Large margin proximal non-parallel support vector classifiers*

Mingzeng Liu¹ and Yuanhai Shao^{*2}

 ¹ School of Mathematics and Physics Science, Dalian University of Technology at Panjin, Dalian, 124221, P.R.China mzliu@dlut.edu.cn
 ² School of Economics and Management, Hainan University, Haikou, 570228, P.R.China shaoyuanhai21@163.com

Abstract. In this paper, we propose a novel large margin proximal non-parallel twin support vector machine for binary classification. The significant advantages over twin support vector machine are that the structural risk minimization principle is implemented and by adopting uncommon constraint formulation for the primal problem, the proposed method avoids the computation of the large inverse matrices before training which is inevitable in the formulation of twin support vector machine. In addition, the dual coordinate descend algorithm is used to solve the optimization problems to accelerate the training efficiency. Experimental results exhibit the effectiveness and the classification accuracy of the proposed method.

Keywords: Support vector machine \cdot Non-parallel SVM \cdot Structural risk minimization principle \cdot Dual coordinate descend algorithm.

1 Introduction

As a powerfull tool in machine learning, support vector machines (SVMs) have been gained a greate deal of attention in wide variety of fields [1–5]. The classical support vector classifier (SVC) is trying to maximize the margin between two parallel hyperplanes, which results in solving a convex quadratic programming problem (QPP). Furthermore, some non-parallel hyperplane classifiers such as the generalized eigen-value proximal support vector machine (GEPSVM) and twin support vector machine (TWSVM) have been proposed in [4] and [5]. TWSVM is to search two non-parallel proximal hyperplanes such that each hyperplane is closer to one class and as far as possible from the other one. In fact, as an efficient generalization of the classical SVC, TWSVMs have been studied extensively [6–14], which need to solve two samll QPPs in contrast with the

^{*} Corresponding author: shaoyuanhai21@163.com. Supported by the Hainan Provincial Natural Science Foundation of China (No.118QN181), the Scientific Research Foundation of Hainan University, the National Natural Science Foundation of China (Nos. 11501310, 61703370, and 61603338).

classical SVC. This paper proposes a novel large margin proximal non-parallel support vector machine for binary classification (PNSVM). The main contributions of PNSVM are:

- PNSVM minimizes the structural risk by imposing a regularization term.
- PNSVM is proximal both classes and maximizes the corresponding margin, while TWSVM is proximal each class and far away from the other class.
- PNSVM can be solved efficiently with dual coordinate descend method [15].

2 Related Work

Benefiting from its excellent generalization performance of twin support vector machines, the approaches of constructing the non-parallel hyperplanes have received extensive attention [6, 8, 9, 11–14, 16]. Shao et al. [6] present a variant of GEPSVM based on the difference measure, which seems to be superior in classification accuracy and computation time. For the imbalanced data classification, Shao et al. [8] suggest an efficient weighted Lagrangian twin support vector machine (WLTSVM), which is robust to ouliers and overcomes the bias phenomenon in the original TWSVM. Liu et al. [9] present a support vector machine for large scael regression based on the minimization of deviation distribution. Tian et al. [11] propose a novel nonparallel SVM for binary classification, which implements the structural risk minimization and is suitable for large scale problems. Shao et al. [16] present a sparse L_q -norm least squares support vector machine, where feature selection and prediction are performed simultaneously.

3 PNSVM

3.1 Linear PNSVM

Linear PNSVM is to find two non-parallel hyperplanes $f_1(\boldsymbol{x}) = \boldsymbol{w}_1^T \boldsymbol{x} + b_1 = 0$ and $f_2(\boldsymbol{x}) = \boldsymbol{w}_2^T \boldsymbol{x} + b_2 = 0$ by solving the following two problems:

$$\min_{\boldsymbol{w}_{1}, b_{1}, \boldsymbol{p}, \boldsymbol{q}_{1}, \rho} \frac{\frac{1}{2} \boldsymbol{p}^{T} \boldsymbol{p} + \frac{1}{2} c_{1} (\|\boldsymbol{w}_{1}\|^{2} + b_{1}^{2}) + c_{2} \boldsymbol{e}_{2}^{T} \boldsymbol{q}_{1} + c_{3} \rho_{1}}{s.t.} \qquad s.t. \qquad \boldsymbol{A} \boldsymbol{w}_{1} + \boldsymbol{e}_{1} b_{1} = \boldsymbol{p}, \qquad (1)$$

$$\rho_{1} \boldsymbol{e}_{2} \geq -(\boldsymbol{B} \boldsymbol{w}_{1} + \boldsymbol{e}_{2} b_{1}) + \boldsymbol{q}_{1} \geq \boldsymbol{e}_{2}, \ \boldsymbol{q}_{1} \geq 0, \quad \rho_{1} \geq 1,$$

and

where $c_i, i = 1, 2, \dots, 6$ are parameters.

The formulation given at (1) can be understood as follows: The first term in (1) is the sum of squared distances from $f(\mathbf{x}) = \mathbf{w}_1 \mathbf{x} + b_1 = 0$ to points of

positive class, whose minimization means to keep $f(\mathbf{x}) = \mathbf{w}_1 \mathbf{x} + b_1 = 0$ close to the positive class. The minimization of the second term in (1) implies that the structural risk principle is implemented by the term $\frac{1}{2}c_2(||\mathbf{w}_1^2 + b_1^2||)$. The second constraints in (1) require the hyperplane $f(\mathbf{x}) = \mathbf{w}_1^T \mathbf{x} + b_1$ to be at a distance of at least 1 and at most ρ_1 from points of negative class. The third term in (1) tries to minimize mis-classification. The last term in (1) requires the points of negative class to be a distance no far away from the hyperplane. A similar interpretation may also be given to the formulation given at (2).

The Lagrangian of the formulation (1) is given by

$$\mathbf{L}(\boldsymbol{w}_{1}, b_{1}, \boldsymbol{p}, \boldsymbol{q}_{1}, \rho_{1}; \boldsymbol{\alpha}_{1}, \boldsymbol{\alpha}_{2}, \boldsymbol{\alpha}_{3}, \boldsymbol{\alpha}_{4}, \beta) \\ = \frac{1}{2}c_{1}(\|\boldsymbol{w}_{1}\|^{2} + b_{1}^{2}) + \frac{1}{2}\boldsymbol{p}^{T}\boldsymbol{p} + c_{2}\boldsymbol{e}_{2}^{T}\boldsymbol{q}_{1} + c_{3}\rho_{1} + \boldsymbol{\alpha}_{1}^{T}(\boldsymbol{A}\boldsymbol{w}_{1} + \boldsymbol{e}_{1}b_{1} - \boldsymbol{p}) \\ - \boldsymbol{\alpha}_{2}^{T}(-(\boldsymbol{B}\boldsymbol{w}_{1} + \boldsymbol{e}_{2}b_{1}) + \boldsymbol{q}_{1} - \boldsymbol{e}_{2}) - \boldsymbol{\alpha}_{3}^{T}(\rho_{1}\boldsymbol{e}_{2} + (\boldsymbol{B}\boldsymbol{w}_{1} + \boldsymbol{e}_{2}b_{1}) - \boldsymbol{q}_{1}) \\ - \boldsymbol{\alpha}_{4}^{T}\boldsymbol{q}_{1} - \beta(\rho_{1} - 1)$$

where $\boldsymbol{\alpha}_1 \in \mathbf{R}^{m_1 \times 1}, \boldsymbol{\alpha}_2 \in \mathbf{R}^{m_2 \times 1}, \boldsymbol{\alpha}_3 \in \mathbf{R}^{m_2 \times 1}, \boldsymbol{\alpha}_4 \in \mathbf{R}^{m_1 \times 1}, \beta \in \mathbf{R}$ are the vectors of Lagrange multipliers. The optimality conditions for $\boldsymbol{w}_1, \boldsymbol{b}_1, \boldsymbol{p}, \boldsymbol{q}_1, \rho_1$ are given by

$$\nabla_{\boldsymbol{w}_1} \mathbf{L} = c_1 \boldsymbol{w}_1 + \boldsymbol{A}^T \boldsymbol{\alpha}_1 + \boldsymbol{B}^T \boldsymbol{\alpha}_2 - \boldsymbol{B}^T \boldsymbol{\alpha}_3 = 0$$
(3)

$$\nabla_{b_1} \mathbf{L} = c_1 b_1 + \boldsymbol{e}_1^T \boldsymbol{\alpha}_1 + \boldsymbol{e}_1^T \boldsymbol{\alpha}_2 - \boldsymbol{e}_2^T \boldsymbol{\alpha}_3 = 0$$
(4)

$$\nabla_{\boldsymbol{p}} \mathbf{L} = \boldsymbol{\alpha}_1 - \boldsymbol{p} = 0 \tag{5}$$

$$\nabla_{\boldsymbol{q}_1} \mathbf{L} = c_2 \boldsymbol{e}_2 - \boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3 - \boldsymbol{\alpha}_4 = 0 \tag{6}$$

$$\nabla_{\rho_1} \mathbf{L} = c_3 - \boldsymbol{e}_2^T \boldsymbol{\alpha}_3 - \beta = 0 \tag{7}$$

$$\boldsymbol{\alpha}_2 \ge 0, \boldsymbol{\alpha}_3 \ge 0, \boldsymbol{\alpha}_4 \ge 0, \beta \ge 0.$$
(8)

Then substituting (3) and (7) into the Lagrangian, we obtain the dual problem of the problem

$$\min_{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3} f(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3) = \frac{1}{2} (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \boldsymbol{\alpha}_3^T) \bar{\boldsymbol{H}} (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \boldsymbol{\alpha}_3^T)^T - [\boldsymbol{0}, \boldsymbol{0}, -c_1 \boldsymbol{e}_1] (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \boldsymbol{\alpha}_3^T)^T$$

s.t.
$$0 \le \boldsymbol{\alpha}_2 \le (c_2 + c_3) \boldsymbol{e}_2$$
(9)
$$0 \le \boldsymbol{\alpha}_3 \le c_3 \boldsymbol{e}_2,$$

where

$$ar{H} = egin{bmatrix} m{A}m{A}^T + c_1m{I} & m{A}m{B}^T & -m{A}m{B}^T \ m{B}m{B}^T & -m{B}m{B}^T \ m{B}m{B}^T & -m{B}m{B}^T \ m{B}m{B}^T & -m{B}m{B}^T \ m{B}^T \$$

The dual of the problem (2) is

4 M.Z. Liu, Y.H. Shao

where

$$ilde{m{H}} = egin{bmatrix} m{B}m{B}^T + c_4m{I} - m{B}m{A}^T & m{B}m{A}^T \ -m{A}m{B}^T & m{A}m{A}^T & -m{A}m{A}^T \ m{A}m{B}^T & -m{A}m{A}^T \ m{A}m{B}^T & -m{A}m{A}^T \ m{A}m{A}^T \end{bmatrix} + egin{bmatrix} m{e}_2 m{e}_2^T & -m{e}_2 m{e}_1^T & m{e}_2 m{e}_1^T \ -m{e}_1 m{e}_1^T & -m{e}_1 m{e}_1^T \ m{e}_1 m{e}_1^T & -m{e}_1 m{e}_1^T \ m{e}_1 m{e}_1^T \ m{e}_1 m{e}_1^T \end{bmatrix} .$$

Once the (\boldsymbol{w}_1, b_1) and (\boldsymbol{w}_2, b_2) are obtained from the (9) and (10) by means of the dual coordinate descent method [15], a new input $\boldsymbol{x} \in \mathbf{R}^n$ is assigned to class i (i = +1, -1) by Class $i = \arg \min_{k=1,2} \frac{|\boldsymbol{w}_k^T \boldsymbol{x} + b_k|}{\|\boldsymbol{w}_k\|}$.

3.2 Kernel PNSVM

Kernel PNSVM searches the following two kernel-generated surfaces instead of hyperplanes $K(\boldsymbol{x}^T, \boldsymbol{C}^T)\boldsymbol{w}_1 + b_1 = 0$ and $K(\boldsymbol{x}^T, \boldsymbol{C}^T)\boldsymbol{w}_2 + b_2 = 0$, where $\boldsymbol{C}^T = [\boldsymbol{A} \ \boldsymbol{B}]^T \in \mathbf{R}^{n \times \ell}$. The optimization problems are

$$\min_{\boldsymbol{w}_{1}, b_{1}, \boldsymbol{p}, \rho, \boldsymbol{q}_{1}} \frac{\frac{1}{2}\boldsymbol{p}^{T}\boldsymbol{p} + \frac{1}{2}c_{1}(\|\boldsymbol{w}_{1}\|^{2} + b_{1}^{2}) + c_{2}\boldsymbol{e}_{2}^{T}\boldsymbol{q}_{1} + c_{3}\rho_{1}}{s.t.} \\
s.t. \quad K(\boldsymbol{A}, \boldsymbol{C}^{T})\boldsymbol{w}_{1} + \boldsymbol{e}_{1}b_{1} = \boldsymbol{p}, \qquad (11) \\
\rho_{1}\boldsymbol{e}_{2} \geq -(k(\boldsymbol{B}, \boldsymbol{C}^{T})\boldsymbol{w}_{1} + \boldsymbol{e}_{2}b_{1}) + \boldsymbol{q}_{1} \geq \boldsymbol{e}_{2}, \ \boldsymbol{q}_{1} \geq 0, \quad \rho_{1} \geq 1,$$

and

$$\min_{\boldsymbol{w}_{2}, b_{2}, \boldsymbol{q}, \rho_{2}, \boldsymbol{q}_{2}} \frac{\frac{1}{2}\boldsymbol{q}^{T}\boldsymbol{q} + \frac{1}{2}c_{4}(\|\boldsymbol{w}_{2}\|^{2} + (b_{2})^{2}) + c_{5}\boldsymbol{e}_{1}^{T}\boldsymbol{q}_{2} + c_{6}\rho_{2} \\
s.t. \quad K(\boldsymbol{B}, \boldsymbol{C}^{T})\boldsymbol{w}_{2} + \boldsymbol{e}_{2}b_{2} = \boldsymbol{q}, \qquad (12) \\
\rho_{2}\boldsymbol{e}_{1} \ge (K(\boldsymbol{A}, \boldsymbol{C}^{T})\boldsymbol{w}_{2} + e_{1}b_{2}) + \boldsymbol{q}_{2} \ge \boldsymbol{e}_{1}, \ \boldsymbol{q}_{2} \ge 0, \rho_{2} \ge 1.$$

where $c_i, i = 1, \dots, 6$ are parameters.

4 Experimental results

In this section, the UCI data sets are chosen to demonstrate the performance of our PNSVM compared with SVC, TWSVM and WLTSVM [13]. The methods are implemented by Matlab 9.0 running on a PC with an Intel(R) Core Duo i7(2.70GHZ) with 32 GB RAM. Our PNSVM is solved by the dual coordinate descend algorithm and SVC and TWSVM are solved by the optimization toolbox QP in Matlab. The classification accuracy is measured by the standard 10-fold cross-validation.

The classification accuracy, computation time, and optimal values of $c_1 (= c_2), c_4 = (c_5)$ and c_3 and c_6 in our PNSVM are listed in Table 1. The parameters from PNSVM, TWSVM, SVC and WLTSVM are searched in the range $\{2^{2i}|i = -8, \dots, 8\}$, and the parameters c_3 and c_6 of PNSVM are selected from the set $\{0.01, 0.1, 1, 10, 100\}$. Table 1 shows the comparison results of all four methods. It can be seen from Table 1 that the accuracy of the proposed linear PNSVM is significantly better than that of the linear TWSVM on

5

Datasets	PNSVM		TWSVM		SVC		WLTSVM	
	accuracy%	time(s)	accuracy%	time(s)	$\operatorname{accuracy}\%$	time(s)	accuracy%	time(s)
	$c_1 = c_2/c_4 = c_5$	c_{3}/c_{6}						
Hepatitis	$ 84.93{\pm}0.03$	0.0005	82.89 ± 6.30	0.281	$84.13 {\pm} 5.58$	1.170	84.39 ± 1.35	0.0062
(155×19)	$2^{14}/2^{12}$	0.1/0.1						
BUPA liver	$70.98{\pm}0.01$	0.0346	66.40 ± 7.74	0.840	$67.78 {\pm} 5.51$	3.540	68.08 ± 1.32	0.0076
(345×6)	$2^{10}/2^2$	10/100						
Heart-Stat	84.41 ± 0.05	0.0012	$84.44{\pm}6.80$	0.454	$83.12 {\pm} 5.41$	1.584	$ 84.96{\pm}0.58$	0.0051
$-\log(270 \times 14)$	$2^{-6}/2^{16}$	10/10						
Heart-c	86.57 ± 0.07	0.0026	84.86 ± 6.27	0.516	$83.33{\pm}5.64$	2.193	$ 92.05{\pm}0.39$	0.0063
(303×14)	$2^{0}/2^{-8}$	0.1/100						
Votes	$96.43{\pm}0.05$	0.0043	95.85 ± 2.75	1.851	$95.80{\pm}2.65$	3.192	95.93 ± 0.21	0.0049
(435×16)	$2^{-6}/2^{-2}$	10/0.01						
WPBC	$84.72{\pm}0.06$	0.0253	83.68 ± 5.73	0.560	$83.30{\pm}4.53$	2.094	79.18 ± 1.73	0.0778
(198×34)	$2^{-14}/2^{-14}$	0.01/100						
Sonar	76.19 ± 0.15	2.4545	77.00 ± 6.10	0.375	$80.13{\pm}5.43$	0.941	78.07 ± 1.41	0.0172
(208×60)	$2^2/2^{-12}$	100/100						
lonosphere	$ 88.89{\pm}0.08$	0.04	88.48 ± 5.74	0.969	$88.20{\pm}4.51$	4.120	87.31 ± 0.64	0.0097
(351 ± 34)	$2^{16}/2^{-2}$	0.01/100		,			•	

 Table 1. The comparison results of linear classifiers.



Fig. 1. The influence on Accuracy(AC) of parameters c_3 and c_6 of linear PNSVM.

most of data sets. And our PNSVM is the fastest on most of data sets. Fig. 1 exhibits the influence on Accuracy of parameters c_3 and c_6 for linear PNSVM. It can be observed that the choice of c_3 and c_6 affects the results dramatically, which implies that adjusting the parameters c_3 and c_6 is practical selection in real applications. Table 2 is concerned with our kernel PNSVM, TWSVM and SVC and WLTSVM. The Gaussian kernel $K(\boldsymbol{x}, \boldsymbol{x}') = e^{-\mu ||\boldsymbol{x}-\boldsymbol{x}'||^2}$ is used. The kernel parameter μ is also searched from the sets $\{2^i | i = -8, \dots, 8\}$. The classification accuracy and computation times for all three methods are also listed in Table 2.

 Table 2. The comparison results of kernel classifiers.

Datasets	PNSVM		TWSVM		SVC		WLTSVM	
	accuracy%	time(s)	accuracy%	time(s)	accuracy%	time(s)	$\operatorname{accuracy}\%$	time(s)
	$c_1 = c_2/c_4 = c_5$	$c_3/c_6/\mu$						
Hepatitis	$84.29{\pm}0.07$	0.0051	83.39 ± 7.31	0.797	$84.13 {\pm} 6.25$	1.300	$84.28 {\pm} 0.47$	0.0057
(155×19)	$2^{-2}/2^{-14}$	$0.01/0.01/2^2$						
BUPA liver	69.72 ± 0.01	0.0135	67.83 ± 6.49	2.700	68.32 ± 7.20	5.248	$\textbf{73.18}{\pm}\textbf{0.59}$	0.0046
(345×6)	$2^{-14}/2^8$	$100/0.01/2^{-5}$						
Heart-Stat	$86.68{\pm}0.06$	0.1703	82.96 ± 4.67	1.130	$83.33 {\pm} 9.11$	6.100	$85.12 {\pm} 0.86$	0.0154
$-\log(270 \times 14)$	$2^{-10}/2^{-10}$	$10/1 / 2^1$						
Heart-c	$89.28{\pm}0.05$	0.0111	83.83 ± 5.78	2.141	$83.68 {\pm} 5.67$	3.800	$85.98 {\pm} 0.62$	0.0059
(303×14)	$2^{-12}/2^4$	$1/100/2^{-3}$						
Votes	$96.53{\pm}0.03$	0.0181	94.91 ± 4.37	3.540	$95.64{\pm}7.23$	7.783	$96.39 {\pm} 0.14$	0.0475
(435×16)	$2^{-8}/2^0$	$0.1/0.01/2^{-2}$						
WPBC	$84.39{\pm}0.06$	0.0096	81.28 ± 5.92	1.305	$80.18 {\pm} 6.90$	4.141	$79.85 {\pm} 0.43$	0.0085
(198×34)	$2^{-10}/2^{12}$	$10/10/2^0$			•			
Sonar	$90.01{\pm}0.06$	0.0077	89.64 ± 6.11	2.630	88.93 ± 10.43	5.302	$88.73 {\pm} 1.27$	0.0237
(208×60)	$2^{-8}/2^{-6}$	$1/10/2^{-2}$						
lonosphere	$ 91.81{\pm}0.03$	0.0608	$ 87.46\pm3.40$	5.576	$90.20 {\pm} 4.51$	15.71	$90.45{\pm}0.47$	0.0039
(351 ± 34)	$2^{0}/2^{2}$	$10/100/2^{-3}$						

5 Conclusions

For binary classification problem, a novel large margin proximal non-parallel twin support vector machine was proposed in this paper. The main contribution are that the structural risk minimization principle is implemented by introducing a regularization term in the primal problems of our PNSVM and the dual formulation of PNSVM avoids the computation of inverse matrices and speeds up the training efficiency. Experimental results of our PNSVM and SVC and TWSVM and WLTSVM, have been made on several data sets, implying that the proposed method is not only faster but also exhibits better generalization. It should be pointed out that there are six parameters in our PNSVM, so the parameter selection and more efficient algorithm [17] is a practical problem and should be addressed in the future.

References

- 1. Vapnik, V.N.: Statistical Learning Theory, Wiley, New York (1998).
- Deng, N.Y., Tian, Y.J., Zhang, C.H.: Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions, CRC Press, New York (2012).
- Noble, W.S.: Support vector machine applications in computational biology. Kernel methods in computational biology 71–92 (2004).
- Mangasarian, O.L., Wild, E.W.: Multisurface proximal support vector machine classification via generalized eigenvalues. IEEE transactions on pattern analysis and machine intelligence 28(1), 69–74 (2006).
- Jayadeva, R.K., Khemchandani, R., Chandra, S.: Twin support vector machines for pattern classification. IEEE transactions on pattern analysis and machine intelligence 29(5) 905–910 (2007).
- Shao, Y.H., Zhang, C.H., Wang, X.B., Deng, N.Y.: Improvements on Twin Support Vector Machines. IEEE transactions on Neural Networks 22(6) 962–968 (2011).
- Shao, Y.H., Wang, Z., Chen, W.J., Deng, N.Y.: A regularization for the projection twin support vector machine. Knotledge-Based Systems 37 203–210 (2013).
- Shao, Y.H., Chen, W.J., Zhang, J.J., Wang, Z., Deng, N.Y.: An efficient weighted Largrangian twin support vector machine for imbalanced data classification. Pattern Recognition 47(9) 3158–3167 (2014).
- Liu, M.Z., Shao, Y.H., Wang, Z., Li, C.N., Chen, W.J.: Minimum deviation distribution machine for large scale regression. Knowledge-Based Systems 146 167–180 (2018).
- Qi, Z.Q., Tian, Y.J., Shi, Y.: Robust twin support vector machine for pattern classification. Patterm Recognition 46(1) 305–316 (2013).
- Tian, Y.J., Qi, Z.Q., Ju, X.C., Shi, Y., Liu, X.H.: Nonparallel support vector machines for pattern classification. IEEE transactions on cybernetics 44(7) 1067– 1079 (2014).
- Tian, Y.J., Ping, Y.: Large-scale linear nonparallel support vector machine solver. Neural Networks, 50 166–174 (2014).
- Shao, Y.H., Chen, W.J., Wang, Z., Li, C.N., Deng, N.Y.: Weighted linear loss twin support vector machine for large-scale classification. Knowledge-Based Systems 73 276–288 (2015).
- Shao, Y.H., Chen, W.J., Deng, N.Y.: Nonparallel hyperplane support vector machine for binary classification problems. Information Sciences 263 22–35 (2014).
- Hsieh, C.J., Chang, K.W., Lin, C.J., Keerthi, S.S., Sundararajan, S.: A dual coordinate descent method for large-scale linear SVM. In: Proceedings of the 25th international conference on Machine learning, pp. 408–415. ACM, (2008).
- Shao, Y.H., Li, C.N., Liu, M.Z., Wang, Z., Deng, N.Y.: Sparse L_q-norm least square support vector machine with feature selection. Pattern Recognition, **78** 167–181 (2018).
- 17. Wang, Z., Shao, Y.H., Bai, L., Liu, L.M., Deng, N.Y.: Stochastic gradient twin support vector machine for large scale problems. arXiv preprint arXiv:1704.05596.