 1,\dots,n_i$ and $i = 1,\dots,N$; $\tilde{Z} = \mathrm{diag}(Z_1,\dots,Z_N)$ is a $N_n \times qN$ block-diagonal matrix with $Z_i = (z_{i1},\dots,z_{in_i})^t$, $\tilde{B} = \mathrm{diag}(B_1,\dots,B_N)$ is a $qN \times qN$ block-diagonal matrix, $\tilde{R} = \mathrm{diag}(R_1,\dots,R_N)$ is a $N_n \times N_n$ block-diagonal matrix.

Now consider an input with covariates (x_*, z_*), which is either a training example or a new (test) example. By using the estimated values for $\hat{\bar{\beta}}$, $\hat{\alpha}$ from Eq. (4), the resulting mixed-effect LS-SVM regression equation is obtained [1,2]:

$$\hat{y}(x_*, z_*) = \hat{b}_0 + \hat{\beta}^t x_{1*} + \sum_{i=1}^{N} \left(\sum_{j=1}^{n_i} \hat{\alpha}_{ij} K(x_{2ij}, x_{2*}) + \hat{b}_i^t z_* \right), \tag{6}$$

where $\hat{b}_i = \frac{1}{\lambda_1} B_i Z_i^t \hat{\alpha}_i$.

4. Derivation of confidence and prediction intervals

At any given covariate vector $(x, z) \in \mathcal{R}^p \times \mathcal{R}^q$, after accounting for both fixed and random effects, we assume the response y is produced by the following model

$$y(x, z) = m(x, z) + \sigma(x, z)\epsilon, \tag{7}$$

where $E(\epsilon) = 0$, $Var(\epsilon) = 1$, and (x, z) and ϵ are statistically independent. Under mild regularity conditions ϵ can be modeled as a standard normal random variable thanks to the law of large numbers, but for derivations in this section, we consider ϵ with a general distribution. To derive the confidence intervals for the estimator $\hat{m}(x, z)$, we need to find a bound h_η such that

$$Prob(\sup_{x,z} |\hat{m}(x, z) - m(x, z)| \leqslant h_\eta) \geqslant 1 - \eta, \tag{8}$$

where $\eta \in (0, 1)$.

4.1. Bias estimator for semiparametric mixed-effect LS-SVM regression

To derive analytical equations of the confidence and prediction intervals, it is important to obtain the bias of the estimator and correct such a bias. Compared to the bootstrap procedures [11,12], using analytical formulas for confidence and prediction bands is computationally efficient and provides more insights into the uncertainty associated to the estimator. Extending the definition used in [9], we first show that the semiparametric mixed-effect LS-SVM regression is a linear smoother as follows.

Definition IV.1. (Linear Smoother). An estimator $\hat{m}$ of a mixed-effect regression function m is a linear smoother if, for each fixed-effect covariate $x \in \mathcal{R}^p$ and random-effect covariate $z \in \mathcal{R}^q$, there exists $L(x, z) = (l_1(x, z), \dots, l_{N_n}(x, z))^t$ such that

$$\hat{m}(x, z) = L^t(x, z)\bar{y}, \tag{9}$$

where $l_i(x, z) : \mathcal{R}^p \times \mathcal{R}^q \to \mathcal{R}$, and $\bar{y}$ is as defined above.