CrossMark

Applying 1-norm SVM with squared loss to gene selection for cancer classification

Li Zhang[1] · **Weida Zhou**[2] · **Bangjun Wang**[1] · **Zhao Zhang**[1] · **Fanzhang Li**[1]

Abstract Gene selection methods available have high computational complexity. This paper applies an 1-norm support vector machine with the squared loss (1-norm SVMSL) to implement fast gene selection for cancer classification. The 1-norm SVMSL, a variant of the 1-norm support vector machine (1-norm SVM) has been proposed. Basically, the 1-norm SVMSL can perform gene selection and classification at the same. However, to improve classification performance, we only use the 1-norm SVMSL as a gene selector, and adopt a subsequent classifier to classify the selected genes. We perform extensive experiments on four DNA microarray data sets. Experimental results indicate that the 1-norm SVMSL has a very fast gene selection speed compared with other methods. For example, the 1-norm SVMSL is almost an order of magnitude faster than the 1-norm SVM, and at least four orders of magnitude faster than SVM-RFE (recursive feature elimination), a state-of-the-art method.

Keywords Support vector machine · Gene selection · Cancer classification · 1-norm support vector machine · Orthogonal matching pursuit

✉ Li Zhang
zhangliml@suda.edu.cn

1   School of Computer Science and Technology, Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou 215006, Jiangsu, China

2   AI Speech Ltd., Suzhou 215123, Jiangsu, China

1 Introduction

Microarray technology can simultaneously access the expression level of thousands of genes for DNA, which can be used for identifying and classifying cancer diseases [6, 16, 19, 20, 26, 27, 36, 37, 39]. Many methods have been applied to cancer classification, such as support vector machines(SVMs) [4, 16], and k nearest neighbor (kNN) [4, 19]. Generally, a DNA microarray data set consists of a large number of genes and a relatively small number of samples for cancer classification. Among these genes, not all of them are useful for cancer classification tasks. To find the useful genes in the gene expression data, it is necessary to implement gene selection or feature selection for cancer classification based on gene expression data.

Many gene selection method have been proposed, such as genetic algorithm-based methods [19, 26], and SVMs-based methods [4, 10, 12, 16–18, 38, 46]. In this paper, we focus on the gene selection methods based on SVMs. An embedded feature selection algorithm based on SVMs was proposed in [18], which can adaptively identify important features through introducing data driven weights, and simultaneously implement classification and gene selection. However, this method requires adjusting more parameters, and its performance largely depends on the parameters. The $\ell_2 - \ell_0$ SVM, a variant of SVM, can be applied to feature selection [31]. How to design the objective function based on the standard SVM is the key and difficult issue of this kind of algorithm.

Lee and Leu proposed a hybrid feature selection algorithm based on SVM for microarray data analysis [17]. Specifically, this method uses the genetic algorithm to generate a number of subsets of genes, the chi square test to

select a proper number of the top-ranked genes for data analysis and SVM to verify the efficiency of the selected genes. Weston et al. proposed the wrapper feature selection algorithm based on SVM, which finds useful genes by minimizing bounds on the leave-one-out error using gradient descent [38].

Guyon et al. proposed SVM-RFE (recursive feature elimination) feature selection algorithm, which is the most representative gene selection algorithm based on SVMs [16]. Duan et al. presented a new gene selection method based on SVM-RFE, called MPSVM-RFE. At each step, MPSVM-RFE trains multiple linear SVMs on subsamples of training data and computes the feature ranking scores from statistical analysis of the weight vectors [10]. To deal with multiclass cancer classification tasks, a multi-class SVM-RFE (MSVM-RFE) [46], multiple SVM-RFE (MPSVM-RFE) [40] and a multiple support vector data description recursive feature elimination (MSVDD-RFE) [4] have been developed. Once gene selection is achieved, a classifier should be applied subsequently. Although SVM-RFE methods can select useful genes for cancer classification, the computational complexity of these methods are very high when the gene number is huge.

Fortunately, the 1-norm SVM can be applied to perform gene selection fast [12] and feature selection for conjoint analysis [22]. The 1-norm SVM is a variant of SVM [2, 23, 43, 45, 47]. It is well-known that the objective function of SVM consists of 2-norm regularization penalty and the hinge loss [13, 29, 34, 35]. The 1-norm SVM is to simultaneously minimize the 1-norm regularization penalty and the hinge loss. Since the 1-norm regularization penalty as a substitution of the 0-norm regularization penalty can also induce a sparse solution, it has shown that the 1-norm SVM has a better sparse representation than SVM [43].

Fung and Mangasarian used the Newton method to minimize an exterior penalty problem constructed from the dual problem of the 1-norm SVM, which is called NLPSVM, and applied NLPSVM to gene selection [12]. Since NLPSVM with the linear kernel directly picks up the useful genes and does not use the recursive feature elimination scheme, NLPSVM is much faster than SVM-RFE methods. In addition, NLPSVM can can classify microarray data in its own way. As a variant of the 1-norm SVM, the 1-norm SVM with a squared loss (1-norm SVMSL) was proposed in [42]. Since the 1-norm SVMSL uses orthogonal matching pursuit (OMP) to find an approximate solution, its training speed is very fast. It is the approximate solution that would cause loss on the classification performance.

To fast select gene and achieve better classification performance, this paper designs a cancer classification framework. The 1-norm SVMSL is first used to select useful genes, and then a subsequent classifier is applied on the reduced data. We perform extensive experiments on DNA microarray data sets.

The rest of the paper is organized as follows. In Section 2, we address related work on gene selection based on SVMs. Section 3 gives the framework of this paper. Experimental results on the real-world data sets are given in Section 4. Section 5 concludes this paper.

2 Related work

This section introduces gene selection methods based on SVMs. We discuss SVM-RFE and the 1-norm SVM, respectively. Generally, cancer classification can be simply described as follows. Given a set of training samples $X = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^D$ denotes the gene expression of a patient, $y_i \in \{-1, +1\}$ is the label of $\mathbf{x}_i$, D is the number of genes (or the dimension of samples), and n is the number of samples, we try to find genes from this training set which are useful for discriminating cancer diseases. In different datasets, the label y_i of $\mathbf{x}_i$ has different meanings. In the Breast Cancer dataset [33], for example, the label "+1" means "non-relapse" and the label "-1" means "relapse".

2.1 SVM-RFE

Support vector machine, the state of the art learner in machine learning and pattern recognition, is famous for its good generalization performance, sparse model representation, and nonlinear learning ability [13, 29, 34, 35].

As a classifier, SVM requires a set of training samples X, and solves the following primal optimization problem:

$$\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}\|\mathbf{w}\|_2^2 + C \sum_{i=1}^n \xi_i \tag{1}$$

$$subject\ to \quad y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 - \xi_i,$$
$$\xi_i \geq 0, i = 1, \cdots, n$$

where ξ_i is the slack variable, $\|\cdot\|$ denotes the 2-norm of a vector, $C > 0$ is the regularization factor which controls the balance between the empirical risk and the capacity of SVM, and $\mathbf{w}$ and b are respectively the weight vector and the threshold in the hypothesis function

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{2}$$

where $\mathbf{w} = [w_1, w_2, \cdots, w_D]^T$. As long as we get $\mathbf{w}$ and b, we can assign a label for any sample $\mathbf{x}$ using

$$f^*(\mathbf{x}) = sign\left(f(\mathbf{x})\right) = sign\left(\mathbf{w}^T\mathbf{x} + b\right) \tag{3}$$

where $sign(\cdot)$ is the sign function.

To implement gene selection for cancer classification, SVM-RFE was proposed in [16]. In a sequential backward elimination way, SVM-RFE iteratively deletes redundant genes and retains useful genes. In each iteration, SVM-RFE trains a linear SVM model with the reserved genes, and deletes a gene according to weight scores w_j^2, $j = 1, \cdots , D$. The weight score for gene j is $c_j = w_j^2$. The higher c_j , the greater the importance of the jth gene is. The detail algorithm for SVM-RFE is shown in Algorithm 1 [16]. Note that the order of elements in R is very important in Step 2(d). The last element in R means that the corresponding gene has the highest score and is the most important one.

A novel gene selection method, MPSVM-RFE was proposed based on SVM-RFE [10]. In each iteration, MPSVM-RFE computes the ranking scores from a statistical analysis of the weight vectors in multiple linear SVMs trained on original training subsets. CV (cross-validation) is adopted to resample the training data. Thus, the computational complexity of MPSVM-RFE is much higher than SVM-RFE.

Algorithm 1 SVM-RFE

Input: Training samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$;
Output: Ranked gene list R.
1. Initialize: The selected gene subset $S = \{1, \cdots , D\}$ and the ranked gene list $R = \emptyset$.
2. Repeat until $S = \emptyset$:
  (a). Train a linear SVM with genes in the set S to get the weight $\mathbf{w}$;
  (b). Compute the ranking weight scores for each gene in the set S. Namely,
$$c_j = w_j^2, j \in S$$
  (c). Find the gene with the smallest ranking weight score:
$$p = \arg\min_{j \in S} c_j$$
  (d). Update the ranked gene list R by adding the index of gene p into it,
$$R \leftarrow R \cup \{p\}$$
  (e). Update the selected gene subset S by removing the index of gene p from it,
$$S \leftarrow S - \{p\}$$

2.2 1-norm SVM

Since the 1-norm SVM has sparsity, the linear 1-norm SVM can be used for feature selection or variable selection [3]. As we know, there are some different LP forms for the 1-norm SVM [2, 8, 23, 24, 43, 45, 47]. However, these forms are proved to be essentially equivalent while selecting proper parameters. Here we introduce the LP form presented in [43], which can be solved by the simplex method, the interior-point method, NLPSVM [12], the column generation Newton (CGN) method [44] and the row-column generation (RCG) method [41]. Among these methods, NLPSVM is relatively fast in the linear case. CGN and RCG were proposed for improving the training speed in the nonlinear case.

Consider the training sample set of two classes $X = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i \in \mathbb{R}^D$ and $y_i \in \{1, -1\}$. The 1-norm SVM is to simultaneously minimize the 1-norm regularization penalty and the hinge loss, which can be expressed as a linear programming problem. The primal problem of the 1-norm SVM is:

$$\min \sum_{j=1}^{D} \left(\alpha_j^+ + \alpha_j^- \right) + \sigma \left(\beta^+ + \beta^- \right) + C \sum_{i=1}^{n} \gamma_i$$
$$s.t. \quad y_i \left[\mathbf{x}_i^T \left(\alpha^+ - \alpha^- \right) + \left(\beta^+ - \beta^- \right) \right] \geq 1 - \gamma_i \qquad (4)$$
$$\beta^+, \beta^- \geq 0, \gamma_i \geq 0, i = 1, \cdots , n$$
$$\alpha_j^+, \alpha_j^- \geq 0, j = 1, \cdots , D$$

where $C > 0$ is the regularization factor, $\sigma > 0$ is a small constant, $\alpha^+ = \left[\alpha_1^+, \cdots , \alpha_D^+ \right]^T$, $\alpha^- = \left[\alpha_1^-, \cdots , \alpha_D^- \right]^T$, $\beta^+ \in \mathbb{R}$ and $\beta^- \in \mathbb{R}$ are model coefficients for the 1-norm SVM, and $\gamma = [\gamma_1, \cdots , \gamma_\ell]^T \in \mathbb{R}^n$ is a loss vector.

Naturally, (4) could be rewritten in matrix form

$$\min_{\mathbf{u}} \ c^T \mathbf{u}$$
$$s.t. \quad A u \geq \mathbf{b} \qquad (5)$$
$$u \geq \mathbf{0}$$

where the variable vector $\mathbf{u} = [(\alpha^+)^T, (\alpha^-)^T, \beta^+, \beta^-, \gamma^T]^T$, $\mathbf{c} = \left[\mathbf{1}^T, \mathbf{1}^T, \sigma, \sigma, C\mathbf{1}^T \right]^T$, $\mathbf{b} = \mathbf{1}$, the constraint coefficient matrix $\mathbf{A} = \left[\mathbf{D}_y \mathbf{X}, -\mathbf{D}_y \mathbf{X}, \mathbf{y}, -\mathbf{y}, \mathbf{I} \right]$, $\mathbf{X} \in \mathbb{R}^{n \times D}$ is the sample matrix in which each sample is a row vector, $\mathbf{y} = [y_1, y_2, \cdots , y_n]^T$, $\mathbf{D}_y$ is the diagonal matrix with the diagonal line of $\mathbf{y}$, and $\mathbf{1}$ and $\mathbf{I}$ are the column vector of all ones and the identity matrix with the appropriate size, respectively.

The decision function of the linear 1-norm SVM takes the form:

$$f^*(\mathbf{x}) = sign \left(\left(\alpha^+ - \alpha^- \right)^T \mathbf{x} + \left(\beta^+ - \beta^- \right) \right)$$
$$= sign \left(\alpha^T \mathbf{x} + \beta \right) \qquad (6)$$

where $\alpha = \alpha^+ - \alpha^-$ and $\beta = \beta^+ - \beta^-$.

Once the problem (5) is solved, we can get α and β. Since most elements in α are zero, gene selection is automatically implemented by searching nonzero coefficients in α. If $\alpha_j \neq 0$, then the j gene is selected.

The 1-norm SVM solved by NLPSVM is much faster than SVM-RFE when dealing with gene selection since it does not require a recursive process. The complexity of each iteration in NLPSVM is $O(min(n, D)^3)$ when applying the Sherman-Morrison-Woodbury identity equation, where n is the number of samples and D is the number of genes.

3 Framework of the proposed method

The classification framework based on the 1-norm SVMSL is shown in Fig. 1, where the linear SVM is taken as an illustration of subsequent classifiers, X is the given training sample set, and X' is the reduced training sample set. We first introduce the 1-norm SVMSL, and then respectively discuss gene selection and classification stages.

3.1 1-norm SVMSL

Given a training set $X = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ with $\mathbf{x} \in \mathbb{R}^D$, our goal is to select d genes from D ones with $d << D$ in the stage of gene selection. The 1-norm SVMSL is applied to select the d genes.

By replacing the hinge loss with the squared loss and ignoring σ, we can rewrite (4) and obtain the optimization problem for the 1-norm SVMSL:

$$\min \left\| \alpha^+ \right\|_1 + \left\| \alpha^- \right\|_1 + \sigma \left(\beta^+ + \beta^- \right) + C \left\| \mathbf{1} - \mathbf{D}_y \left(\mathbf{X}\alpha + \beta \right) \right\|_2^2$$
$$s.t.\ \beta^+ \geq 0,\ \beta^- \geq 0,\ \alpha^+ \geq \mathbf{0},\ \alpha^- \geq \mathbf{0} \tag{7}$$

where $\|\cdot\|_1$ denotes the 1-norm of a vector. The standard matrix formulation of (7) can be expressed as

$$\min_{\mathbf{u}}\ \tau \|\mathbf{u}\|_1 + \frac{1}{2} \|\mathbf{c} - \mathbf{A}\mathbf{u}\|_2^2$$
$$s.t.\ \mathbf{u} \geq \mathbf{0} \tag{8}$$

where $\mathbf{u} = \left[(\alpha^+)^T, (\alpha^-)^T, \beta^+, \beta^- \right]^T$, $\mathbf{A} = \left[\mathbf{D}_y \mathbf{X}, -\mathbf{D}_y \mathbf{X}, \mathbf{y}, -\mathbf{y} \right]$, $\mathbf{c} = \mathbf{1}$, and $\tau > 0$ is a non-negative real parameter which has the same role as the parameter C in both SVM and the 1-norm SVM.

In sparse signal reconstruction, the problem (8) is equivalent to both the least absolute shrinkage and selection operator (LASSO), and the quadratically constrained linear problem (QCLP) under some conditions [1]. Thus, orthogonal matching pursuit (OMP) which has been applied to approximately solve QCLP [7, 9, 32] can also be used to approximately solve the 1-norm SVMSL. In addition, OMP needs only the parameter d instead of τ in solving (8).

3.2 Gene selection and classification

When the problem (8) is solved approximately by using OMP, we can get the coefficient vector vector α. Similar to NLPSVM, the 1-norm SVMSL can automatically select genes. However, since the solution to the 1-norm SVMSL is only an approximate one, the classification performance obtained by directly using α is not consistent with one generated by an exact solution. The experimental results in [44] also show that the performance of approximate algorithms could be better or worse than that of exact algorithms.

To avoid this situation, we can use a subsequent classifier to classify the reduced dataset, such as support vector machine, and nearest neighbor. As shown in Fig. 1, the framework shows the training and test processes, which are respectively discussed as follows.

The training process is given in Algorithm 2, which requires training both the 1-norm SVMSL and the linear SVM. Let $X = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ denote the given training sample set. Assume that we need to select d genes from D ones, where $d << D$. The 1-norm SVMSL is first trained by using X and then outputs the selected gene subset S with $|S| = d$. According to the subset S, we can select useful genes and generate a new training set $X' = \{(\mathbf{x}_i', y_i)\}_{i=1}^n$, where $\mathbf{x}_i' \in \mathbb{R}^d$. The subsequent classifier, for example the linear SVM, is trained by using X'.

For an unseen sample $\mathbf{x}$, we first reduce it to $\mathbf{x}'$ employing S and then apply the trained model on $\mathbf{x}'$ to assign a class label for $\mathbf{x}$.


Algorithm 2 Training process of the proposed method

Input: Training samples $X = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, the reserved gene number d;

Output: Selected gene subset S and the trained classifier model.

1. Initialize the selected gene subset $S = \emptyset$;
2. Train a linear 1-norm SVMSL with all genes using OMP to get α;
3. For $j = 1, \cdots, D$
   If $\alpha_j \neq 0$, THEN $S \leftarrow S \cup \{j\}$.
4. Generate a new training set X' according to S;
5. Train a classifier with X';
6. Return S and the trained classifier model.

Fig. 1 Classification framework based on 1-norm SVMSL
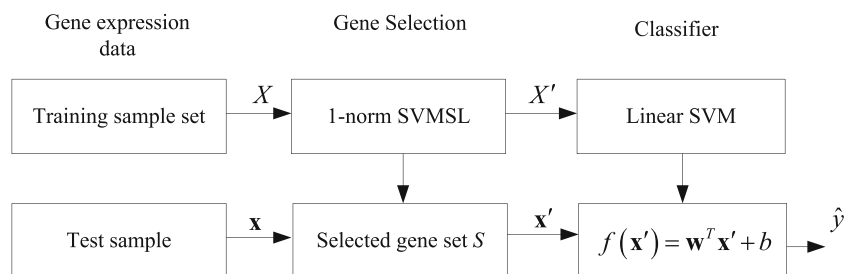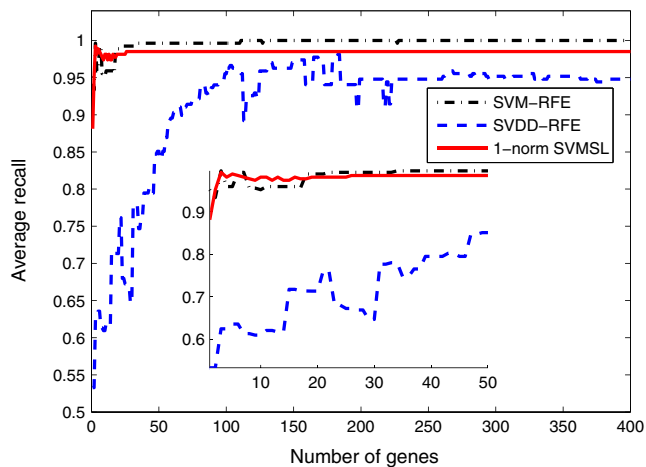
Fig. 2 Average recall vs. gene number with the linear SVM on the Lung cancer dataset

4 Results and discussions

This section discusses experimental results of the proposed method on cancer classification using gene expression data and compares it with classical gene selection methods, such as SVM-RFE.

4.1 Experiment settings

Four public gene microarray datasets available are used to validate the performance of these compared methods, including Leukemia-ALLAML [14], Lung Cancer [15], Central Nervous System (CNS) dataset [25], and Prostate Cancer [28]. These four datasets can be downloaded from http://datam.i2r.a-star.edu.sg/datasets/krbd/. All data here are normalized such that each gene has mean 0 and variance 1 according to the way mentioned in [11].

Since a subsequent classifier is required, we here provide two classifiers, including SVM, and the nearest neighbor (NN) to observe the performance of the 1-norm SVMSL. The compared methods include SVM-RFE+SVM/NN [16], SVDD-RFE+SVM/NN [4], the 1-norm SVM [12, 43], the

1-norm SVMSL [42], the 1-norm SVM+SVM/NN, and the proposed 1-norm SVMSL+SVM/NN. Since the 1-norm SVM is similar to the 1-norm SVMSL, SVM or NN could be subsequent to it. Thus, we have the compared method, the 1-norm SVM+SVM/NN. Note that there are many methods for solving the 1-norm SVM, we choose two ways, or the simplex method and the Newton method. The 1-norm SVM with the simplex method is an exact solution one, and with the Newton method is an approximate one. For simplicity, the 1-norm SVM with the simplex method is still denoted by the 1-norm SVM. The 1-norm SVM with the Newton method is NLPSVM proposed in [12].

All these gene selection methods use the linear kernel. The regularization parameter C for SVM-RFE is selected from the set $\{0.5, 1, 2, 10, 100\}$, for SVDD-RFE from the set $\{1/n, 0.1, 0.5, 1\}$, and for 1-norm SVM from the set $\{0.5, 1, 2, 10, 100\}$ according to the empirical evidence. The regularization parameter is approximately selected by using 3-fold cross validation on the training set and the corresponding classifiers. For example, C for SVM-RFE is selected by using SVM. For the 1-norm SVMSL, the only parameter is d which determines the selected gene number.

We use recalls to evaluate the performance of these methods. The recall of the jth class is defined as

$$Recall_j = \frac{TP_j}{TP_j + FN_j} \tag{9}$$

where TP_j and FN_j are the number of correctly classified samples and the number of incorrectly classified samples in class j, respectively.

All programs are written in MATLAB or/and C. Both SVM and SVDD are implemented by using the LIBSVM package [5], the 1-norm SVM is implemented by using GLPK package [21], NLPSVM is implemented by using the code provided in [12]. In addition, the 1-norm SVMSL is solved by using OMP in the SparseLab package [30]. All numerical experiments are performed on a personal computer with a 2.93GHz Intel(R) Core(T)2 Duo CPU and 2G bytes of memory. This computer runs on Windows 7, with MATLAB 7.10 compiler installed.

Table 1 Comparison of gene selection methods with SVM on the Lung cancer dataset

Method	# Gene	ADCA recall(%)	MPM recall(%)	Average recall (%)	CPU time (s)
None+SVM	Full(12533)	99.25	100.00	99.63	0.3276
SVM-RFE+SVM	111	100.00	100.00	100.00	1.1509×10^3
SVDD-RFE+SVM	183	96.27	100.00	98.13	270.2028
1-norm SVM	24	97.76	100.00	98.88	0.7488
1-norm SVM+SVM	24	98.51	100.00	99.25	0.7644
NLPSVM	33	97.76	93.33	95.55	1.2948
NLPSVM+SVM	33	97.01	100.00	98.51	1.3416
1-norm SVMSL	3	97.76	100.00	98.88	0.0164
1-norm SVMSL+SVM	3	98.51	100.00	99.25	0.0195
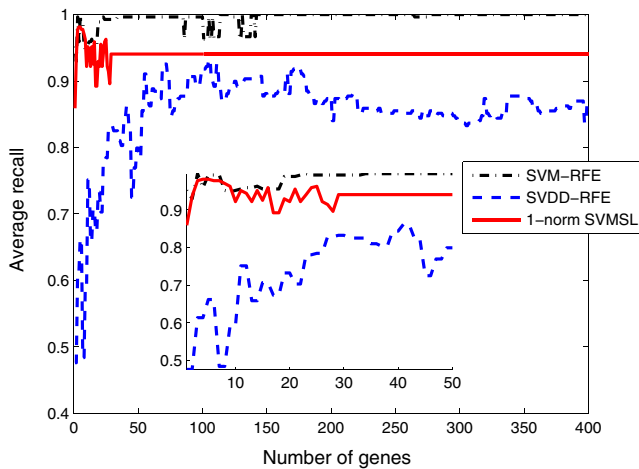
Fig. 3 Average recall vs. gene number with NN on the Lung cancer dataset

4.2 Lung cancer

Lung cancer consists of 181 tissue samples of which 31 samples are malignant pleural mesothelioma (MPM), and 150 adenocarcinoma(ADCA) of the lung. The training set contains 32 of them, 16 MPM and 16 ADCA. The remaining 149 samples are used for test, with 15 MPM and 134 ADCA. Each sample is described by 12533 genes.

4.2.1 SVM classifier

On the training set, we determine the parameter for SVM-RFE $C = 10$, for SVDD-RFE $C = 1$, and for the 1-norm SVM, NLPSVM and the 1-norm SVMSL $C = 2$. Note that SVM-RFE and SVDD-RFE can rank genes, but the 1-norm SVMSL and the 1-norm SVM can not. However, 1-norm SVMSL has a free parameter d which can represent the number of selected genes. Thus, we can select the last d genes in R for the first two methods and set different d for the 1-norm SVMSL. Let d vary from 1 to 400, which follows the experiment setting in [46]. The average recall vs. d for the three methods is shown in Fig. 2.

When d is small, say $d < 20$, the 1-norm SVMSL is better than SVM-RFE and SVDD-RFE. When $d > 25$, the performance of the 1-norm SVMSL is unchanged. In addition, the maximum selected gene number is 29 for the 1-norm SVMSL on the Lung cancer dataset. Thus, the 1-norm SVMSL has a stable performance when $d \geq 29$. The 1-norm SVM automatically determines the number of selected genes which is the number of non-zero coefficients in its model. On this dataset, 24 genes are selected by the 1-norm SVM, and 33 by NLPSVM.

Table 1 lists the best average performance and the corresponding gene number and CPU time for all these methods, where CPU time includes gene selection, classifier training and test. Observation on Table 1 indicates that the 1-norm SVMSL can very quickly pick up genes, and the use of the subsequent classifier can also improve the average recall from 98.99% to 99.25%. Actually, for the 1-norm SVM-like methods, the use of SVM improves their classification performance. The 1-norm SVMSL is almost four orders of magnitude faster than SVM-RFE, three orders faster than SVDD-RFE, two orders faster than NLPSVM, and one order faster than the 1-norm SVM. In addition, note that the number of useful genes is relatively small for the 1-norm SVMSL, only three.

4.2.2 NN classifier

The subsequent classifier used here is NN. Similary, let d vary from 1 to 400 for three methods, SVM-RFE, SVDD-RFE and the 1-norm SVMSL. The average recall vs. d is shown in Fig. 3. We can see that the 1-norm SVMSL has good average recall when d is small, say $d < 10$.

Table 2 lists the best average performance and the corresponding gene number and CPU time for all these methods, where CPU time includes gene selection and classification process. Since the 1-norm SVM, NLPSVM and the 1-norm SVMSL can implement gene selection and classification at the same time, their results are the same as those

**Table 2** Comparison of gene selection methods with NN on the Lung cancer dataset

Method	# Gene	ADCA recall(%)	MPM recall(%)	Average recall (%)	CPU time (s)
None+NN	Full(12533)	93.33	97.76	95.55	0.4836
SVM-RFE+NN	114	100.00	100.00	100.00	1.1509×10^3
SVDD-RFE+NN	105	100	86.57	93.28	270.1965
1-norm SVM	24	97.76	100.00	98.88	1.0920
1-norm SVM+NN	24	100.00	98.51	99.25	1.0920
NLPSVM	33	97.76	93.33	95.55	1.3260
NLPSVM+NN	33	93.33	97.01	95.17	1.3260
1-norm SVMSL	3	97.76	100.00	98.88	0.0116
1-norm SVMSL+NN	4	100	96.27	98.13	0.0120

Table 3 Genes selected by the 1-norm SVMSL on the Lung cancer dataset

Index	Gene ID	Official symbol	Description
7249	23194	FBXL7	F-box and leucine-rich repeat protein 7 [Homo sapiens (human)]
10817	5080	PAX6	Paired box 6 [Homo sapiens (Human)]
2521	1212	CLTB	Clathrin, light chain B [Homo sapiens (human)]
7388	3710	ITPR3	Inositol 1,4,5-trisphosphate receptor, type 3 [Homo sapiens (human)]

Table 4 Average correlation coefficient and distribution of best gene subsets on the Lung cancer dataset

Method	# Gene pair	$\bar{\rho}$	Coefficient distribution (%)				
			0.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
SVM-RFE	6150	0.4732	2.83	25.45	53.01	17.92	0.79
SVDD-RFE	16653	0.1956	57.94	31.45	9.35	1.23	0.04
1-norm SVM	276	0.4879	2.17	18.48	51.09	27.17	1.09
NLPSVM	528	0.4392	3.98	32.20	47.54	16.29	0.00
1-norm SVMSL	3	0.1476	66.67	33.33	0.00	0.00	0.00

(a) SVM (b) NN
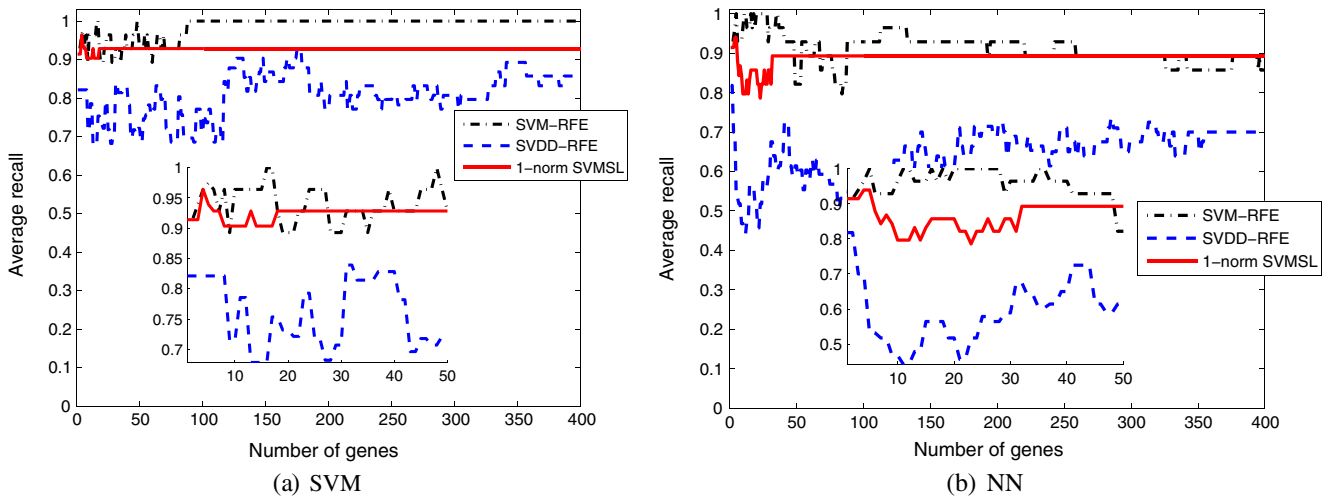
Fig. 4 Average recall vs. gene number on the Leukemia-ALLAML dataset

Table 5 Comparison on the Leukemia-ALLAML dataset

Method	# Gene	AML recall(%)	ALL recall(%)	Average recall (%)	CPU time (s)
None+SVM	Full(7129)	71.43	100.00	85.71	0.078
None+NN	Full(7129)	90.00	57.14	73.57	0.0468
SVM-RFE+SVM	16	100.00	100.00	100.00	365.4999
SVM-RFE+NN	5	100.00	100.00	100.00	365.4951
SVDD-RFE+SVM	175	85.71	100.00	92.86	89.6128
SVDD-RFE+NN	1	85.00	78.57	81.79	89.6079
1-norm SVM	22	92.86	100.00	96.43	0.7332
1-norm SVM+SVM	22	85.71	100.00	92.86	0.7644
1-norm SVM+NN	22	100.00	85.71	92.86	0.7332
NLPSVM	36	71.43	100.00	80.71	0.8268
NLPSVM+SVM	36	85.71	100.00	92.86	0.8268
NLPSVM+NN	36	95.00	57.14	76.07	0.8268
1-norm SVMSL	2	92.86	95.00	93.93	0.0154
1-norm SVMSL+SVM	4	90.86	100.00	96.43	0.0170
1-norm SVMSL+NN	4	95.00	92.86	93.93	0.0156

Table 6 Genes selected by the 1-norm SVMSL on the Leukemia-ALLAML dataset

Index	Gene ID	Official symbol	Description
4847	7791	ZYX	Zyxin [Homo sapiens (human)]
1779	4353	MPO	Myeloperoxidase [Homo sapiens (Human)]
1710	7389	UROD	Uroporphyrinogen decarboxylase [Homo sapiens (human)]
2205	1523	CUX1	Cut-like homeobox 1 [Homo sapiens (human)]

in Table 1. Unfortunately, the use of NN can not improve the classification performance as we expect. Only for the 1-norm SVM, NN improves its classification performance. Obviously, the rank about the CPU time is the same as the previous experiment.

4.2.3 Selected genes by the 1-norm SVMSL

On the Lung cancer dataset, the 1-norm SVMSL can at most select 29 genes based on the training set. However, from Tables 1 and 2, we can see that only three or four genes are enough for discriminating ADCA and MPM, 3 for the SVM classifier and 4 for the NN classifier. Table 3 shows the information on four selected genes, where the first three genes are only for SVM. These genes are protein coding genes from homo sapiens, or human.

Gene selection methods aim to select genes which are less correlated to each other to reduce the redundancy. Statistical methods offer a way to measure a linear correlation between genes. For any gene pair (G_i, G_j), we can compute the correlation coefficient between them. Namely,

$$\rho_{ij} = \frac{Cov(G_i, G_j)}{\sqrt{D(G_i)D(G_j)}} \tag{10}$$

where $Cov(\cdot, \cdot)$ denotes the covariance function, and $D(\cdot)$ denotes the variance function and $0 \leq \rho_{ij} \leq 1$. G_i is linearly uncorrelated to G_j if and only if $\rho_{ij} = 0$.

Based on Table 1, we check the linear correlation of selected genes corresponding to the best performance. To see the coefficient distribution, we discrete the correlation coefficient into five intervals $[0.0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$ and $[0.8, 1)$, and compute the probability in the corresponding interval. Table 4 gives the

coefficient distribution and average correlation coefficient, which is defined as

$$\bar{\rho} = \frac{2}{d(d-1)} \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \rho_{ij} \tag{11}$$

The number of Gene pair in Table 4 is equal to $d(d-1)/2$ and the detail value for d is given in Table 1. The correlation coefficients for the 1-norm SVMSL are less than 0.4, which indicates that this method can select less correlated genes indeed. For Table 2, we have the similar conclusion.

In summary, we have the following conclusions based on the experimental results on the Lung cancer dataset:

- Although SVM-RFE +SVM/NN has the best average recall, which is slightly superior to the 1-norm SVMSL +SVM/NN, the speed of the 1-norm SVMSL is greatly superior to SVM-RFE. In addition, the 1-norm SVMSL +SVM/NN achieves its best performance with a small number of genes, which is useful for speeding the test process and data storage.
- Compared with SVDD-RFE, the 1-norm SVMSL has an advantage over both speed and performance.
- The 1-norm SVMSL has a similar performance to the 1-norm SVM and a much better performance than NLPSVM. However, the 1-norm SVMSL needs less time for selecting genes than these two methods.
- As the subsequent classifier, SVM is much better than NN because SVM improves the average recall of the 1-norm SVMSL.
- The genes selected by the 1-norm SVMSL are mostly less correlated. Some gene pairs selected by other methods have high correlation. For example, 0.79 % gene pairs in SVM-RFE have large correlation coefficients which locates in the interval $[0.8, 0.9)$.

Table 7 Average correlation coefficient and distribution of best gene subsets on the Leukemia-ALLAML dataset

Method	# Gene pair	$\bar{\rho}$	Coefficient distribution (%)				
			0.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
SVM-RFE	120	0.4324	6.67	32.50	40.00	17.50	3.33
SVDD-RFE	15225	0.2990	37.07	32.50	20.53	8.84	0.82
1-norm SVM	231	0.3690	16.02	37.66	34.63	9.52	2.16
NLPSVM	630	0.3957	10.63	37.30	39.84	11.75	0.48
1-norm SVMSL	6	0.1641	0.00	100.00	0.00	0.00	0.00
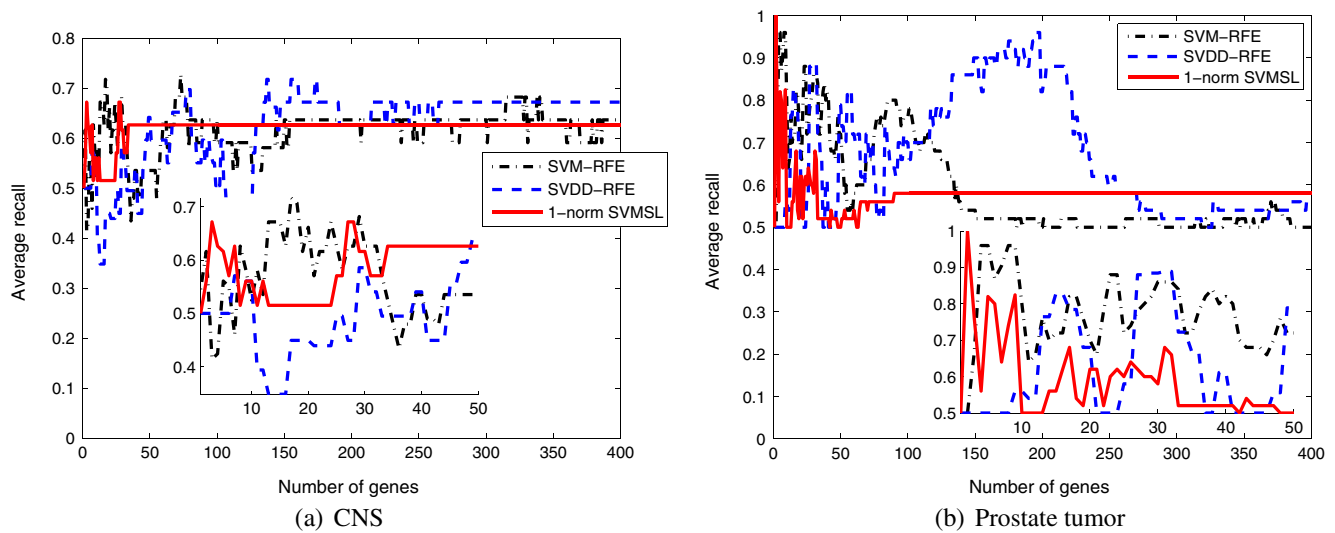
Fig. 5 Average recall vs. gene number when taking SVM as the classifier

4.3 Leukemia-ALLAML

Leukemia-ALLAML dataset has 72 samples belonging to two classes, or ALL (Acute Lymphoblastic Leukemia) and AML (Acute Myeloid Leukemia). The training set consists of 38 bone marrow samples (27 ALL and 11 AML), over 7129 probes from 6817 human genes. In addition, 34 test samples are provided, with 20 ALL and 14 AML.

d varies from 1 to 400. The average recall vs. d obtained by SVM-RFE, SVDD-RFE and the 1-norm SVMSL is shown in Fig. 4. We have the similar conclusion as before. The 1-norm SVMSL achieves its best performance when d is small.

Table 5 shows the best average performance and the corresponding gene number and CPU time for all these methods. Observation on Table 5 also implies that the 1-norm SVMSL can select gene subsets fast. SVM-RFE has the best average recall among all the methods, followed by the 1-norm SVMSL+SVM and the 1-norm SVM. In addition, NN can not improve the performance of the 1-norm SVM,

NLPS-VM and the 1-norm SVMSL, but SVM can improve the performance of both NLPSVM and the 1-norm SVMSL.

On this dataset, the 1-norm SVMSL can at most select 32 genes based on the training set. Among the 32 genes, four genes are enough for discriminating ALL and AML. Table 6 shows the information on the four genes. The first two genes, ZYX and MPO are often presented in visualizations, see http://www.biolab.si/supp/bi-cancer/projections/info/leukemia.htm.

Now, we validate the linear correlation of best gene genes which means subsets with highest average recall obtained by SVM. Table 7 lists the coefficient distribution and average correlation coefficient. The correlation coefficients for the 1-norm SVMSL are all between 0.2 to 0.4, which shows that the genes selected by the 1-norm SVMSL are not strongly correlated.

Experimental conclusions on the Leukemia-ALLAML dataset are consistent with those on the Lung cancer dataset. We do not discuss more.

Table 8 Comparison on the CNS dataset

Method	# Gene	Failure recall(%)	Survivor recall(%)	Average recall (%)	CPU time (s)
None+SVM	Full(7129)	100.00	18.18	59.09	0.0936
SVM-RFE+SVM	73	100.00	45.45	72.73	479.52
SVDD-RFE+SVM	137	88.89	54.55	71.72	96.44
1-norm SVM	30	88.89	27.27	58.08	0.7020
1-norm SVM+SVM	30	88.89	27.27	58.08	0.7020
NLPSVM	159	33.33	63.64	48.48	0.9516
NLPSVM+SVM	159	100.00	18.18	59.09	0.9516
1-norm SVMSL	3	88.89	36.36	62.13	0.0153
1-norm SVMSL+SVM	3	88.89	45.45	67.17	0.0173

Table 9 Comparison on the Prostate tumor dataset

Method	# Gene	Prostate recall(%)	Normal recall(%)	Average recall (%)	CPU time (s)
None+SVM	Full(12600)	100	8.00	54.00	0.4992
SVM-RFE+SVM	4	100.00	92.00	96.00	4.46×10^3
SVDD-RFE+SVM	195	100.00	92.00	96.00	496.08
1-norm SVM	51	100.00	52.00	76.00	2.6364
1-norm SVM+SVM	51	100.00	36.00	68.00	2.6364
NLPSVM	87	100.00	88.00	94.00	5.1012
NLPSVM+SVM	87	100.00	24.00	62.00	5.1012
1-norm SVMSL	2	100.00	84.00	92.00	0.2103
1-norm SVMSL+SVM	2	100.00	100.00	100.00	0.2142

4.4 More datasets

In the following, we perform experiments on another two gene expression datasets, including CNS and Prostate tumor. The two datasets are described as follows:

– CNS dataset contains 60 patient samples, 21 are survivors (alive after treatment) and 39 are failures (succumbed to their disease). There are 7129 genes in the dataset. The training set consists of the first 10 survivors and 30 failures, the other 11 survivors and 9 failures are testing points.

– Prostate tumor dataset has a training set which contains 52 prostate tumor samples and 50 non-tumor (labeled as "Normal") prostate samples with around 12600 genes. An independent set of testing samples is also prepared, which is from a different experiment and has a nearly 10-fold difference in overall microarray intensity from the training data. Besides, we have removed extra genes

contained in the testing samples. There are 25 tumor and 9 normal samples in the test set.

As showed above, NN does not outperform SVM. Thus, we only consider SVM as the subsequent classifier. The setting of parameters is the same as before.

The average recall vs. d obtained by SVM-RFE, SVDD-RFE and the 1-norm SVMSL is shown in Fig. 5. If the selected gene number is big enough, SVDD-RFE can obtained good performance on the CNS and Prostate datasets. For the 1-norm SVMSL, its best performance still can be achieved when d is small.

Tables 8 and 9 show the experimental results on the two datasets, respectively. If we do not perform gene selection, the average recall obtained by SVM is very low on both datasets. SVDD-RFE is compared to SVM-RFE when SVDD-RFE selects more genes. On the two datasets, the 1-norm SVMSL+SVM is much better than other 1-norm SVM-like methods.

(a) CNS (b) Prostate tumor
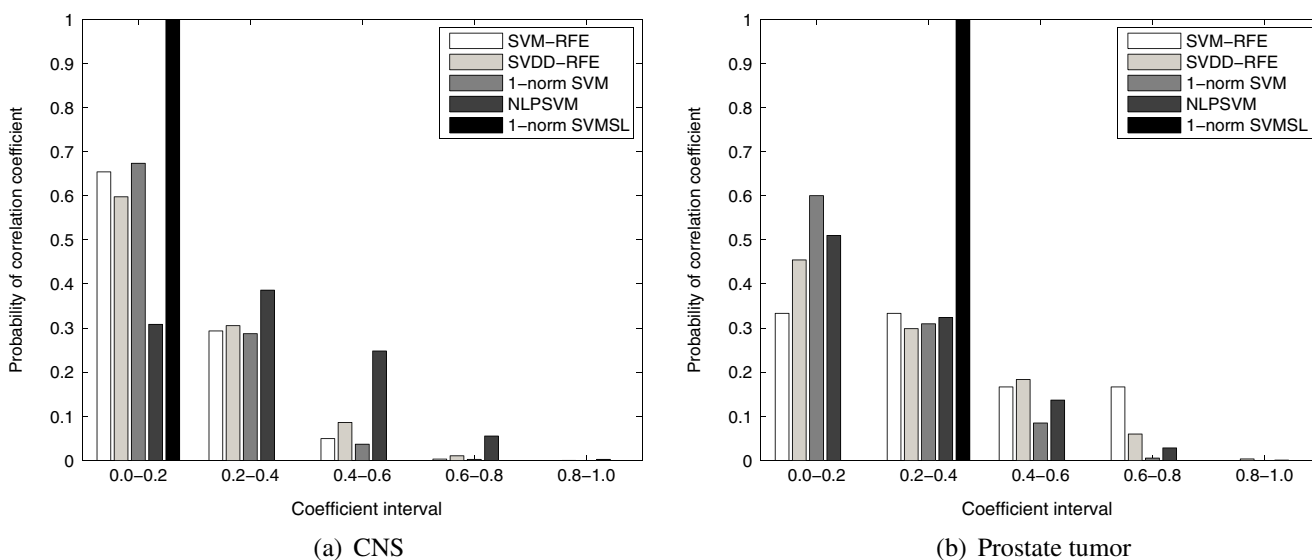
Fig. 6 Correlation coefficient distribution

The correlation coefficient between selected genes is also shown in Fig. 6. Similarly, the linear correlation between genes selected by the 1-norm SVMSL is very low on both the CNS and Prostate tumor datasets.

5 Conclusions

This paper proposes a new strategy for cancer classification based on gene expression data. A fast gene selection way is implemented by using the 1-norm SVMSL. The 1-norm SVMSL is an variant of the 1-norm SVM. Although the 1-norm SVMSL can perform gene selection and classification at the same time, we use OMP to approximate its solution which may be not so exact to perform classification. Thus, a subsequent classifier is used to discriminating genes selected by the 1-norm SVMSL. In theory, any classifier dealing with cancer classification can be used as the subsequent classifier. Experimental results shows SVM is a proper one.

In our experiments, we compare the 1-norm SVMSL with SVM-RFE, SVDD-RFE, the 1-norm SVM and NLPSVM on four datasets. SVM-RFE outperforms other methods on classification performance in three of four datasets, but it takes a long time to rank genes, say more than 1,000 sec. for 12533 genes. Note that the 1-norm SVMSL is at least four order of magnitude faster than SVM-RFE. In addition, the 1-norm SVMSL+SVM has an acceptable performance by reference to SVM-RFE+SVM. Compared to the 1-norm SVM and NLPSVM, the 1-norm SVMSL is almost an order of magnitude faster than it. On four datasets, the 1-norm SVMSL is better than or compared to both the 1-norm SVM and NLPSVM. In addition, we test the linear correlation between the selected genes. Among these gene selection methods, the linear correlation is the weakest for the 1-norm SVMSL. In other words, genes selected by the 1-norm SVMSL are not redundant and less correlated.

It is notable that the 1-norm SVMSL only needs a small number of genes to achieve its best performance, which would contribute to fast test and storage. However, we can not determine how many genes we should select, or determine the value of d in the 1-norm SVMSL. If we select all which the 1-norm SVMSL generates, we can not guarantee the classification performance, especially in case of the Prostate tumor dataset. In the future, we try to find a way to determine the optimal gene number for the 1-norm SVMSL.

References

1. Becker S, Bobin J, Candès E (2009) NESTA: A fast and accurate first-order method for sparse recovery. SIAM J Imaging Sci 4(1):1–39
2. Bennett KP (1999) Combining support vector and mathematical programming methods for classification. In: Schölkopf B, Burges CJC, Smola AJ (eds) Advances in Kernel methods: support vector learning. MIT Press, Cambridge, pp 307–326
3. Bi J, Bennett KP, Embrechts M, Breneman CM, Song M (2003) Dimensionality reduction via sparse support vector machines. J Mach Learn Res 3:1229–1243
4. Cao J, Zhang L, Wang B, Li F, Yang J (2015) A fast gene selection method for multi-cancer classification using multiple support vector data description. J Biomed Inform 53(2):381–389
5. Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/cjlin/libsvm
6. Cui Y, Zheng CH, Yang J, Sha W (2013) Sparse maximum margin discriminant analysis for feature extraction and gene selection on gene expression data. Comput Biol Med 43(7):933–941
7. Davis G, Mallat S, Avellaneda M (1997) Greedy adaptive approximation. J Construtive Approx 13:57–98
8. Demiriz A, Bennett KP, Shawe-Taylor J (2002) Linear programming Boosting via column generation. Mach Learn 46(1):225–254
9. Donoho D, Elad M, Temlyakov V (2006) Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans Inf Theory 52:6–18
10. Duan KB, Rajapakse JC, Wang H, Azuaje F (2005) Multiple SVM-RFE for gene selection in cancer classification with expression data. IEEE Trans Nanobioscience 4(3):228–234
11. Dudoit S, Fridlyand J, Speed TP (2002) Comparison of discrimination methods for the classification of tumors using gene expression data. J Amer Stat Assoc 97(457):77–87
12. Fung GM, Mangasarian OL (2004) A feature selection Newton method for support vector machine classification. Comput Optim Appl 28:185–202
13. Girosi F (1998) An equivalence between sparse approximation and support vector machines. Neural Comput 10(6):1455–1480
14. Golub T, Slonim D, Tamayo P, Huard C, Gaasenbeek M, Mesirov J, Coller H, Loh M, Downing J, Caligiuri M, Bloomfield C, Lander ES (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286(5436):531–537
15. Gordon G, Jensen R, Hsiao L, Gullans S, Blumenstock J, Ramaswamy S, Richards W, Sugarbaker D, Bueno R (2002) Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Res 62(17):4963–4937
16. Guyon I, Weston J, Barnhill S, Vapink V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46(1-3):389–422
17. Lee C, Leu Y (2011) A novel hybrid feature selection method for microarray data analysis. Appl Soft Comput 11(1):208–213
18. Li JT, Jia YM, Li WL (2011) Adaptive huberized support vector machine and its application to microarray classification. Neural Comput Appl 20(1):123–132
19. Li L, Weinberg CR, Darden TA, Pedersen LG (2001) Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method. Bioinformatics 17(12):1131–1142
20. Liu H, Liu L, Zhang H (2010) Ensemble gene selection by grouping for microarray data classification. J Biomed Inform 43(1):81–87
21. Makh A (2012) GLPK (GNU linear programming kit). http://www.gnu.org/software/glpk/glpk.html

22. Maldonado S, Montoya R, Lpez J (2017) Embedded heterogeneous feature selection for conjoint analysis: A SVM approach using L1 penalty. Appl Intell 46:775–787

23. Mangasarian OL (2000) Generalized support vector machines. In: Smola A, Bartlett P, Schölkopf B, Schuurmans D (eds) Advances in large margin classifiers. MIT Press, Cambridge, pp 135–146

24. Mangasarian OL (2006) Exact 1-norm support vector machines via unconstrained convex differentiable minimization. J Mach Learn Res 7:1517–1530

25. Pomeroy S, Tamayo P, Gaasenbeek M, Sturla L, Angelo M, McLaughlin M, Kim J, Goumnerova L, Black P, Lau C, Allen J, Zagzag D, Olson J, Curran T, Wetmore C, Biegel J, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis D, Mesirov J, Lander E, Golub T (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 415(6870):436–442

26. Shah S, Kusiak A (2007) Cancer gene search with data-mining and genetic algorithms. Comput Biol Med 37(2):251–261

27. Shen Q, Mei Z, Ye BX (2009) Simultaneous genes and training samples selection by modified particle swarm optimization for gene expression data classification. Comput Biol Med 39(7):646–649

28. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP (2002) Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 1(2):203–209

29. Steinwart I (2003) Sparseness of support vector machines. J Mach Learn Res 4(6):1071–1105

30. Stodden V, Carlin L, David Donoho EA (2007) SparseLab: Seeking sparse solutions to linear systems of equations. http://sparselab.stanford.edu/

31. Thi HAL, Tao PD, Thiao M (2016) Efficient approaches for L2-L0 regularization and applications to feature selection in SVM. Appl Intell 45(2):1–17

32. Tropp J, Gilbert A (2007) Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inform Theroy 53(12):4655–4666

33. Van't Veer LJ, Dai H, van de Vijver M, He Y, Hart A, Mao M, Peterse H, Van der Kooy K, Marton M, Witteveen A, Schreiber G, Kerkhoven R, Roberts C, Linsley P, Bernards R, Friend S (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530–536

34. Vapnik V (1998) Statistical learning theory. Wiley, New York

35. Vapnik V (1999) The overview of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999

36. Wang C, Cao L, Miao B (2013) Optimal feature selection for sparse linear discriminant analysis and its applications in gene expression data. Comput Stat Data Anal 66(10):140–149

37. Wang HQ, Wong HS, Zhu H, Yip TT (2009) A neural network-based biomarker association information extraction approach for cancer classification. J Biomed Inform 42(4):654–666

38. Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V (2000) Feature selection for SVMs. NIPS 12:668–672

39. Wong HS, Wang HQ (2008) Constructing the gene regulation-level representation of microarray data for cancer classification. J Biomed Inform 41(1):95–105

40. Zhang L, Huang X (2015) Multiple SVM-RFE for multi-class gene selection on dna microarray data. In: Proceedings of 2015 international joint conference on neural networks, pp 897–902

41. Zhang L, Zhou W (2013) Analysis of programming properties and the rowccolumn generation method for 1-norm support vector machines. Neural Netwx 48(12):32–43

42. Zhang L, Zhou W, Zhang Z, Yang J (2015) A fast approximation algorithm for 1-norm svm with squared loss. In: Proceedings of 2015 IEEE International Joint Conference on Neural Networks (IJCNN)

43. Zhang L, Zhou W (2010) On the sparseness of 1-norm support vector machines. Neural Netw 23(3):373–385

44. Zhang L, Zhou W (2013) A fast algorithm for kernel 1-norm support vector machines. Knowl-Based Syst 52(16):223–235

45. Zhou W, Zhang L, Jiao L (2002) Linear programming support vector machines. Pattern Recogn 35(12):2927–2936

46. Zhou X, Tuck DP (2007) MSVM-RFE: Extensions of SVM-RFE for multi-class gene selection on dna microarray data. Bioinformatics 23(9):1106C1114

47. Zhu J, Rosset S, Hastie T, Tibshirani R (2004) 1-norm support vector machines. In: Thrun S, Saul L, Schölkopf B (eds) Advances in Neural Information Processing Systems 16. MIT Press, Cambridge, pp 49–56

Li Zhang received the B.S. degree in 1997 and the Ph.D. degree in 2002 in electronic engineering from Xidian University, Xi'an, China. Now she is a full professor with the School of Computer Science and Technology, Soochow University, Suzhou, China. She was a postdoctor at the Institute of Automation, Shanghai Jiao Tong University, Shanghai, China, from 2003 to 2005. She worked as an associate professor at the Institute of Intelligent Information Processing, Xidian University, Xi'an, China, from 2005 to 2010. She was a visiting professor at Yuan Ze University, Taiwan, from February to May 2010. She has authored/co-authored more than 100 technical papers published in journals and conferences. Her research interests have been in the areas of machine learning, pattern recognition, neural networks and intelligent information processing.

Weida Zhou received the B.S. in 1996 and the Ph.D. degree in 2003 in electronic engineering from Xidian University, Xi'an, China. He was an associate professor at the School of Electronic Engineering at Xidian University, Xi'an, China from 2006 to 2009. From 2009, he works for AI Speech Ltd., Suzhou 215123, Jiangsu, China. His research interests include machine learning, learning theory and intelligent information processing.

Bangjun Wang received the M.S. degree in information management from the Wuhan University of Technology, Wuhan, in 2004. From 2004, he worked in Soochow University. He is currently working toward the Ph.D. degree in Beijing Jiaotong University, Beijing, China. His research interests include machine learning and Layered learning.

Fanzhang Li received the Master's degree in engineering from the Department of Computer Science, University of Science and Technology, Hefei, China, in 1999. He is currently a Full Professor with the School of Computer Science and Technology, Soochow University, Suzhou, China, and is also an Adjunct Professor with Beijing Jiaotong University, China. His research interests include Lie Group machine learning, data mining, and dynamic fuzzy logic.

Zhao Zhang received the Ph.D. degree in computer engineering from the Department of Electronic Engineering (EE), City University of Hong Kong, Kowloon Tong, China, in 2013. Currently, he is an Associate Professor with the School of Computer Science and Technology, Soochow University, Suzhou, China. He was a Visiting Research Engineer at the Learning and Vision Research Group, National University of Singapore, Singapore, from February to May 2012. He then visited the National Laboratory of Pattern Recognition (NLPR), Chinese Academy of Sciences, Beijing, China, from September to December 2012. He has authored/co-authored more than 35 technical papers published in prestigious international journals and conferences, including six IEEE/Association for Computing Machinery Transactions articles. His research interests include pattern recognition, machine learning and data mining, and computer vision.