FAST M-TLS-RSVR AND M-LS- ε -TSVR FOR MULTI-OUTPUT REGRESSION

Chunhui Zhao and Liya Fan

School of Mathematics Sciences, Liaocheng University, Liaocheng, P. R. China

Abstract

This paper focuses on research multi-output regression problems and proposes two novel fast learning algorithms named as fast multi-output twin least squares regularized SVR (FM-TLS-RSVR) and fast multi-output least squares ε -TSVR (FM-LS- ε -TSVR), respectively. The main advantage of the proposed methods is to consider the cross relations among output vectors as a whole and avoid the singularity of matrices. In addition, the proposed FM-TLS-RSVR also possesses the sparsity. Experiment results indicate that FM-TLS-RSVR and FM-LS- ε -TSVR are two effective and competitive multi-output regressors.

Keywords: multi-output regression problem, least-squares technique, support vector machine, KKT conditions, regression accuracy.

*Corresponding author.

E-mail address: fanliya63@126.com (Liya Fan).

1. Introduction

We know that there are a plenty of multi-output (MO) regression problems in real-world applications, such as large tanker motion dynamics identification [1], hand movement trajectory decoding [2] and so on. However, there are no many methods for dealing with this kind of problems. Existing methods can be roughly divided into two categories. One is artificial neural networks and another is based on single-output support vector machine for regression (SVR). Classical SVR roots from statistical learning theory and seeks the best generalization performance of decision model by minimizing the training error, which has shown a better performance than other methods because its computational complexity does not depend on the dimensionality of the input space and can avoid multiple local extreme values [3]. Despite its potential advantages, the conventional formulation of SVR only has one dimensional output, which cannot directly deal with MO regression problems.

In order to deal with MO regression problems, Prez-Cruz et al. [4] developed M-SVR. But according to the experimental results in [5], M-SVR is somewhat sensitive to the perturbation of hyper-parameters in some small scale sample problems. In 2013, Zhang et al. [6] proposed a multiple single-output SVR (multiple S-SVR) method, which is to identify a single-output model for each output separately and then combines every individual model to produce a final MO model. Unfortunately, M-SVR and multiple S-SVR are all to ignore the cross relations among output variables and then decrease the accuracy of regression.

We all know that twin least squares regularized SVR (TLS-RSVR) [7] and ε -twin SVR (ε -TSVR) [8] are two commonly used single-output regressors and are all an extension and improvement of TSVR presented by Peng [9]. Different from only empirical risk minimization being implemented in TSVR, the structural risk minimization is also

implemented in ε -TSVR by introducing the regularization term in primal problems. This paper mainly studies the multi-output versions of TLS-RSVR and ε -TSVR and propose two novel fast MO regressors named as fast multi-output twin least squares regularized SVR (FM-TLS-RSVR) and fast multi-output least squares ε -TSVR (FM-LS- ε -TSVR), respectively. The main advantage of the proposed methods is to consider the cross relations among output vectors as a whole rather than separation and avoid the singularity of matrices. In addition, FM-TLS-RSVR also possesses the sparsity. In order to verify the effectiveness of the proposed regressors, a series of comparative experiments are performed on six multi-output data sets.

The rest of the paper is organized as follows. single-output TLS-RSVR and single-output ε -TSVR are briefly recalled in Section 2. FM-TLS-RSVR and FM-LS- ε -TSVR are presented in Section 3 and Section 4, respectively. A series of comparative experiments are performed in Section 5 and some conclusions are given in Section 6.

2. Preliminaries

This section briefly recalls single-output TLS-RSVR and single-output ε -TSVR, for details, see [7, 8]. Given a single-output data set $T_S = \{(x_i, y_i)\}_{i=1}^l \in R^d \times R$ and an MO data set $T_M = \{(x_i, y_i)\}_{i=1}^l \in R^d \times R^m$, where $x_i \in R^d$ is the i -th input sample and $y_i \in R$ and $y_i \in R^m$ are the output corresponding to x_i . We denote by $X = [x_1, \dots, x_l] \in R^{d \times l}$, $y = (y_1, \dots, y_l)^T \in R^l$ and $Y = [y_1, \dots, y_l]^T \in R^{l \times m}$ the sample matrix, the output vector and the output matrix, respectively. Let $e_l = (1, \dots, 1)^T \in R^l$ and $E = \text{repmat}(e_l, 1, m) \in R^{l \times m}$, where $\text{repmat}(A, m, n)$ denotes a large block matrix consisting of an $m \times n$ tiling of copies of A . Let $k : R^d \times R^d \rightarrow R$ be a kernel function with reproducing kernel Hilbert space (RKHS) H and nonlinear feature mapping

$\varphi : R^d \rightarrow H$. According to the properties of kernel function, it has $H = \text{span}\{\varphi(x_1), \dots, \varphi(x_l)\}$ and the kernel matrix $K = [k(x_i, x_j)] \in R^{l \times l}$. Let K_i denote the i -th column of K , $K_x = (k(x_1, x), \dots, k(x_l, x))^T$ and $Z = [\varphi(x_1), \dots, \varphi(x_l)]$, then $K = Z^T Z$.

For a matrix $A \in R^{p \times q}$ with $\text{rank}(A) = r \leq \min\{p, q\}$, its condensed singular value decomposition (SVD) is defined as $A = U_A \sum_A V_A^T$, where $U_A \in R^{p \times r}$ and $V_A \in R^{q \times r}$ are column orthogonal matrices, $\sum_A = \text{diag}(\sigma_1(A), \dots, \sigma_r(A))$ and $\sigma_1(A) \geq \dots \geq \sigma_r(A) > 0$ are all nonzero singular values of the matrix A . The Frobenius norm of the matrix A is defined by $\|A\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2(A)}$. The inner product of two same order matrices $A, B \in R^{p \times q}$ is defined by $\langle A, B \rangle = \text{tr}(A^T B)$, where $\text{tr}(\cdot)$ denotes the trace of a matrix. It is evident that $\|A\|_F^2 = \text{tr}(A^T A) = \langle A, A \rangle$.

2.1. Simple-output TLS-RSVR

Simple-output TLS-RSVR seeks a pair of nonparallel nonlinear up- and down-bound functions $f_1(x) = w_1^T \varphi(x) + b_1$ and $f_2(x) = w_2^T \varphi(x) + b_2$, where $w_1, w_2 \in H$ and $b_1, b_2 \in R$ are, respectively, coefficient vectors and thresholds, by considering the following two quadratic programming problems (QPPs):

$$\begin{aligned} \min_{w_1, b_1, \xi} \quad & \frac{1}{2} (\|w_1\|^2 + b_1^2) + \frac{c_1}{2} \|\xi_1\|^2 \\ \text{s.t.} \quad & (Z^T w_1 + b_1 e_l) - (y - \varepsilon_1 e_l) = \xi_1, \\ \min_{w_2, b_2, \eta} \quad & \frac{1}{2} (\|w_2\|^2 + b_2^2) + \frac{c_2}{2} \|\xi_2\|^2 \\ \text{s.t.} \quad & (y + \varepsilon_2 e_l) - (Z^T w_2 + b_2 e_l) = \xi_2, \end{aligned} \tag{1}$$

where $c_1, c_2, \varepsilon_1, \varepsilon_2 > 0$ are the users' parameters and $\xi_1, \xi_2 \in R^l$ are slack vectors. By translating the problems in (1) into unconstrained optimization problems and letting the derivatives of the objective functions with respect to (w_1, b_1) and w_2, b_2 are zeros, respectively, we can get the optimal (w_1^*, b_1^*) and (w_2^*, b_2^*) . Consequently, the regression function can be obtained by $f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}K_x^T(w_1^* + w_2^*) + \frac{1}{2}(b_1^* + b_2^*)$.

2.2. Single-output ε -TSVR

Single-output ε -TSVR seeks a pair of nonparallel nonlinear up- and down-bound functions $f_1(x) = w_1^T \phi(x) + b_1$ and $f_2(x) = w_2^T \phi(x) + b_2$, where $w_1, w_2 \in H$ and $b_1, b_2 \in R$ are respectively coefficient vectors and thresholds, by considering the following two QPPs:

$$\begin{aligned} \min_{w_1, b_1, \xi_1} \quad & \frac{c_1}{2} (\|w_1\|^2 + b_1^2) + \frac{1}{2} \|(y - \varepsilon_1 e_l) - (Z^T w_1 + b_1 e_l)\|^2 + c_2 e_l^T \xi_1 \\ \text{s.t.} \quad & (Z^T w_1 + b_1 e_l) - (y - \varepsilon_1 e_l) \geq \xi_1, \quad \xi_1 \geq 0, \\ \min_{w_2, b_2, \xi_2} \quad & \frac{c_3}{2} (\|w_2\|^2 + b_2^2) + \frac{1}{2} \|(y + \varepsilon_2 e_l) - (Z^T w_2 + b_2 e_l)\|^2 + c_4 e_l^T \xi_2 \\ \text{s.t.} \quad & (y + \varepsilon_2 e_l) - (Z^T w_2 + b_2 e_l) \geq \xi_2, \quad \xi_2 \geq 0, \end{aligned} \quad (2)$$

where $c_1, c_2, c_3, c_4, \varepsilon_1, \varepsilon_2 > 0$ are the users' parameters and $\xi_1, \xi_2 \in R^l$ are slack vectors. Different from translating the problems in (1) into unconstrained optimization problems, here the Wolfe dual forms of the problems in (2) are considered and then the optimal (w_1^*, b_1^*) and (w_2^*, b_2^*) can be gotten, respectively. Consequently, the regression function can be obtained by $f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}K_x^T(w_1^* + w_2^*) + \frac{1}{2}(b_1^* + b_2^*)$.

3. M-TLS-RSVR

This section tries to extend TLS-RSVR into multi-output version and propose a fast multi-output regression method named as FM-TLS-RSVR. For given MO data set T_M and kernel function $k : R^d \times R^d \rightarrow R$, M-TLS-RSVR hopes to seek a pair of nonparallel nonlinear up- and down-bound functions $f_1(x) = W_1^T \phi(x) + b_1$ and $f_2(x) = W_2^T \phi(x) + b_2$, where $W_j = [w_{j1}, \dots, w_{jm}]$ with $w_{ji} \in H$ and $b_j = (b_{j1}, \dots, b_{jm})^T \in R^m$ for $j = 1, 2$ are, respectively, weight matrices and threshold value vectors, by considering the following a pair of optimization questions:

$$\begin{aligned} \min_{W_1, b_1, \Xi_1} \quad & \frac{1}{2} (\|W_1\|_F^2 + \|b_1\|^2) + \frac{c_1}{2} \|\Xi_1\|_F^2 \\ \text{s.t.} \quad & (Z^T W_1 + B_1) - (Y - \varepsilon_1 E) = \Xi_1, \\ \min_{W_2, b_2, \Xi_2} \quad & \frac{1}{2} (\|W_2\|_F^2 + \|b_2\|^2) + \frac{c_2}{2} \|\Xi_2\|_F^2 \\ \text{s.t.} \quad & (Y + \varepsilon_2 E) - (Z^T W_2 + B_2) = \Xi_2, \end{aligned} \quad (3)$$

where $c_1, c_2, \varepsilon_1, \varepsilon_2 > 0$ are the users' parameters, $B_j = \text{repmat}(b_j^T, l, 1) \in R^{l \times m}$ for $j = 1, 2$ and $\Xi_j = [\xi_{j1}, \dots, \xi_{jm}] \in R^{l \times m}$ for $j = 1, 2$ are slack matrices. By solving the problems in (3), we can obtain the regression function $f(x) = \frac{1}{2} (f_1(x) + f_2(x))$. Next, before proposing FM-TLS-RSVR, we first introduce a classical method for solving the problems in (3).

3.1. Classical algorithm

The main idea of classical algorithm is to translate weight matrices W_1, W_2 into coefficient matrices and then solve the problems in (3) by considering the equivalent unconstrained optimization problems. Due to $H = \text{span}\{\phi(x_1), \dots, \phi(x_l)\}$ and $w_{ji} \in H$, there are coefficient matrices

$C_j \in R^{l \times m}$ such as $W_j = ZC_j$ for $j = 1, 2$. Consequently, $W_j^T W_j = C_j^T K C_j$, $Z^T W_j = K C_j$, $\|B_j\|_F^2 = \text{tr}(B_j B_j^T) = l \|b_j\|^2$, $f_1(x) = C_1^T K_x + b_1$, $f_2(x) = C_2^T K_x + b_2$ and the problems in (3) can be equivalently translated into the following unconstrained optimization problems:

$$\begin{aligned} \min_{C_1, B_1} F_1(C_1, B_1) &= \frac{1}{2} (\|ZC_1\|_F^2 + \frac{1}{l} \|B_1\|_F^2) + \frac{c_1}{2} \|(Y - \varepsilon_1 E) - (KC_1 + B_1)\|_F^2, \\ \min_{C_2, B_2} F_2(C_2, B_2) &= \frac{1}{2} (\|ZC_2\|_F^2 + \frac{1}{l} \|B_2\|_F^2) + \frac{c_2}{2} \|(Y + \varepsilon_2 E) - (KC_2 + B_2)\|_F^2. \end{aligned} \quad (4)$$

Letting $\partial F_1 / \partial C_1 = \partial F_1 / \partial B_1 = 0$ and $G = \begin{bmatrix} K \\ I_l \end{bmatrix} \in R^{2l \times l}$, we can deduce that

$$\begin{bmatrix} C_1^* \\ B_1^* \end{bmatrix} = \left(\begin{bmatrix} c_1^{-1} K & 0 \\ 0 & (lc_1)^{-1} I_l \end{bmatrix} + GG^T \right)^{-1} \begin{bmatrix} K(Y - \varepsilon_1 E) \\ Y - \varepsilon_1 E \end{bmatrix}, \quad (5)$$

where without loss of generality the symmetric nonnegative definite matrix $\begin{bmatrix} c_1^{-1} K & 0 \\ 0 & (lc_1)^{-1} I_l \end{bmatrix}$ is assumed to be nonsingular; otherwise, it can be regularized. With the similar way, we can also get

$$\begin{bmatrix} C_2^* \\ B_2^* \end{bmatrix} = \left(\begin{bmatrix} c_2^{-1} K & 0 \\ 0 & (lc_2)^{-1} I_l \end{bmatrix} + GG^T \right)^{-1} \begin{bmatrix} K(Y + \varepsilon_2 E) \\ Y + \varepsilon_2 E \end{bmatrix}. \quad (6)$$

By means of (5) and (6), the regression function can be gotten. The specific procedure is as follows.

Algorithm 1. Classical M-TLS-RSVR (CM-TLS-RSVR)

Step 1. Given an MO data set T_M and a proper kernel function $k : R^d \times R^d \rightarrow R$.

Step 2. Select suitable model parameters and kernel parameters.

Step 3. Calculate (C_1^*, B_1^*) and (C_2^*, B_2^*) by using (5) and (6), respectively.

Step 4. Calculate b_1^* and b_2^* , which are the first row of B_1^* and B_2^* , respectively.

Step 5. Construct regression function $f(x) = \frac{1}{2}(C_1^* + C_2^*)^T K_x + \frac{1}{2}(b_1^* + b_2^*)$.

From the models in (4), we can see that CM-TLS-RSVR loses the sparsity.

3.2. Fast algorithm

Different from classical algorithm, in order to preserve the sparsity, fast algorithm directly solves the problems in (3) by means of KKT conditions. To this end, consider the Lagrange function of the first problem in (3):

$$L(W_1, b_1, \Xi_1, A_1) = \frac{1}{2}(\|W_1\|_F^2 + \|b_1\|^2) + \frac{c_1}{2}\|\Xi_1\|_F^2 - \langle A_1, \Xi_1 - (Z^T W_1 + B_1) + (Y - \varepsilon_1 E) \rangle,$$

where $A_1 = [\alpha_{11}, \dots, \alpha_{1l}]^T \in R^{l \times m}$ is the Lagrangian multiplier matrix. Due to

$$\begin{aligned} \langle A_1, B_1 \rangle &= \text{tr}(A_1 B_1^T) = \text{tr}([\alpha_{11}, \dots, \alpha_{1l}]^T [b_1, \dots, b_l]) = \text{tr} \begin{bmatrix} a_{11}^T b_1 \\ \vdots \\ a_{1l}^T b_l \end{bmatrix} \\ &= \sum_{i=1}^l a_{1i}^T b_i = (\sum_{i=1}^l a_{1i}^T) b_1, \end{aligned}$$

the Lagrange function can be expressed as

$$L(W_1, b_1, \Xi_1, A_1) = \frac{1}{2} (\|W_1\|_F^2 + \|b_1\|^2) + \frac{c_1}{2} \|\Xi_1\|_F^2 - \langle A_1, \Xi_1 \rangle + \langle ZA_1, W_1 \rangle \\ + \left(\sum_{i=1}^l a_{1i}^T \right) b_1 - \langle A_1, (Y - \varepsilon_1 E) \rangle.$$

Letting $\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial \Xi_1} = \frac{\partial L}{\partial A_1} = 0$, we can deduce that $W_1 = -ZA_1$,

$f_1(x) = -A_1^T K_x + b_1$, $f_2(x) = -A_2^T K_x + b_2$, and

$$\begin{bmatrix} A_1^* \\ (b_1^*)^T \end{bmatrix} = \begin{bmatrix} e^T & 1 \\ c_1^{-1} I_l + K & -e \end{bmatrix}^{-1} \begin{bmatrix} 0_m^T \\ \varepsilon_1 E - Y \end{bmatrix}. \quad (7)$$

With the similar way, we can also get

$$\begin{bmatrix} A_2^* \\ (b_2^*)^T \end{bmatrix} = \begin{bmatrix} e^T & -1 \\ c_2^{-1} I_l + K & e \end{bmatrix}^{-1} \begin{bmatrix} 0_m^T \\ (Y + \varepsilon_2 E) \end{bmatrix}, \quad (8)$$

for the second problem in (3). Consequently, the regression function can be gotten. The specific procedure is as follows.

Algorithm 2. Fast M-TLS-RSVR (FM-TLS-RSVR)

Step 1. Given an MO data set T_M and a proper kernel function $k : R^d \times R^d \rightarrow R$.

Step 2. Select suitable model parameters and kernel parameters.

Step 3. Calculate (A_1^*, b_1^*) and (A_2^*, b_2^*) by using (7) and (8), respectively.

Step 5. Construct regression function $f(x) = -\frac{1}{2} (A_1^* + A_2^*)^T K_x + \frac{1}{2} (b_1^* + b_2^*)$.

Compared with CM-TLS-RSVR, we can see that FM-TLS-RSVR possesses the sparsity and avoid the singularity of matrices.

4. FM-LS- ε -TSVR

This section mainly extends ε -TSVR into multiple-output version and proposes another fast algorithm named as FM-LS- ε -TSVR. Unless special instructions, all symbols used in this section are same as in Section 3. M-LS- ε -TSVR hopes to seek a pair of nonparallel nonlinear up- and down-bound functions $f_1(x) = W_1^T \varphi(x) + b_1$ and $f_2(x) = W_2^T \varphi(x) + b_2$ by considering the following a pair of optimization problems:

$$\begin{aligned} \min_{W_1, b_1, \Xi_1} \quad & \frac{c_1}{2} (\|W_1\|_F^2 + \|b_1\|^2) + \frac{1}{2} \| (Y - \varepsilon_1 E) - (Z^T W_1 + B_1) \|_F^2 + c_2 \|\Xi_1\|_F^2 \\ \text{s.t.} \quad & (Z^T W_1 + B_1) - (Y - \varepsilon_1 E) = \Xi_1, \\ \min_{W_2, b_2, \Xi_2} \quad & \frac{c_3}{2} (\|W_2\|_F^2 + \|b_2\|^2) + \frac{1}{2} \| (Y + \varepsilon_2 E) - (Z^T W_2 + B_2) \|_F^2 + c_4 \|\Xi_2\|_F^2 \\ \text{s.t.} \quad & (Y + \varepsilon_2 E) - (Z^T W_2 + B_2) = \Xi_2. \end{aligned} \quad (9)$$

We can easily show that the classical algorithm for implementing M-LS- ε -TSVR is same as CM-TLS-RSVR. We only discuss the fast implementation algorithm in this section. With the similar way in Subsection 3.2, considering the Lagrangian function of the first problem in (9)

$$\begin{aligned} L(W_1, B_1, \Xi_1, A_1) = & \frac{c_1}{2} (\|W_1\|_F^2 + l^{-1} \|B_1\|_F^2) + \frac{c_2}{2} \|\Xi_1\|_F^2 + \frac{1}{2} \| (Z^T W_1 + B_1) \\ & - (Y - \varepsilon_1 E) \|_F^2 - \langle A_1, \Xi_1 - (Z^T W_1 + B_1) + (Y - \varepsilon_1 E) \rangle, \end{aligned}$$

and letting $\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial B_1} = \frac{\partial L}{\partial \Xi_1} = \frac{\partial L}{\partial A_1} = 0$, we can deduce that $W_1^* = ZM_1$

and

$$\begin{aligned} M_1 = & (K + (l + c_1(1 + c_2)^{-1})I_l)^{-1} (Y - \varepsilon_1 E), \quad B_1^* = l(K + (l + c_1(1 + c_2)^{-1})I_l)^{-1} \\ & (Y - \varepsilon_1 E). \end{aligned} \quad (10)$$

With the similar way, we can also deduce that $W_2^* = ZM_2$ and

$$M_2 = (K + (l + c_3(1 - c_4)^{-1})I_l)^{-1}(Y + \varepsilon_2 E), \quad B_1^* = l(K + (l + c_3(1 - c_4)^{-1})I_l)^{-1} \\ (Y + \varepsilon_2 E), \quad (11)$$

for the second problem in (9). By using (10) and (11), the regression function can be gotten. The specific procedure is as follows.

Algorithm 3. Fast M-LS- ε -TSVR (FM-LS- ε -TSVR)

Step 1. Given an MO data set T_M and a proper kernel function $k : R^d \times R^d \rightarrow R$.

Step 2. Select suitable model parameters and kernel parameters.

Step 3. Calculate (M_1, B_1^*) and (M_2, B_2^*) by using (10) and (11), respectively.

Step 4. Calculate b_1^* and b_2^* , which are the first row of B_1^* and B_2^* , respectively.

Step 5. Construct regression function $f(x) = \frac{1}{2}(M_1 + M_2)^T K_x + \frac{1}{2}(b_1^* + b_2^*)$.

From the above analysis, we can see that FM-TLS-RSVR avoid the singularity of matrices but does not possess the sparsity.

5. Experiments

In order to verify the effectiveness of the proposed FM-TLS-RSVR and FM-LS- ε -TSVR, a series of comparative experiments with CM-TLS-RSVR are performed on Corn (m5, mp5, and mp6) [10], regression, robustness to correlated noise and synthetic six multi-output data sets in this section.

Corn data set contains 700 data, each one contains 80 attributes measured by 3 different near-infrared spectrometers (m5, mp5, and mp6). All spectra are used respectively. Moisture, oil, protein, and starch represent four dependent output variables, where the first three variables are used by m5 and all of them by mp5 and mp6. 500 data are chosen randomly as training samples and the rest as testing samples in each experiment. Regression data set includes 600 data, which are generated by using the following three functions:

$$f_1(x) = \sum_{i=1}^m x_i^2, \quad f_2(x) = 20 - 20e^{-0.2\sqrt{\sum_{i=1}^m x_i/m}} - e^{\sum_{i=1}^m \cos(2\pi x_i)/m} + e,$$

$$f_3(x) = \sum_{i=1}^m (x_i^2 - 10 \cos(2\pi x_i) + 10),$$

for $x \in R^m$. In experiments, take $m = 3$ and $x \in [0, 1]^3$, and 300 data are chosen randomly as training samples and the rest as testing samples. In addition, the output values are normalized to $[-1, 1]$. Robustness to correlated noise data set includes 100 data, which are generated by using the following two functions:

$$\begin{cases} f_1(x) = 0.25[1.5(1 - x_1) + e^{2x_1-1} \sin(3\pi(x_1 - 0.6)^2) \\ \quad + e^{3(x_2-0.5)} \sin(4\pi(x_2 - 0.9)^2)] + 0.1, \\ f_2(x) = \sin(2\pi x_1) + 4(x_2 - 0.5)^2 + 0.1, \end{cases}$$

for $x \in [0, 1]^2$. data are chosen randomly as training samples and the rest as testing samples in each experiment. Synthetic data set contains 1000 noisy data, which are generated by using the following simulated two-output time series process:

$$\begin{cases} y_1(k) = 0.1 \sin(\pi y_2(k-1)) + (0.8 - 0.5 \exp(-y_1^2(k-1)))y_1(k-1) \\ \quad - (0.3 + 0.9 \exp(-y_1^2(k-1)))y_1(k-2) + \varepsilon_1(k), \\ y_2(k) = 0.6y_2(k-1) + 0.2y_2(k-1)y_2(k-2) + 1.2 \tanh(y_1(k-2)) + \varepsilon_2(k), \end{cases}$$

with the initial conditions $y_1(0) = y_1(-1) = y_2(0) = y_2(-1) = 0$, where the zero-mean Gaussian noise vector $\varepsilon(k) = (\varepsilon_1(k), \varepsilon_2(k))^T$ has a covariance matrix μI_2 . The data set $\{x(k), y(k)\}_{k=1}^{1000}$ are generated by $x(k) = (y_1(k-1), y_1(k-2), y_2(k-1), y_2(k-2))^T \in R^4$ and $y(k) = (y_1(k), y_2(k))^T \in R^2$ with $\mu = 0.01$. In each experiment, 500 data are selected randomly as training samples and the rest as testing samples.

All the predictive error involved is testing error, that is, the predictive error on testing sets. Each experiment result is the average result of ten times repeat experiments with different training samples and testing samples. In addition, Gaussian radial basis function (RBF) kernel $k(x_i, x_j) = \exp(-p\|x_i - x_j\|^2)$ with kernel parameter $p > 0$ is used in all experiments. The three indices are used for measuring the predictive error: root mean square error (RMSE) $= \sqrt{\frac{1}{l} \sum_{i=1}^l \|y_i - \hat{y}_i\|^2}$, mean absolute error (MAE) $= \frac{1}{l} \sum_{i=1}^l \frac{\|y_i - \hat{y}_i\|}{\|y_i + \hat{y}_i\|}$, and mean relative error (MRE) $= \frac{1}{l} \sum_{i=1}^l \frac{\|y_i - \hat{y}_i\|}{\|y_i\|}$, where y_i and $\hat{y}_i$ denote the true output and predicted output corresponding to the input x_i , respectively.

We know that the selection of parameters affects the performance of the algorithms. In order to facilitate comparison, the model parameters c_1, c_2, c_3, c_4 and $\varepsilon_1, \varepsilon_2$ and kernel parameter p are all selected by grid search method from 2^{-5} to 2^{15} , from 10^{-4} to 10^2 , and from 2^{-15} to 2^3 , respectively. The selected results are listed in Table 1. The experiment results with selected parameters are listed in Table 2. We can see from Table 2 that the learning time of FM-TLS-RSVR and FM-LS- ε -TSVR are all faster than that of CM-TLS-RSVR. For Corn(m5) and synthetic two data sets, FM-TLS-RSVR is significantly better than CM-TLS-RSVR and FM-LS- ε -TSVR on regression accuracy. But for Corn(mp5) and

regression two data sets, FM-LS- ε -TSVR is better than CM-TLS-RSVR and FM-TLS-RSVR. For Corn(mp6) data set, FM-LS- ε -TSVR is also better than CM-TLS-RSVR and FM-TLS-RSVR for RMAE and MMRE two indices. So, we can conclude that the proposed FM-TLS-RSVR and FM-LS- ε -TSVR are two effective and competitive multi-output regressors.

Table 1. Optimal selection of parameters

Data set	Algorithm	(c_1, c_2)	(c_3, c_4)	$(\varepsilon_1, \varepsilon_2)$	p
Corn(m5)	CM-TLS-RSVR	(0.001, 0.001)	—	(1, 1)	0.001
	FM-TLS-RSVR	(0.001, 0.001)	—	(0.001, 0.001)	0.001
	FM-LS- ε -TSVR	(0.1, 0.1)	(0.001, 0.001)	(0.01, 0.001)	0.01
Corn(mp5)	CM-TLS-RSVR	(0.001, 0.001)	—	(0.01, 0.01)	8
	FM-TLS-RSVR	(1, 1)	—	(1, 1)	8
	FM-LS- ε -TSVR	(0.001, 0.001)	(0.001, 0.001)	(0.1, 0.01)	0.0001
Corn(mp6)	CM-TLS-RSVR	(0.1, 0.001)	—	(0.1, 0.1)	8
	FM-TLS-RSVR	(0.0001, 0.01)	—	(0.001, 0.001)	0.001
	FM-LS- ε -TSVR	(0.0001, 0.0001)	(0.1, 0.1)	(0.1, 0.0001)	0.1
Regression	CM-TLS-RSVR	(0.001, 0.001)	—	(1, 1)	0.01
	FM-TLS-RSVR	(0.001, 0.001)	—	(0.001, 0.001)	0.001
	FM-LS- ε -TSVR	(0.001, 0.001)	(0.001, 0.001)	(0.001, 0.001)	0.01
Robustness	CM-TLS-RSVR	(0.001, 0.001)	—	(1, 1)	0.001
	FM-TLS-RSVR	(0.001, 0.001)	—	(0.1, 0.1)	8
	FM-LS- ε -TSVR	(0.001, 0.001)	(0.001, 0.001)	(0.001, 0.001)	0.01
Synthetic	CM-TLS-RSVR	(0.01, 0.01)	—	(0.001, 0.001)	1
	FM-TLS-RSVR	(0.01, 0.01)	—	(0.001, 0.001)	0.1
	FM-LS- ε -TSVR	(0.001, 0.001)	(0.001, 0.001)	(0.001, 0.001)	1

Table 2. Experiment results on 6 data sets

Data set	Algorithm	RME	MAE	MMRE	Time(s)
Corn(m5)	CM-TLS-RSVR	0.0064	0.0053	0.0081	1.5243
	FM-TLS-RSVR	0.0031	0.0022	0.0019	0.9557
	FM-LS- ϵ -TSVR	0.0106	0.0046	0.0220	0.9372
Corn(mp5)	CM-TLS-RSVR	0.0015	0.0036	0.0147	1.4496
	FM-TLS-RSVR	0.0099	0.0104	0.5775	0.9669
	FM-LS- ϵ -TSVR	9.34×10^{-5}	2.93×10^{-4}	7.11×10^{-5}	0.9261
Corn(mp6)	CM-TLS-RSVR	4.38×10^{-4}	0.0079	0.1338	1.4508
	FM-TLS-RSVR	1.19×10^{-4}	0.0034	0.0099	0.9480
	FM-LS- ϵ -TSVR	9.78×10^{-5}	0.0047	0.0067	0.9292
Regression	CM-TLS-RSVR	0.0365	0.0025	0.0099	0.4621
	FM-TLS-RSVR	0.0514	0.0150	0.0302	0.0165
	FM-LS- ϵ -TSVR	0.0075	5.97×10^{-4}	2.23×10^{-4}	0.3830
Robustness	CM-TLS-RSVR	0.0464	0.0143	0.0271	0.0329
	FM-TLS-RSVR	0.0320	0.0150	0.0304	0.0226
	FM-LS- ϵ -TSVR	0.0583	0.0028	0.0042	0.0273
Synthetic	CM-TLS-RSVR	0.0489	0.0156	0.0313	1.38694
	FM-TLS-RSVR	0.0129	0.0058	0.0132	0.9355
	FM-LS- ϵ -TSVR	0.0136	0.0166	0.0334	0.9116

6. Conclusion

Different from general single-output LS-SVR-type and LS-TSVR-type regressors translating primal problems into unconstrained optimization problems that may bring the singularity of matrices and then increase the learning time of regressors, FM-TLS-RSVR and FM-LS- ϵ -TSVR presented in this paper avoid the singularity problem by means of KKT conditions. In addition, FM-TLS-RSVR and FM-LS- ϵ -TSVR consider the relations among output vectors as a whole rather than separation. This

improve the regression accuracy of algorithms. We know that there are many single-output SVR-type and TSVR-type regressors, which are invalid for directly processing multi-output regression problems. By means of the similar way used in this paper, whether we can extend part of them to deal with multi-output problems will be our next work.

Acknowledgements

This work is supported by the Natural Science Foundation of Shandong Province (Grant No: ZR2016AM24) and the Natural Science Foundation of Liaocheng University (Grant No: 318011507).

References

- [1] Ning Wang, Min Han Nuo Dong and Meng Joo Er, Constructive multi-output extreme learning machine with application to large tanker motion dynamics identification, *Neurocomputing* 128 (2014), 59-72.
- [2] Pan Tian, Jie Hu, Jin Qi, Peng Xia and Ying-Hong Peng, Application of multi-output support vector regression on EMGs to decode hand continuous movement trajectory, *Bio-Medical Materials and Engineering* 26 (2015), 575-582.
- [3] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, 2000.
- [4] F. Prez-Cruz, G. Camps-Valls, E. Soria-Olivas et al., Multi-dimensional function approximation and regression estimation, *Artificial Neural Networks ICANN (2002)*, Springer, Berlin Heidelberg (2002), 757-762.
- [5] Wentao Mao, Jiucheng Xu, Chuan Wang and Longlei Dong, A fast and robust model selection algorithm for multi-input multi-output support vector machine, *Neurocomputing* 130 (2014), 10-19.
- [6] Li Zhang, Wei-Da Zhou, Pei-Chann Chang, Ji-Wen Yang and Fan-Zhang Li, Iterated time series prediction with multiple support vector regression models, *Neurocomputing* 99 (2013), 411-422.
- [7] Yongping Zhao, Jing Zhao and Min Zhao, Twin least squares support vector regression, *Neurocomputing* 118 (2013), 225-236.
- [8] Y. H. Shao, C. H. Zhang et al., An ε -twin support vector machine for regressor, *Neural Net.* 23 (2013), 175-185.

- [9] X. Peng, TSVR: An efficient twin support vector machine for regression, *Neural Networks* 23(3) (2010), 365-372.
- [10] S. Xu, X. An, X. Xiao, L. Zhu and L. Li, Multi-output least-squares support vector regression machines, *Pattern Recognition Letters* 34 (2013), 1078-1084.