MULTI-OUTPUT LEAST SQUARES TSVR

Min Hou and Liya Fan

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

_____

Abstract

In order to improve the regression accuracy of MLS-SVR presented by Xu et al. [10] with a fast training algorithm, this paper proposes a new regression method named as multi-output least squares twin SVR (MLS-TSVR) for multi-output problems. The main advantage of the proposed method is that the relations among output weight vectors are considered as a whole rather than decomposed in MLS-SVR. Experimental results show that MLS-TSVR has a good property on regression accuracy.

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

_____

1. Introduction

In many real-world applications, nonlinear black-box modelling based on machine learning is widely used as an effective soft-sensing technique. Especially, it generally needs to estimate and predict several variables or

_____

*Corresponding author.

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

targets in the fields of system identification and state estimation, such as large tanker motion dynamics identification [1], biophysical parameter evaluation from remote sensing images [2], hand movement trajectory decoding [3] and so on. In this scenario, systems output is a vector, which is called multi-output (MO) regression problem. It is of very important significance to improve the precision and speed of MO modelling.

Currently, there are no much results for directly researching MO regression problems. Existing methods can be roughly divided into two categories. One is feed forward neural networks, such as ELM and some variants [4-7], and another is based on single-output support vector regression machine (S-SVR), such as M-SVR [8] and multiple S-SVR model (multiple S-SVR) [9]. Unfortunately, M-SVR and multiple S-SVR are all to ignore the cross relations among output variables and then decrease the accuracy of regression. In 2013, Xu et al. [10] proposed multi-output least squares SVR (MLS-SVR).

We all know that single-output twin support vector regression machine (TSVR) presented by Peng [11] indirectly optimizes the regression function through the ε-insensitive down- and up-bound functions. This results in solving a pair of smaller sized quadratic programming problems (QPPs) instead of a larger QPP in SVR. Later, ε-TSVR is proposed by Shao et al. [12] based on TSVR for single-output regression problems. 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. Experiment results indicate that TSVR and ε-TSVR are all better than S-SVR on regression accuracy.

Motivated by works above, in this paper, we study the twin version of MLS-SVR by means of single-output TSVR and ε-TSVR and propose a novel multi-output regression method named as multi-output least squares twin SVR (MLS-TSVR). In order to verify the effectiveness of MLS-TSVR, a series of comparative experiments with MLS-SVR are

performed. The rest of the paper is organized as follows. Some basic notions and basic results are briefly recalled in Section 2. MLS-TSVR is proposed in Section 3 and comparative experiments on 9 data sets are performed in Section 4. Some conclusions are given in Section 5.

2. Preliminaries

This section briefly recalls TSVR, ε-TSVR and MLS-SVR, for details, see [10-12]. 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 outputs corresponding to x_i. Let $X = [x_1, \cdots, x_l] \in R^{d \times l}$ and $k : R^d \times R^d \to R$ be a kernel function with reproducing kernel Hilbert space (RKHS) H and nonlinear feature mapping $\varphi : R^d \to H$. According to the properties of kernel function, it has $H = \text{span}\{\varphi(x_1), \cdots, \varphi(x_l)\}$ and the kernel matrix $K = [k(x_i, x_j)] \in R^{l \times l}$.

For a matrix $A \in R^{p \times q}$ with $\text{rank}(A) = r$, 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, $\sum_A = \text{diag}(\sigma_1(A), \cdots, \sigma_r(A))$ and $\sigma_1(A) \geq \cdots \geq \sigma_r(A) > 0$ 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 $< A, B > = Tr(A^T B)$, where $Tr(\cdot)$ denotes the trace of a matrix. It is evident that $\|A\|_F^2 = Tr(A^T A) = < A, A >$.

For data set T_S, let $y = (y_1, \cdots, y_l)^T \in R^l$. Let $e_l = (1, \cdots, 1)^T \in R^l$, $K_x = [k(x, x_1), \cdots, k(x, x_l)]$ for $x \in R^l$ and I_l denotes the l order identity matrix. For data set T_M, let $Z = [\varphi(x_1), \cdots, \varphi(x_l)]$ and $Y = [y_1, \cdots, y_l]^T \in R^{l \times m}$, then the kernel matrix $K = Z^T Z$. Let repmat(A, m, n) denote a large block matrix consisting of an $m \times n$ tiling of copies of A.

2.1. TSVR

For data set T_S, TSVR seeks a pair of nonparallel nonlinear up- and down-bound functions $f_1(x) = K_x w_1 + b_1$ and $f_2(x) = K_x w_2 + b_2$, where $w_1, w_2 \in R^l$ are coefficient vectors and $b_1, b_2 \in R$ are thresholds, by considering the following two QPPs:

$$\min_{w_1, b_1, \xi} \frac{1}{2} \|y - e_l \varepsilon_1 - (Kw_1 + e_l b_1)\|^2 + c_1 e_l^T \xi$$

$$s.t. \ y - (Kw_1 + e_l b_1) \geq e_l \varepsilon_1 - \xi, \ \xi \geq 0,$$

$$\min_{w_2, b_2, \eta} \frac{1}{2} \|y + e_l \varepsilon_2 - (Kw_2 + e_l b_2)\|^2 + c_2 e_l^T \eta \tag{1}$$

$$s.t. \ (Kw_2 + e_l b_2) - y \geq e_l \varepsilon_2 - \eta, \ \eta \geq 0,$$

where $c_1, c_2, \varepsilon_1, \varepsilon_2 > 0$ are the users' parameters and $\xi, \eta \in R^l$ are slack vectors. By solving respectively the Wolfe dual forms of the problems in (1), we can get $[(w_1^*)^T \ b_1^*]^T = (H^T H)^{-1} H^T (f - \alpha^*)$ and $[(w_2^*)^T \ b_2^*]^T = (H^T H)^{-1} H^T (h + \beta^*)$, where $H = [K e] \in R^{l \times (l+1)}$, $f = y - e_l \varepsilon_1$, $h = y + e_l \varepsilon_2$ and α^* and β^* are respectively optimal solutions of the dual forms. Consequently, the regression function can be gotten by $f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2} K_x (w_1^* + w_2^*) + \frac{1}{2}(b_1^* + b_2^*)$.

2.2. ε-TSVR

Different from TSVR, ε-TSVR seeks a pair of nonparallel nonlinear up-and down-bound functions $f_1(x) = K_x w_1 + b_1$ and $f_2(x) = K_x w_2 + b_2$ by considering the following two QPPs:

$$
\min_{w_1, b_1, \xi} \frac{1}{2} c_3(\|w_1\|^2 + b_1^2) + \frac{1}{2} \|y - (Kw_1 + e_l b_1)\|^2 + c_1 e_l^T \xi
$$

$$
s.t. \ y - (Kw_1 + e_l b_1) \geq -e_l \varepsilon_1 - \xi, \ \xi \geq 0,
$$

$$
\min_{w_2, b_2, \eta} \frac{1}{2} c_4(\|w_2\|^2 + b_2^2) + \frac{1}{2} \|(Kw_2 + e_l b_2) - y\|^2 + c_2 e_l^T \eta \tag{2}
$$

$$
s.t. \ (Kw_2 + e_l b_2) - y \geq -e_l \varepsilon_2 - \eta, \ \eta \geq 0.
$$

Similar to TSVR, by solving the Wolfe dual forms of the problems in (2), we can get the regression function $f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2} K_x(w_1^* + w_2^*)$

$+ \frac{1}{2}(b_1^* + b_2^*)$, where $[(w_1^*)^T \ b_1^*]^T = (H^T H + c_3 I_{l+1})^{-1} H^T(y - \alpha^*)$,

$[(w_2^*)^T \ b_2^*]^T = (H^T H + c_4 I_{l+1})^{-1} H^T(y + \beta^*)$ and α^* and β^* are respectively optimal solutions of the dual forms.

2.3. MLS-SVR

For data set T_M, MLS-SVR seeks the regression function $f(x) = W^T \varphi(x) + b$, where $W = [w_1, \cdots, w_m]$ with $w_i \in H$ and $b = (b_1, \cdots, b_m)^T \in R^m$ are respectively unknown weighted matrix and bias vector, by considering the following optimization problem:

$$
\min_{W, b, \Xi} \frac{1}{2} \|W\|_F^2 + \frac{\gamma}{2} \|\Xi\|_F^2
$$

$$
s.t. \ Y = Z^T W + \text{repmat}(b^T, l, 1) + \Xi, \tag{3}
$$

where $\Xi \in R^{l \times m}$ is a slack matrix, $\gamma > 0$ is a user' parameter and repmat$(b^T, l, 1) = e_l b^T$. By decomposing each weight vector w_i into $w_0 + v_i$, the model (3) can be equivalently translated into

$$\min_{W_0, V, b, \Xi} \frac{1}{2} \|W_0 + V\|_F^2 + \frac{\gamma}{2} \|\Xi\|_F^2$$
$$s.t. \ Y = Z^T W_0 + Z^T V + e_l b^T + \Xi, \tag{4}$$

where $W_0 = [w_0, \cdots, w_0] = \text{repmat}(w_0, 1, m)$, $V = [v_1, \cdots, v_m]$ and $W = W_0 + V$. Due to $\|W_0\|_F^2 = m\|w_0\|^2$, $Tr(V^T W_0) = e_m^T V^T w_0$, and $\|W_0 + V\|_F^2 = m\|w_0\|^2 + 2 < V, w_0 e_m^T > + \|V\|_F^2$, the model (4) can be simplified as

$$\min_{w_0, V, b, \Xi} \frac{1}{2} \|w_0\|^2 + \frac{\lambda}{2m} \|V\|^2 + \frac{\gamma}{2} \|\Xi\|_F^2$$
$$s.t. \ Y = Z^T W_0 + Z^T V + e_l b^T + \Xi. \tag{5}$$

In model (5), there is a potential assumption that $w_0^T v_i = 0$ for all $i = 1, \cdots, m$, that is, w_0 is vertical to all v_i. In fact, this is not true in general. So, the model (5) is only an approximate model of the model (4) and may lead to a decline in the accuracy of regression. Let $A = [\alpha_1, \cdots, \alpha_m] \in R^{l \times m}$ be a Lagrange multipliers matrix of the model (5) and $\alpha = vec(A) \in R^{lm}$. By using KKT conditions, we can get $b^* = (P^T M^{-1} P)^{-1} P^T M^{-1} \widehat{y}$, $\alpha^* = M^{-1} \widehat{y} - M^{-1} P b^*$, where $P = \text{blockdiag}(e_l, \cdots, e_l) \in R^{ml \times m}$, $\Omega = \text{repmat}(K, m, m) \in R^{ml \times ml}$, $Q = \text{blockdiag}(K, \cdots, K) \in R^{ml \times ml}$, $M = \Omega + \frac{m}{\lambda} Q + \gamma^{-1} I_{ml} \in R^{ml \times ml}$ and $\widehat{y} = vec(Y) \in R^{lm}$. Consequently, the regression function can be obtained by using the following procedure:

Algorithm 1. MLS-SVR

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 vector $s = M^{-1} \widehat{y}$ and matrix $B = M^{-1} P$.

Step 4. Calculate matrix $C = (P^T B)^{-1}$.

Step 5. Calculate $b^* = CB^T \widehat{y}$ and $\alpha^* = s - Bb^*$.

Step 6. Calculate A^* by willing α^* matrix.

Step 7. Construct regression function $f(\widetilde{x})^T = K_{\widetilde{x}} \text{repmat}(A^* e_m, 1, m)$
$+ \frac{m}{\lambda} K_{\widetilde{x}} A^* + (b^*)^T$ for $\widetilde{x} \in R^d$.

3. MLS-TSVR

In this section, we discuss the twin version of MLS-SVR by means of single-output TSVR and ε-TSVR and propose a novel multi-output regression method MLS-TSVR.

The idea of MLS-TSVR is to seek a pair of nonparallel nonlinear up- and down-bound functions $f_1(x) = W_1^T \varphi(x) + b_1$ and $f_2(x) = W_2^T \varphi(x) + b_2$, where $W_j = [w_1^j, \cdots, w_m^j]$ with $w_i^j \in H$ and $bj = (b_1^j, \cdots, b_m^j)^T \in R^m$ are respectively unknown weighted matrix and bias vector for $j = 1, 2$, by considering the following two optimization problems:

$$
\begin{aligned}
&\min_{W_1, B_1, \Xi_1} \frac{1}{2} \|W_1\|_F^2 + \frac{\gamma_1}{2} \|\Xi_1\|_F^2 \\
&\text{s.t. } \Xi_1 = (Z^T W_1 + B_1) - (Y - \varepsilon_1 E), \\
&\min_{W_2, B_2, \Xi_2} \frac{1}{2} \|W_2\|_F^2 + \frac{\gamma_2}{2} \|\Xi_2\|_F^2 \\
&\text{s.t. } \Xi_2 = (Y + \varepsilon_2 E) - (Z^T W_2 + B_2),
\end{aligned}
\tag{6}
$$

where $E = \text{repmat}(e_l, m, 1) \in R^{l \times m}$ and $B_j = \text{repmat}(b_j^T, l, 1)$ for $j = 1, 2$. It is evident that the models in (6) can be translated into two unconstraint optimization problems:

$$\min_{W_1, B_1} F_1(W_1, B_1) = \frac{1}{2} \|W_1\|_F^2 + \frac{\gamma_1}{2} \|(Y - \varepsilon_1 E) - (Z^T W_1 + B_1)\|_F^2,$$

$$\min_{W_2, B_2} F_2(W_2, B_2) = \frac{1}{2} \|W_2\|_F^2 + \frac{\gamma_2}{2} \|(Y + \varepsilon_2 E) - (Z^T W_2 + B_2)\|_F^2. \tag{7}$$

Since $H = \text{span}\{\varphi(x_1), \cdots, \varphi(x_l)\}$, we can assume that $W_j = Z\Lambda_j$, where $\Lambda_j \in R^{l \times m}$ is coefficient matrix. Consequently, $\|W_j\|_F^2 = Tr(\Lambda_j^T K \Lambda_j)$, $Z^T W_j = K\Lambda_j$ for $= 1, 2$ and the up- and down-bound functions and the models in (7) can be respectively rewritten as $f_1(x)^T = K_x \Lambda_1 + b_1^T$, $f_2(x)^T = K_x \Lambda_2 + b_2^T$ and

$$\min_{\Lambda_1, B_1} F_1(\Lambda_1, B_1) = \frac{1}{2} \|Z\Lambda_1\|_F^2 + \frac{\gamma_1}{2} \|(Y - \varepsilon_1 E) - (K\Lambda_1 + B_1)\|_F^2,$$

$$\min_{\Lambda_2, B_2} F_2(\Lambda_2, B_2) = \frac{1}{2} \|Z\Lambda_2\|_F^2 + \frac{\gamma_2}{2} \|(Y + \varepsilon_2 E) - (K\Lambda_2 + B_2)\|_F^2. \tag{8}$$

Letting $\dfrac{\partial F_1}{\partial \Lambda_1} = \dfrac{\partial F_1}{\partial B_1} = \dfrac{\partial F_2}{\partial \Lambda_2} = \dfrac{\partial F_2}{\partial B_2} = 0$, we can deduce respectively that

$$\begin{bmatrix} \Lambda_1^* \\ B_1^* \end{bmatrix} = (\begin{bmatrix} \gamma_1^{-1}K & 0 \\ 0 & 0 \end{bmatrix} + GG^T)^{-1} \begin{bmatrix} K(Y - \varepsilon_1 E) \\ Y - \varepsilon_1 E \end{bmatrix}, \tag{9}$$

$$\begin{bmatrix} \Lambda_2^* \\ B_2^* \end{bmatrix} = (\begin{bmatrix} \gamma_2^{-1}K & 0 \\ 0 & 0 \end{bmatrix} + GG^T)^{-1} \begin{bmatrix} K(Y - \varepsilon_2 E) \\ Y - \varepsilon_2 E \end{bmatrix}, \tag{10}$$

where $G = [K, I_l] \in R^{2l \times l}$ and without loss of generality, the symmetric nonnegative definite matrices $\begin{bmatrix} \gamma_1^{-1}K & 0 \\ 0 & 0 \end{bmatrix} + GG^T$ and $\begin{bmatrix} \gamma_2^{-1}K & 0 \\ 0 & 0 \end{bmatrix} + GG^T$ are assumed as nonsingular; otherwise, they can be regularized. By using (9) and (10), the regression function can be gotten. The specific procedure is as follows:

Algorithm 2. MLS-TSVR

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

Step 2. Select suitable model parameters, regularization parameter and kernel parameters.

Step 3. Calculate matrices Λ_1^*, Λ_2^* and vectors b_1^*, b_2^* by using (9) and (10).

Step 4. Calculate the up- and down-bound functions $f_1(x)^T = K_x \Lambda_1^* + (b_1^*)^T$ and $f_2(x)^T = K_x \Lambda_2^* + (b_2^*)^T$.

Step 5. Construct the regression function $f(x) = \frac{1}{2}(f_1(x) + f_2(x))$ for $x \in R^d.$

Compared with MLS-SVR, MLS-TSVR does not need to assume $w_i = w_0 + v_i$ with $w_0^T v_i = 0$ for all $i = 1, \cdots, m$, which makes the models in (6) being exact models rather than approximate models.

4. Experiments

To verify the effectiveness of the proposed MLS-TSVR, in this section, a series of comparative experiments with MLS-SVR are performed on Corn (m5, mp5, and mp6) [10], synthetic [10] with parameter $c = 0.01$, 0.02, 0.03, 0.04 and Wind (Wind 1 and Wind 2) [13] nine MO data sets. The detailed characters of these data sets are listed in Table 1, where l, d, and m denote the number of data, the dimension of the input samples, and the dimension of the outputs, respectively.

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. Synthetic data set contains 1000 noisy data, which are generated by using a simulated two-output time series process as follows:

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

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 $x(k) = (y_1(k-1), y_1(k-2), y_2(k-1), y_2(k-2))^T$ and $y(k) = (y_1(k), y_2(k))^T$. In each experiment, 500 data are selected randomly as training samples and the rest as testing samples, and consider the cases with $c = 0.01$, 0.02, 0.03, and 0.04, respectively. Wind data set consists of 452 data, in which 300 data are selected randomly as training samples and the rest as testing samples in each experiment.

All the predictive error involved is testing error, that is, the predictive error on testing sets. All the experiments are implemented in MATLAB (R2010a) running on a PC with system configuration Intel (R) Core(TM) i3 (2.53GHz) with 2GB of RAM, and each of the results is the average result of ten times repeat experiments with different training samples and testing samples. In addition, Gaussian radial basis function (RBF) kernel $k(x_i, x_j) = \exp(-\dfrac{\|x_i - x_j\|^2}{2\sigma^2})$ with kernel parameter $\sigma > 0$ is used in all experiments.

The three indices are used for measuring the predictive error: root mean square error $(\text{RMSE}) = \sqrt{\frac{1}{l} \sum_{i=1}^{l} \| y_i - \widehat{y_i} \|}$, mean absolute error $(\text{MAE}) = \frac{1}{l} \sum_{i=1}^{l} \frac{\| y_i - \widehat{y_i} \|}{\| y_i + \widehat{y_i} \|}$, and mean relative error $(\text{MRE}) = \frac{1}{l} \sum_{i=1}^{l} \frac{\| y_i - \widehat{y_i} \|}{\| y_i \|}$, where y_i and $\widehat{y_i}$ 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 $\gamma, \gamma_1, \gamma_2$ and kernel parameter σ are all selected by grid search method from 10^{-3} to 10^3. The selected results are listed in Table 1.

Table 1. Optimal selection of parameters

Data set	Algorithm	γ	σ	γ_1	γ_2	ξ_1	ξ_2
Corn	MLS-SVR	10^2	10^{-1}	–	–	–	–
(m5)	MLS-TSVR	–	10^{-2}	10^2	10^0	10^{-2}	10^{-2}
Corn	MLS-SVR	10^3	10^{-3}	–	–	–	–
(mp5)	MLS-TSVR	–	10^{-2}	10^2	10^2	10^{-2}	10^{-2}
Corn	MLS-SVR	10^1	10^{-2}	–	–	–	–
(mp6)	MLS-TSVR	–	10^{-2}	10^2	10^2	10^{-1}	10^{-2}
Wind 1	MLS-SVR	10^2	10^{-2}	–	–	–	–
	MLS-TSVR	–	10^{-1}	10^{-2}	10^1	10^2	10^2
Wind 2	MLS-SVR	10^{-2}	10^{-2}	–	–	–	–
	MLS-TSVR	–	10^{-1}	10^0	10^1	10^{-2}	10^{-2}
Synthetic	MLS-SVR	10^2	10^{-3}	–	–	–	–
($c = 0.01$)	MLS-TSVR	–	10^0	10^{-1}	10^{-2}	10^{-2}	10^0
Synthetic	MLS-SVR	10^2	10^{-3}	–	–	–	–
($c = 0.02$)	MLS-TSVR	–	10^0	10^{-2}	10^{-2}	10^{-2}	10^1
Synthetic	MLS-SVR	10^2	10^{-1}	–	–	–	–
($c = 0.03$)	MLS-TSVR	–	10^0	10^{-2}	10^{-2}	10^1	10^1
Synthetic	MLS-SVR	10^2	10^{-2}	–	–	–	–
($c = 0.04$)	MLS-TSVR	–	10^0	10^{-2}	10^{-2}	10^1	10^2

The experiment results with selected parameters are listed in Table 2, from which we can see that MLS-TSVR is significantly better than MLS-SVR on the regression accuracy, especially for Wind 1 and Wind 2 on RMSE and MRE two indices and for synthetic with c = 0.02, 0.03, and 0.04 on MAE index. So, we can conclude that MLS-TSVR is an more effective and efficient regressor for multi-output regression problems.

Table 2. Comparison results on 9 data sets

Data set	Algorithm	RMSE	MAE	MRE
Corn	MLS-SVR	3.6804E-04	6.8021E-05	2.3801E-05
(m5)	MLS-TSVR	**2.4021E-04**	**1.2543E-05**	**1.1086E-05**
Corn	MLS-SVR	6.5366E-04	0.0012	0.0063
(mp5)	MLS-TSVR	**4.9198E-04**	**4.5230E-04**	**0.0027**
Corn	MLS-SVR	4.1710E-04	0.0021	0.3678
(mp6)	MLS-TSVR	**1.0730E-04**	**9.5166E-04**	**0.0022**
Wind 1	MLS-SVR	2.5507E+04	0.0025	3.5515E+05
	MLS-TSVR	**1.0113**	**0.0013**	**0.0023**
Wind 2	MLS-SVR	2.5507E+04	0.0025	6.5656E+06
	MLS-TSVR	**0.6181**	**0.0020**	**0.0038**
Synthetic	F-MLS-SVR	0.0152	0.0015	0.0042
($c = 0.01$)	M-TLS-SVR	**0.0123**	**0.0014**	**0.0027**
Synthetic	MLS-SVR	0.0400	0.0019	0.0050
($c = 0.02$)	MLS-TSVR	**0.0298**	**8.7039E-04**	**0.0012**
Synthetic	MLS-SVR	0.0321	0.0013	0.0019
($c = 0.03$)	MLS-TSVR	**0.0160**	**3.0239E-04**	**3.4617E-04**
Synthetic	MLS-SVR	0.0075	0.0015	0.0049
($c = 0.04$)	MLS-TSVR	**0.0011**	**2.4299E-05**	**8.3030E-06**

5. Conclusion

This paper presents a new multi-output regression method MLS-TSVR, which improves the regression accuracy of MLS-SVR by considering the relations among output weight vectors as a whole rather than decomposition forms. We know that there are many single-output TSVR-type regressors, which are invalid for directly processing MO regression problems. By means of the discussion method used in this paper, whether we can extend part of them to deal with MO problems will be our next work.

Acknowledgements

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]   D. Tuia, J. Verrelst, L. Alonso et al., Multioutput support vector regression for remote sensing biophysical parameter estimation, Geoscience and Remote Sensing Letters 8 (2011), 804-808.

[3]   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.

[4]   Junpeng Li, Changchun Hua, Yinggan Tang and Xinping Guan, A fast training algorithm for extreme learning machine based on matrix decomposition, Neurocomputing 173 (2016), 1951-1958.

[5]   Wentao Mao, Shengjie Zhao, Xiaoxia Mu and Haicheng Wang, Multi-dimensional extreme learning machine, Neurocomputing 149 (2015), 160-170.

[6]   P. Halko, N. Martinsson and J. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, SIAM Rev. 53(2) (2011), 217-288.

[7]   Alexandros Iosifidis, Anastasios Tefas and Ioannis Pitas, Regularized extreme learning machine for multi-view semi-supervised action recognition, Neurocomputing 145 (2014), 250-262.

[8]   F. Prez-Cruz, G. Camps-Valls, E. Soria-Olivas et al., Multi-dimensional function approximation and regression estimation, Artificial Neural Networks ІICANN (2002), Springer, Berlin Heidelberg, (2002), 757-762.

[9]   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.

[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.

[11]   X. Peng, TSVR: An efficient twin support vector machine for regression, Neural Networks 23 (2010), 365-372.

[12]   Y. H. Shao, C. H. Zhang et al., An ε-twin support vector machine for regressor, Neural Netw. 23 (2013), 175-185.

[13]   http://archive.ics.uci.edu/ml/datasets

■