MATEC Web of Conferences **128**, 02015 (2017) DOI: 10.1051/matecconf/201712802015
EITCE 2017

An Updated Projection Twin Support Vector Machine for Classification

Xiaopeng Hua[1,a] and Sen Xu[1]

[1]*School of Information and Engineering, Yancheng Institute of Technology, 224001 Yancheng, China*

Abstract. Based on projection twin support vector machine (PTSVM) and its extensions, this paper describes an updated PTSVM (UPTSVM) for classification. Compared with existing PTSVMs, UPTSVM has its own advantages. First, similar to the standard support vector machine (SVM), UPTSVM maintains the consistency of the optimization problems in the linear and nonlinear case, which results in the nonlinear formulations can be directly turned into the linear ones. Nevertheless, the existing PTSVMs lose the consistency because of using empirical kernel to construct nonlinear formulations. Second, UPTSVM avoids the inverse of kernel matrixes in the course of solving dual problems, which indicates it can not only reduce computing time but also save storage space. Third, UPTSVM can be practically proved equivalent to the PTSVM with regularization (RPTSVM). Experimental results on lots of data sets show the virtue of the presented method.

1 Introduction

Recently, proposed projection twin support vector machine (PTSVM) [1], which is on the base of multi-weight vector support vector machine (MVSVM) [2] and twin support vector machine (TWSVM) [3], has got the attention of increasing researchers. For binary classification problems, PTSVM has the intension of searching two projection axes same as MVSVM and formulates two small scale optimiaztion problems similar to TWSVM. Experimental results in [1] indicatie that PTSVM has comparable generalization ability than MVSVM and TWSVM in some aspects.

For the objective of improving the generalization ability of PTSVM furtherly, a method named as PTSVM with regularization term (RPTSVM) is proposed in [4]. Compared with PTSVM, RPTSVM has better classification capability because of considering the minimization of the structual risk similar to standard SVM [5]. For furthering the local learning ability of PTSVM, a weighted PTSVM (WPTSVM) [6] is proposed by adding the weight of each sample to the primal formulations of PTSVM. Experimental results validate WPTSVM can obtain better classification performance in dealing with manifold data sets. Shao et al. proposed a least squares PTSVM (LSPTSVM) [7], which can reducing the compiting time to a great extent by changing the inequality constrains in PTSVM to the equality ones and turning the L1 norm in the objective functions of PTSVM into L2 norm. Just like WPTSVM improves PTSVM, Hua et al. present a weighted LSPTSVM with local information, termed as (LIWLSPTSVM) [8], to improve the local learning ability of the LSPTSVM.

However, PTSVM and its extensions can not avoid the inverse of kernel matrixes in the course of training, which indicate they need more computing time and storage space in solving their dual problems. In addition, existing PTSVMs use empirical kernel to construct nonlinear formulations, which leads to the inconsistency of linear and nonlinear optimization problems. For the intention of overcome the above two defects, we propose an updated PTSVM (UPTSVM) in this paper.

The rest of this paper is organized as follows. In section 2 we overview RPTSVM briefly. Section 3 describes our new model, and section 4 shows the experiment results. In section 5 we give a summary of this paper.

2 Projection twin support vector machine with regularization term

Let us consider the classification problem with two classes of training samples, which are organized as a $l_1 \times n$ matrix $\boldsymbol{A}$ with positive samples (class 1) and a $l_2 \times n$ matrix $\boldsymbol{B}$ with negative samples (class 2). Define $\boldsymbol{E}=\boldsymbol{A}-\boldsymbol{e}_1\boldsymbol{e}_1^T\boldsymbol{A}/l_1$, $\boldsymbol{F}=\boldsymbol{B}-\boldsymbol{e}_2\boldsymbol{e}_1^T\boldsymbol{A}/l_1$, $\boldsymbol{G}=\boldsymbol{A}-\boldsymbol{e}_1\boldsymbol{e}_2^T\boldsymbol{B}/l_2$, and $\boldsymbol{H}=\boldsymbol{B}-\boldsymbol{e}_2\boldsymbol{e}_2^T\boldsymbol{B}/l_2$. For the linear classification problem, the formulations of the RPTSVM are

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{w}_1^T\boldsymbol{E}^T\boldsymbol{E}\boldsymbol{w}_1 + C_1\boldsymbol{e}_2^T\boldsymbol{\xi}_2 + \frac{C_2}{2}\|\boldsymbol{w}_1\|^2, \\ \text{s.t.} \quad & -\boldsymbol{F}\boldsymbol{w}_1 + \boldsymbol{\xi}_2 \ge \boldsymbol{e}_2, \quad \boldsymbol{\xi}_2 \ge 0, \end{aligned} \tag{1}$$

and

[a] Corresponding author: xp_hua@163.com

$$\min \quad \frac{1}{2}\boldsymbol{w}_2^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{w}_2 + C_3\boldsymbol{e}_1^T\boldsymbol{\xi}_1 + \frac{C_4}{2}\|\boldsymbol{w}_2\|^2, \qquad \text{s.t.} \quad \boldsymbol{G}\boldsymbol{w}_2 + \boldsymbol{\xi}_1 \ge \boldsymbol{e}_1, \quad \boldsymbol{\xi}_1 \ge 0, \tag{2}$$

where C_1~C_4 are positive parameters, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are slack vectors. The dual problems of Eqs. (1) and (2) are

$$\min \quad \frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{F}(C_2\boldsymbol{A}^T\boldsymbol{A}+\boldsymbol{I})^{-1}\boldsymbol{F}^T\boldsymbol{\alpha} - \boldsymbol{e}_2^T\boldsymbol{\alpha}, \qquad \text{s.t.} \quad \boldsymbol{0} \le \boldsymbol{\alpha} \le C_1\boldsymbol{e}_2. \tag{3}$$

and

$$\min \quad \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{G}(C_4\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{I})^{-1}\boldsymbol{G}^T\boldsymbol{\beta} - \boldsymbol{e}_1^T\boldsymbol{\beta}, \qquad \text{s.t.} \quad \boldsymbol{0} \le \boldsymbol{\beta} \le C_3\boldsymbol{e}_1. \tag{4}$$

Two projection axes are given by

$$\boldsymbol{w}_1 = (C_2\boldsymbol{E}^T\boldsymbol{E}+\boldsymbol{I})^{-1}\boldsymbol{F}^T\boldsymbol{\alpha}, \tag{5}$$

and

$$\boldsymbol{w}_2 = (C_4\boldsymbol{H}^T\boldsymbol{H}+\boldsymbol{I})^{-1}\boldsymbol{G}^T\boldsymbol{\beta}. \tag{6}$$

The decision function is

$$label(\boldsymbol{x}) = \underset{i=1,2}{\arg\min}\left(\left|\frac{\boldsymbol{w}_i^T}{\|\boldsymbol{w}_i\|}(\boldsymbol{x}-m_i)\right|\right), \tag{7}$$

where $\boldsymbol{x}$ is a new sample, m_i is the mean of the ith class samples, and $|\cdot|$ is the 1-norm.

3 Updated projection twin support vector machine (UPTSVM)

3.1. Linear UPTSVM

In the linear case, the optimization problems of UPTSVM are formulated as

$$\min \quad \frac{1}{2}\|\boldsymbol{w}_1\|^2 + \frac{C_2}{2}\boldsymbol{\eta}^T\boldsymbol{\eta} + C_1\boldsymbol{e}_1^T\boldsymbol{\xi}, \qquad \text{s.t.} \quad \boldsymbol{E}\boldsymbol{w}_1 \le \boldsymbol{\eta}, \quad -\boldsymbol{E}\boldsymbol{w}_1 \le \boldsymbol{\eta}, \quad -\boldsymbol{F}\boldsymbol{w}_1 \ge \boldsymbol{e}_2 - \boldsymbol{\xi}, \quad \boldsymbol{\xi} \ge \boldsymbol{0}, \tag{8}$$

and

$$\min \quad \frac{1}{2}\|\boldsymbol{w}_2\|^2 + \frac{C_4}{2}\tilde{\boldsymbol{\eta}}^T\tilde{\boldsymbol{\eta}} + C_3\boldsymbol{e}_2^T\tilde{\boldsymbol{\xi}}, \qquad \text{s.t.} \quad \boldsymbol{H}\boldsymbol{w}_2 \le \tilde{\boldsymbol{\eta}}, \quad -\boldsymbol{H}\boldsymbol{w}_2 \le \tilde{\boldsymbol{\eta}}, \quad \boldsymbol{G}\boldsymbol{w}_2 \ge \boldsymbol{e}_1 - \tilde{\boldsymbol{\xi}}, \quad \tilde{\boldsymbol{\xi}} \ge \boldsymbol{0}, \tag{9}$$

where $\boldsymbol{\xi}$, $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\xi}}$, $\tilde{\boldsymbol{\eta}}$ are slack vectors.

Consider the formulation (8). Compared with the formulation (1), the only change is taking the upper bounds $\boldsymbol{\eta}^T\boldsymbol{\eta}$ instead of $\boldsymbol{w}_1^T\boldsymbol{E}^T\boldsymbol{E}\boldsymbol{w}_1$, where $|\boldsymbol{E}\boldsymbol{w}_1| \le \boldsymbol{\eta}$. Obviously, we can easily prove the equivalence between the formulations (1) and (8). The formulation (9) has similar conclusion.

In order to gain the solutions of (8), we need to construct the Lagrangian function given by

$$L(\boldsymbol{w}_1,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\gamma}) = \frac{1}{2}\|\boldsymbol{w}_1\|^2 + \frac{C_2}{2}\boldsymbol{\eta}^T\boldsymbol{\eta} + C_1\boldsymbol{e}_1^T\boldsymbol{\xi} + \boldsymbol{\alpha}^T(\boldsymbol{E}\boldsymbol{w}_1 - \boldsymbol{\eta}) - \boldsymbol{\alpha}^{*T}(\boldsymbol{E}\boldsymbol{w}_1 + \boldsymbol{\eta}) + \boldsymbol{\beta}^T(\boldsymbol{F}\boldsymbol{w}_1 + e_2 - \boldsymbol{\xi}) - \boldsymbol{\gamma}^T\boldsymbol{\xi}, \tag{10}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^*$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are Lagrangian multipliers. Then the KKT conditions for (8) are given by

$$\frac{\partial L}{\partial \boldsymbol{w}_1} = \boldsymbol{0} \Rightarrow \boldsymbol{w}_1 = (\boldsymbol{\alpha}^* - \boldsymbol{\alpha})^T\boldsymbol{E} - \boldsymbol{\beta}^T\boldsymbol{F}, \tag{11}$$

$$\frac{\partial L}{\partial \boldsymbol{\eta}} = 0 \Rightarrow \boldsymbol{\eta} = \frac{\boldsymbol{\alpha}+\boldsymbol{\alpha}^*}{C_2}, \tag{12}$$

$$\frac{\partial L}{\partial \boldsymbol{\xi}} = 0 \Rightarrow \boldsymbol{0} \le \boldsymbol{\beta} \le C_1\boldsymbol{e}_2, \tag{13}$$

Substituting (11) and (12) into (10) and combining with (13) can generate the dual problem described as

$$\min \quad \frac{1}{2}(\boldsymbol{\alpha}^*-\boldsymbol{\alpha})^T(\boldsymbol{E}\boldsymbol{E}^T + \frac{1}{2C_2}\boldsymbol{I})(\boldsymbol{\alpha}^*-\boldsymbol{\alpha}) - (\boldsymbol{\alpha}^*-\boldsymbol{\alpha})^T\boldsymbol{E}\boldsymbol{F}^T\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{F}\boldsymbol{F}^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\boldsymbol{e}_2, \tag{14}$$

$$\text{s.t.} \quad \boldsymbol{\alpha}^*, \boldsymbol{\alpha} \ge \boldsymbol{0}, \quad \boldsymbol{0} \le \boldsymbol{\beta} \le C_1\boldsymbol{e}_2.$$

Furthermore, define

$$\pi = (\boldsymbol{\alpha}^{*T}, \boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T, \tag{15}$$

$$\kappa = (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{e}_2^T)^T, \tag{16}$$

$$C = (+\infty\boldsymbol{e}_1^T, +\infty\boldsymbol{e}_1^T, C_1\boldsymbol{e}_2^T)^T, \tag{17}$$

and

$$\boldsymbol{Q}_1 = (\boldsymbol{E}\boldsymbol{E}^T + \frac{1}{2C_2}\boldsymbol{I}), \ \boldsymbol{Q}_2 = \boldsymbol{E}\boldsymbol{F}^T, \ \boldsymbol{Q}_3 = \boldsymbol{F}\boldsymbol{F}^T, \tag{18}$$

and

$$\boldsymbol{M}_1 = \begin{pmatrix} \boldsymbol{Q}_1 & -\boldsymbol{Q}_1 \\ -\boldsymbol{Q}_1^T & \boldsymbol{Q}_1 \end{pmatrix}, \boldsymbol{M}_2 = \begin{pmatrix} -\boldsymbol{Q}_2 \\ \boldsymbol{Q}_2 \end{pmatrix}, \boldsymbol{M}_3 = \boldsymbol{Q}_3, \tag{19}$$

$$\Lambda = \begin{pmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{M}_2^T & \boldsymbol{M}_1 \end{pmatrix}, \tag{20}$$

the dual problem (14) can be reformulated as

$$\min \quad \frac{1}{2}\pi^T\Lambda\pi - \kappa^T\pi, \qquad \text{s.t.} \quad \boldsymbol{0} \le \pi \le C. \tag{21}$$

Similarly, the dual problem of (9) can be generated as

$$\min \quad \frac{1}{2}(\hat{\boldsymbol{\alpha}}^*-\hat{\boldsymbol{\alpha}})^T(\boldsymbol{H}\boldsymbol{H}^T + \frac{1}{2C_4}\boldsymbol{I})(\hat{\boldsymbol{\alpha}}^*-\hat{\boldsymbol{\alpha}}) - (\hat{\boldsymbol{\alpha}}^*-\hat{\boldsymbol{\alpha}})^T\boldsymbol{H}\boldsymbol{G}^T\hat{\boldsymbol{\beta}} + \frac{1}{2}\hat{\boldsymbol{\beta}}^T\boldsymbol{G}\boldsymbol{G}^T\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T\boldsymbol{e}_1, \tag{22}$$

$$\text{s.t.} \quad \hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\alpha}} \ge \boldsymbol{0}, \quad \boldsymbol{0} \le \hat{\boldsymbol{\beta}} \le C_3\boldsymbol{e}_2,$$

where $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\alpha}}^*$, and $\hat{\boldsymbol{\beta}}$ and are Lagrangian multipliers. It can also be described as

$$\min \quad \frac{1}{2}\hat{\pi}^T\hat{\Lambda}\hat{\pi} - \hat{\kappa}^T\hat{\pi}, \qquad \text{s.t.} \quad \boldsymbol{0} \le \hat{\pi} \le \hat{C}, \tag{23}$$

where

$$\hat{\pi} = (\hat{\boldsymbol{\alpha}}^{*T}, \hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\beta}}^T)^T, \tag{24}$$

$$\hat{\kappa} = (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{e}_1^T)^T, \tag{25}$$

$$\hat{C} = (+\infty\boldsymbol{e}_2^T, +\infty\boldsymbol{e}_2^T, C_3\boldsymbol{e}_1^T)^T, \tag{26}$$

and

$$\hat{\Lambda} = \begin{pmatrix} \hat{\boldsymbol{M}}_1 & \hat{\boldsymbol{M}}_2 \\ \hat{\boldsymbol{M}}_2^T & \hat{\boldsymbol{M}}_1 \end{pmatrix}, \tag{27}$$

$$\hat{\boldsymbol{M}}_1 = \begin{pmatrix} \hat{\boldsymbol{Q}}_1 & -\hat{\boldsymbol{Q}}_1 \\ -\hat{\boldsymbol{Q}}_1^T & \hat{\boldsymbol{Q}}_1 \end{pmatrix}, \hat{\boldsymbol{M}}_2 = \begin{pmatrix} \hat{\boldsymbol{Q}}_2 \\ -\hat{\boldsymbol{Q}}_2 \end{pmatrix}, \hat{\boldsymbol{M}}_3 = \hat{\boldsymbol{Q}}_3, \tag{28}$$

and

$$\hat{\boldsymbol{Q}}_1 = (\boldsymbol{HH}^T + \frac{1}{2C_4}\boldsymbol{I}),\ \hat{\boldsymbol{Q}}_2 = \boldsymbol{HG}^T,\ \hat{\boldsymbol{Q}}_3 = \boldsymbol{GG}^T. \quad (29)$$

After solving the dual problems (21) and (23), we can obtain the solutions of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. A new point $\boldsymbol{x}$ can be predicted according to

$$label(\boldsymbol{x}) = \arg\min_{i=1,2}(|\,\boldsymbol{w}_i^T(\boldsymbol{x} - m_i)\,|), \quad (30)$$

where m_i is the mean of the ith class samples.

3.2 Nonlinear UPTSVM

In order to solve linearly inseparable problems, we extend the linear UPTSVM to the nonlinear one. Different from existing PTSVMs, we directly formulate the nonlinear dual problems using the kernel function $K(\ ,\ )$ according to the linear formulations (14) and (22), which is similar to the standard SVM. The formulations of nonlinear UPTSVM are

$$\min\ \frac{1}{2}(\boldsymbol{\alpha}^* - \boldsymbol{\alpha})^T(K(\boldsymbol{E},\boldsymbol{E}) + \frac{1}{2C_2}\boldsymbol{I})(\boldsymbol{\alpha}^* - \boldsymbol{\alpha})$$
$$-(\boldsymbol{\alpha}^* - \boldsymbol{\alpha})^T K(\boldsymbol{E},\boldsymbol{F})\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}^T K(\boldsymbol{F},\boldsymbol{F})\boldsymbol{\beta} - \boldsymbol{\beta}^T \boldsymbol{e}_2, \quad (31)$$
$$\text{s.t.}\ \ \boldsymbol{\alpha}^*, \boldsymbol{\alpha} \geq \boldsymbol{0},\ \ \boldsymbol{0} \leq \boldsymbol{\beta} \leq C_1\boldsymbol{e}_2,$$

and

$$\min\ \frac{1}{2}(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})^T(K(\boldsymbol{H},\boldsymbol{H}) + \frac{1}{2C_4}\boldsymbol{I})(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})$$
$$-(\hat{\boldsymbol{\alpha}}^* - \hat{\boldsymbol{\alpha}})^T K(\boldsymbol{H},\boldsymbol{G})\hat{\boldsymbol{\beta}} + \frac{1}{2}\hat{\boldsymbol{\beta}}^T K(\boldsymbol{G},\boldsymbol{G})\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T \boldsymbol{e}_1, \quad (32)$$
$$\text{s.t.}\ \ \hat{\boldsymbol{\alpha}}^*, \hat{\boldsymbol{\alpha}} \geq \boldsymbol{0},\ \ \boldsymbol{0} \leq \hat{\boldsymbol{\beta}} \leq C_3\boldsymbol{e}_2.$$

3.3 Comparision of UPTSVM with RPTSVM

Consider the linear UPTSVM. Compared with the dual problems (3) and (4) in RPTSVM, the dual problems (14) and (22) in UPTSVM avoid the inverse of kernel matrixes in the course of training, which can not only reduce computing time but also save storage space. The nonlinear UPTSVM has the same advantage.

Comparing the dual formulations in nonlinear UPTSVM with ones in linear UPTSVM, we can find that the only difference is the kernel function $K(\ ,\ )$ taken instead of inner product. That is to say UPTSVM maintains the consistency of linear and nonlinear primal formulations in addition to using different kernel functions. However, RPTSVM use empirical kernel to construct nonlinear formulations [4], which leads to the inconsistency of linear and nonlinear optimization problems.

4 Experimental results

For the intention to compare the performance of our NPTSVM with PTSVM and RPTSVM, we conduct experiments on lots of standard datasets used in [13], [14], [15]. All of the classification methods are implemented on a computer with Matlab 7.0. The computer is equipped with Intel P4 processors (2.3 GHz) and 2 GB RAM. For the sake of brevity, C_1=C_2 is set for PTSVM, and C_1=C_3 and C_2=C_4 are set for RPTSVM and our UPTSVM. For the parameter values, we select them from the range $\{2^i|i= -8,-6,\ldots,+8\}$. Table 1 shows the average 10-fold cross-validation results. It discloses that the classification ability of our UPTSVM is comparable than PTSVM and RPTSVM. However, UPTSVM has higher computing complexity, although it does not need to compute the large inverse matrices. This is mainly because UPTSVM has more variables (number of $2l_1+l_2$ or $2l_2+l_1$) than PTSVM and RPTSVM (number of l_1 or l_2) during the learning process.

We also conduct experiments on benchmark datasets in nonlinear case. The RBF kernel exp(-$\|x_i$-$x_j\|$2/ δ) is choosed for all of the methods, the value of δ is selected from the set $\{2^i|i= -1,0,\ldots,+7\}$. Table 2 lists experimental results in nonlinear case. It reveals that the generalization ability of our UPTSVM is better than PTSVM and RPTSVM because of adopting different kernel tricks in constructing nonlinear formulations. It also disclose that UPTSVM is little slower than RPTSVM and PTSVM in terms of training time.

Table 1. Experimental results in linear case.

Datasets	PTSVM Accuracy % Time (s)	RPTSVM Accuracy % Time (s)	UPTSVM Accuracy % Time (s)
hepatitis (155×19)	83.88±6.75 0.0094	84.00±13.06 0.0078	**84.17±12.59** 0.0562
sonar (208×60)	**68.71±14.31** 0.014	67.07±13.42 0.0125	65.71±13.72 0.1123
haberman (306×3)	**74.94±5.48** 0.0265	73.94±5.38 0.0281	74.83±4.25 0.3152
Cleve (296×13)	**82.48±3.87** 0.0218	82.42±5.66 0.0187	82.48±4.16 0.2621
spectf (267×44)	80.05±3.64 0.0281	76.50±12.82 0.0234	**80.51±6.92** 0.2246
glass_7 (214×10)	87.77±6.71 0.0156	**96.77±9.71** 0.0172	91.09±12..56 0.1264
heart (270×13)	82.96±4.74 0.0109	84.82±5.35 0.0172	**84.82±1.99** 0.2059
p_gene (106×57)	62.50±11.24 0.0047	**77.88±10.87** 0.0078	**77.88±10.87** 0.0218
wpbc (198×33)	77.51±6.28 0.0156	77.78±5.67 0.0249	**78.93±7.48** 0.1014
monks3 (432×6)	76.93±9.93 0.0312	80.95±14.75 0.0437	**81.35±11.43** 0.8159
parkinsons (195×22)	**83.59±8.34** 0.0125	83.59±12.24 0.0140	82.73±9.56 0.0952
spect (267×22)	80.87±11.15 0.0250	78.26±17.71 0.0234	**82.88±10.96** 0.2231
votes (435×16)	95.79±2.64 0.0499	**96.03±2.47** 0.0452	94.90±3.07 0.8424
bupa_liver (345×6)	59.40±7.45 0.0265	64.06±4.51 0.0265	**64.28±7.51** 0.4290
australian (690×14)	86.28±5.53 0.1404	86.11±5.72 0.1357	**86.81±4.74** 2.9453
breast (699×9)	96.38±2.92 0.1716	96.23±3.68 0.1700	**97.12±2.26** 3.1559

Table 2. Experimental results in nonlinear case.

Datasets	**PTSVM Accuracy % Time (s)**	**RPTSVM Accuracy % Time (s)**	**UPTSVM Accuracy % Time (s)**
hepatitis (155×19)	83.33±4.47 0.1701	85.17±10.23 0.1607	**85.67±7.61** 0.1825
sonar (208×60)	**73.29±16.95** 0.6536	69.07±20.32 0.6552	70.86±18.13 0.7098
haberman (306×3)	74.56±3.77 0.6427	74.83±5.19 0.6443	**75.17±2.63** 0.8097
Cleve (296×13)	82.48±4.44 0.6178	**84.78±6.00** 0.6131	83.57±4.77 0.7316
spectf (267×44)	80.96±7.81 1.0015	83.18±8.08 1.0218	**83.35±6.80** 1.1248
glass_7 (214×10)	95.32±4.95 0.3011	95.71±5.12 0.312	**96.91±3.38** 0.3713
heart (270×13)	82.22±7.55 0.5008	**84.44±4.32** 0.5101	84.44±4.63 0.5881
p_gene (106×57)	77.13±11.04 0.1591	78.13±14.02 0.1701	**83.88±11.30** 0.1778
wpbc (198×33)	82.83±7.11 0.2746	80.19±6.68 0.2683	**83.41±5.00** 0.3214
monks3 (432×6)	**98.68±2.95** 1.3541	93.83±5.02 1.3853	94.97±3.78 1.7893
parkinsons (195×22)	90.25±11.79 0.2558	89.70±11.71 0.2589	**92.53±9.24** 0.3089
spect (267×22)	84.25±8.33 0.5039	**85.49±5.83** 0.5023	85.49±6.55 0.6053
votes (435×16)	94.6±3.62 1.4227	95.08±2.80 1.4056	**95.79±2.17** 1.8798

5 Conclusions

In this paper, an updated PTSVM (UPTSVM) is presented for improving the performance of existing PTSVMs. Compared with existing PTSVMs, UPTSVM avoids the inverse of kernel matrixes in the course of training and maintains the consistency of linear and nonlinear primal problems. In addition, UPTSVM can be practically proved equivalent to RPTSVM. However, although UPTSVM does not have to compute the large inverse matrices as existing PTSVMs do, it has higher computing complexity. So improving the computational efficiency is our further research goal in the future.

Acknowledgment

This work is partly funded by Jiangsu Provincial Science and Technology Support Project of China (Nos. BY2016065-05, BY2016065-01).

References

1. X. Chen, J. Yang, Q. Ye, J. Liang, Pattern Recogn. **44** 2643 (2011)
2. Q. Ye, C. Zhao, N. Ye, Y. Chen, Pattern Recogn. **31** 2006 (2010)
3. Jayadeva, R. Khemchandai, S. Chandra, IEEE Trans. Pattern Anal. Mach. Intell. **29** 905 (2007)
4. Y.H. Shao, Z. Wang, W.J. Chen, N.Y. Deng, Knowl-Based Syst. **37** 203 (2013)
5. C. Cortes, V.N. Vapnik, Mach. Learn. **20**, 273 (1995)
6. X. Hua, Y. Sun, S. Ding, CAAI Trans. Intell. Syst. **11** 384 (2016)
7. Y. Shao, N. Deng, Z. Yang, Pattern Recogn. **45** 2299 (2012)
8. X. Hua, S. Ding, Neurocomputing **160** 228 (2015)
9. Y. Xu, Z. Yang, X. Pan, IEEE Trans. Neural Netw. and Learn. Syst. **28** 359 (2017)
10. M. Carrasco, J. Lopez, S. Maldonado, Expert Syst. Appl. **54** 95 (2016)
11. X. Peng, D. Xu, L. Kong, D. Chen, Inform. Sci. **340/341** 86 (2016)
12. R. Khemchandani, P. Saigal, S. Chandra, Neural Netw. **79** 97 (2016)
13. F. Alamdar, S. Ghane, A. Amiri, Neurocomputing **186** 8 (2016)
14. F. Dufrenois, J.C. Noyer, Pattern Recogn. **52** 96 (2016)
15. H. Wang, Z. Zhou, Knowl-Based Syst. **128** 125 (2017)