Laplacain Pair-Weight Vector Projection with Adaptive Neighbor Graph for Semi-supervised Learning

Yangtao Xue and Li Zhang[✉]

The School of Computer Science and Technology and Joint International Research Laboratory of Machine Learning and Neuromorphic Computing, Soochow University, Suzhou 215006, China
zhangliml@suda.edu.cn

Abstract. Recently, Laplacian pair-weight vector projection (LapPVP) algorithm was proposed for semi-supervised classification. Although Lap-PVP achieves a good classification performance for semi-supervised learning, it may be sensitive to noise and outliers for using the neighbor graph with a fixed similarity matrix. To remedy it, this paper proposes a novel method named Laplacain pair-weight vector projection with adaptive neighbor graph (ANG-LapPVP), in which the graph induced by the Laplacian manifold regularization is adaptively constructed by solving an optimization problem. For binary classification problems, ANG-LapPVP learns a pair of projection vectors by solving the pair-wise optimal formulations in which we maximize the between-class scatter and minimize both the within-class scatter and the adaptive neighbor graph (ANG) regularization. The ANG regularization is to learn the ANG whose similarity matrix varies with iterations, which may solve the issue of LapPVP. Thus, ANG-LapPVP simultaneously learns adaptive similarity matrices and a pair of projection vectors with an iterative process. Experimental results on an artificial and real-world benchmark datasets show the superiority of ANG-LapPVP compared to the related methods. Thus, ANG-LapPVP is promising in semi-supervised learning.

Keywords: Semi-supervised learning · Binary classification · Manifold regularization · Adaptive neighbor graph

1 Introduction

Recently, more and more research attention acts in for semi-supervised classification tasks. On the basis of graph theory, the manifold regularization is becoming a popular technology to extend supervised learners to semi-supervised ones [2,4,6]. Considering the structure information provided by unlabeled data, semi-supervised learners outperform related supervised ones and have more applications in reality [1,3,12,15].

Because the performance of semi-supervised learners partly depends on the corresponding supervised ones, we need to choose a good supervised learner to

construct a semi-supervised one. Presently, inspired by the idea of non-parallel planes, many supervised algorithms have been designed for dealing with binary classification problems, such as generalized proximal support vector machine (GEPSVM) [9,16], twin support vector machine (TSVM) [7], least squares twin support vector machine (LSTSVM) [8], multi-weight vector support vector machine (MVSVM) [18], and enhanced multi-weight vector projection support vector machine (EMVSVM) [17]. In these learners with non-parallel planes, each plane is as close as possible to samples from its own class and meanwhile as far as possible from samples belonging to the other class. Owing to the outstanding generalization performance of TSVM and LSTSVM, they have been extended to semi-supervised learning by using the manifold regularization framework [3,12]. In [12], Laplacian twin support vector machine (LapTSVM) constructs a more reasonable classifier from labeled and unlabeled data by integrating the manifold regularization. Chen et al. [3] proposed Laplacian least squares twin support vector machine (LapLSTSVM) based on LapTSVM and LSTSVM. Different from LapTSVM, LapLSTVM needs to solve only two systems of linear equations with remarkably less computational time. These semi-supervised learners have proved that the manifold regularization is a reasonable and effective technology. On the basis of EMVSVM and manifold regularization, Laplacian pair-weight vector projection (LapPVP) was extended to semi-supervised learning for binary classification problems [14]. LapPVP achieves a pair of projection vectors by maximizing the between-class scatter and minimizing the within-class scatter and manifold regularization.

These performance of semi-supervised learners is partly related to the neighbor graph induced by the manifold regularization. Generally, the neighbor graph is predefined and may be sensitive to noise and outliers [10]. To improve the robustness of LapPVP, we propose a novel semi-supervised learner, named Laplacain pair-weight vector projection with adaptive neighbor graph (ANG-LapPVP). ANG-LapPVP learns a pair of projection vectors by solving the pair-wise optimal formulations, where we maximize the between-class scatter and minimize both the within-class scatter and the adaptive neighbor graph (ANG) regularization. In ANG, the similarity matrix is not fixed but adaptively learned on both labeled and unlabeled data by solving an optimization problem [11,19,20]. Moreover, the between- and within-class scatter matrices are computed separably for each class, which can strengthen the discriminant capability of ANG-LapPVP. Therefore, it is easy for ANG-LapPVP to handle binary classification tasks and achieve a good performance.

2   Proposed Method

The propose method ANG-LapPVP is an enhanced version of LapPVP. In ANG-LapPVP, we learn an ANG based on the assumption that the smaller the distance between data points is, the greater the probability of being neighbors is. Likewise, ANG-LapPVP is to find a pair of the projection vectors by maximizing between-class scatter and minimizing both the within-class scatter and the ANG regularization.

Let $\mathbf{X} = [\mathbf{X}_\ell; \mathbf{X}_u] \in \mathbb{R}^{n \times m}$ be the training sample matrix, where n and m are the number of total samples and features, respectively; $\mathbf{X}_\ell \in \mathbb{R}^{\ell \times m}$ and $\mathbf{X}_u \in \mathbb{R}^{u \times m}$ are the labeled and unlabeled sample matrices, respectively; ℓ and u are the number of labeled and unlabeled samples, respectively, and $n = \ell + u$. For convenience, we use y_i to describe the label situation of sample $\mathbf{x}_i$. If $y_i = 1$, $\mathbf{x}_i$ is a labeled and positive sample; if $y_i = -1$, $\mathbf{x}_i$ is a labeled and negative sample; if $y_i = 0$, $\mathbf{x}_i$ is unlabeled. Furthermore, the labeled sample matrix $\mathbf{X}_\ell$ can be represented as $\mathbf{X}_\ell = [\mathbf{X}_1; \mathbf{X}_2]$, where $\mathbf{X}_1 = [\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1\ell_1}]^T \in \mathbb{R}^{\ell_1 \times m}$ is the positive sample matrix with a label of 1, $\mathbf{X}_2 = [\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2\ell_2}]^T \in \mathbb{R}^{\ell_2 \times m}$ is the negative sample matrix with a label of -1, $\ell = \ell_1 + \ell_2$, ℓ_1 and ℓ_2 are the number of positive and negative samples, respectively.

2.1 Formulations of ANG-LapPVP

For binary classification tasks, the goal of ANG-LapPVP is to find a pair of projection vectors similar to LapPVP. As mentioned above, the proposed ANG-LapPVP is an enhanced version of LapPVP. To better describe our method, we first briefly introduce LapPVP [14]. For the positive class, LapPVP is to solve the following optimization problem:

$$
\begin{aligned}
\max_{\mathbf{v}_1} \quad & \mathbf{v}_1^T \mathbf{B}_1 \mathbf{v}_1 - \alpha_1 \mathbf{v}_1^T \mathbf{W}_1 \mathbf{v}_1 - \beta_1 \mathbf{v}_1^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{v}_1 \\
s.t. \quad & \mathbf{v}_1^T \mathbf{v}_1 = 1
\end{aligned}
\tag{1}
$$

where $\alpha_1 > 0$ and $\beta_1 > 0$ are regularization parameters, $\mathbf{L}$ is the Laplacian matrix of all training data, and $\mathbf{B}_1$ is the between-class scatter matrix and $\mathbf{W}_1$ is the within-class scatter matrix of the positive class, which can be calculated by

$$
\mathbf{B}_1 = \left(\mathbf{X} - \mathbf{e}\mathbf{u}_1^T\right)^T \left(\mathbf{X} - \mathbf{e}\mathbf{u}_1^T\right)
\tag{2}
$$

and

$$
\mathbf{W}_1 = \left(\mathbf{X}_1 - \mathbf{e}_1\mathbf{u}_1^T\right)^T \left(\mathbf{X}_1 - \mathbf{e}_1\mathbf{u}_1^T\right)
\tag{3}
$$

where $\mathbf{v}_1 \in \mathbb{R}^m$ is the projection vector of the positive class, $\mathbf{u}_1 = \frac{1}{\ell_1}\sum_{i=1}^{\ell_1} \mathbf{x}_{1i}$ is the mean vector of the positive samples, $\mathbf{e}_1 \in \mathbb{R}^{\ell_1}$ and $\mathbf{e} \in \mathbb{R}^\ell$ are the vectors of all ones with different length.

In the optimization problem (1), the laplacian matrix $\mathbf{L}$ is computed in advance and is independent of the objective function. The concept of ANG was proposed in [11], which has been applied to feature selection for unsupervised multi-view learning [19] and semi-supervised learning [20]. We incorporate this concept into LapPVP and form ANG-LapPVA.

Similarly, the pair of projection vectors of ANG-LapPVP is achieved by a pair-wise optimal formulations. On the basis of (1), the optimal formulation of ANG-LapPVP for the positive class is defined as:

$$
\begin{aligned}
\max_{\mathbf{v}_1, \mathbf{S}_1} \quad & \mathbf{v}_1^T \mathbf{B}_1 \mathbf{v}_1 - \alpha_1 \mathbf{v}_1^T \mathbf{W}_1 \mathbf{v}_1 - \beta_1 (\mathbf{v}_1^T \mathbf{X}^T \mathbf{L}_{s_1} \mathbf{X} \mathbf{v}_1 + \gamma_1 \mathbf{S}_1^T \mathbf{S}_1) \\
s.t. \quad & \mathbf{v}_1^T \mathbf{v}_1 = 1, \quad \mathbf{S}_1 \mathbf{e} = \mathbf{e}, \quad \mathbf{S}_1 > 0
\end{aligned}
\tag{4}
$$

where $\mathbf{S}_1$ is the similarity matrix for the positive class, $\mathbf{L}_{s_1}$ is the Laplacian matrix related to $\mathbf{S}_1$ for the positive class, and $\gamma_1 > 0$ is a regularization parameter.

Compared with (1), (4) has a different term, or the third term, which is called the ANG regularization here. $\mathbf{S}_1$ varies with iterations, and then the Laplacian matrix $\mathbf{L}_{s_1} = \mathbf{D}_{s_1} - \mathbf{S}_1$ is changed, where $\mathbf{D}_{s_1}$ is a diagonal matrix with diagonal elements of $(D_{s_1})_{ii} = \sum_j (S_1)_{ij}$. The first and second terms represent the between- and within-class scatters of the positive class, and the regularization parameter α_1 is to make a balance between these two scatters. ANG-LapPVP can keep data points as near as possible in the same class while as far as possible from the other class by maximizing the between-class scatter and minimizing the within-class scatter.

For the negative class, ANG-LapPVP has the following similar problem:

$$\max_{\mathbf{v}_2, \mathbf{S}_2} \mathbf{v}_2^T \mathbf{B}_2 \mathbf{v}_2 - \alpha_2 \mathbf{v}_2^T \mathbf{W}_2 \mathbf{v}_2 - \beta_2 (\mathbf{v}_2^T \mathbf{X}^T \mathbf{L}_{s_2} \mathbf{X} \mathbf{v}_2 + \gamma_2 \mathbf{S}_2^T \mathbf{S}_2)$$
$$s.t. \quad \mathbf{v}_2^T \mathbf{v}_2 = 1, \quad \mathbf{S}_2 \mathbf{e} = \mathbf{e}, \quad \mathbf{S}_2 > 0 \tag{5}$$

where $\mathbf{v}_2$ is the projection vector for the negative class, α_2, β_2, and γ_2 are positive regularization parameters, $\mathbf{S}_2$ is the similarity matrix for the negative class, $\mathbf{L}_{s_2}$ is the Laplacian matrix related to $\mathbf{S}_2$, $\mathbf{B}_2$ and $\mathbf{W}_2$ are the between- and within-class scatter matrices for the negative class, respectively, which can be written as:

$$\mathbf{B}_2 = \left(\mathbf{X} - \mathbf{e}\mathbf{u}_2^T\right)^T \left(\mathbf{X} - \mathbf{e}\mathbf{u}_2^T\right) \tag{6}$$

and

$$\mathbf{W}_2 = \left(\mathbf{X}_2 - \mathbf{e}_2\mathbf{u}_2^T\right)^T \left(\mathbf{X}_2 - \mathbf{e}_2\mathbf{u}_2^T\right) \tag{7}$$

where $\mathbf{u}_2 = \frac{1}{\ell_2} \sum_{i=1}^{\ell_2} \mathbf{x}_{2i}$ is the mean vector of negative samples, $\mathbf{e}_2 \in \mathbb{R}^{\ell_c}$ is the vector of all ones.

2.2   Optimization of ANG-LapPVP

Problems (4) and (5) form the pair of optimization problems for ANG-LapPVP, where projection vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ and similarity matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ are unknown. It is difficult to find the optimal solution to them at the same time. Thus, we use an alternative optimization approach to solve (4) or (5). During the optimization procedure, we would fix a set of variables and solve the other set of ones.

When $\mathbf{S}_1$ and $\mathbf{S}_2$ are fixed, the optimization formulations of ANG-LapPVP can be reduced to

$$\max_{\mathbf{v}_1} \quad \mathbf{v}_1^T \mathbf{B}_1 \mathbf{v}_1 - \alpha_1 \mathbf{v}_1^T \mathbf{W}_1 \mathbf{v}_1 - \beta_1 \mathbf{v}_1^T \mathbf{X}^T \mathbf{L}_{s_1} \mathbf{X} \mathbf{v}_1$$
$$s.t. \quad \mathbf{v}_1^T \mathbf{v}_1 = 1 \tag{8}$$

and

$$\max_{\mathbf{v}_2} \quad \mathbf{v}_2^T \mathbf{B}_2 \mathbf{v}_2 - \alpha_2 \mathbf{v}_2^T \mathbf{W}_2 \mathbf{v}_2 - \beta_2 \mathbf{v}_2^T \mathbf{X}^T \mathbf{L}_{s_2} \mathbf{X} \mathbf{v}_2$$
$$s.t. \quad \mathbf{v}_2^T \mathbf{v}_2 = 1 \tag{9}$$

which are exactly LapPVP.

According to the way in [14], we can find the solutions $\mathbf{v}_1$ and $\mathbf{v}_2$ to (8) and (9), respectively. In [14], (8) and (9) can be respectively converted to the following eigenvalue decomposition problems:

$$
\begin{aligned}
\mathbf{B}_1\mathbf{v}_1 - \alpha_1\mathbf{W}_1\mathbf{v}_1 - \beta_1\mathbf{X}^T\mathbf{L}_{s_1}\mathbf{X}\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \\
\mathbf{B}_2\mathbf{v}_2 - \alpha_2\mathbf{W}_2\mathbf{v}_2 - \beta_2\mathbf{X}^T\mathbf{L}_{s_2}\mathbf{X}\mathbf{v}_2 = \lambda_2\mathbf{v}_2
\end{aligned}
\tag{10}
$$

where λ_1 and λ_2 are eigenvalues for the positive and negative classes, respectively. Thus, the optimal solutions here are the eigenvectors corresponding to the largest eigenvalues.

Once we get the projection vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, we take them as fixed variables, then solve $\mathbf{S}_1$ and $\mathbf{S}_2$. In this case, the optimization formulations of ANG-LapPVP are reduced to

$$
\begin{aligned}
\min_{\mathbf{S}_1} \quad & \mathbf{v}_1^T\mathbf{X}^T\mathbf{L}_{s_1}\mathbf{X}\mathbf{v}_1 + \gamma_1\mathbf{S}_1^T\mathbf{S}_1 \\
s.t. \quad & \mathbf{S}_1\mathbf{e} = \mathbf{e}, \quad \mathbf{S}_1 > 0
\end{aligned}
\tag{11}
$$

and

$$
\begin{aligned}
\min_{\mathbf{S}_2} \quad & \mathbf{v}_2^T\mathbf{X}^T\mathbf{L}_{s_2}\mathbf{X}\mathbf{v}_2 + \gamma_2\mathbf{S}_2^T\mathbf{S}_2 \\
s.t. \quad & \mathbf{S}_2\mathbf{e} = \mathbf{e}, \quad \mathbf{S}_2 > 0
\end{aligned}
\tag{12}
$$

For simplicity, let $(Z_1)_{ij} = ||\mathbf{v}_1^T\mathbf{x}_i - \mathbf{v}_1^T\mathbf{x}_j||^2$ and $(Z_2)_{ij} = ||\mathbf{v}_2^T\mathbf{x}_i - \mathbf{v}_2^T\mathbf{x}_j||^2$. Then matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are constant when $\mathbf{v}_1$ and $\mathbf{v}_2$ are fixed. Thus, (11) and (12) can be rewritten as:

$$
\begin{aligned}
\min_{\mathbf{S}_1} \quad & \frac{1}{2}(\mathbf{S}_1 + \frac{1}{2\gamma_1}\mathbf{Z}_1)^T(\mathbf{S}_1 + \frac{1}{2\gamma_1}\mathbf{Z}_1) \\
s.t. \quad & \mathbf{S}_1\mathbf{e} = \mathbf{e}, \quad \mathbf{S}_1 > 0
\end{aligned}
\tag{13}
$$

and

$$
\begin{aligned}
\min_{\mathbf{S}_2} \quad & \frac{1}{2}(\mathbf{S}_2 + \frac{1}{2\gamma_2}\mathbf{Z}_2)^T(\mathbf{S}_2 + \frac{1}{2\gamma_2}\mathbf{Z}_2) \\
s.t. \quad & \mathbf{S}_2\mathbf{e} = \mathbf{e}, \quad \mathbf{S}_2 > 0
\end{aligned}
\tag{14}
$$

Because (13) and (14) are similar, we describe the optimization procedure only for (13). First, we generate the Lagrangian function of (13) with multipliers δ_1 and ζ_1 as follows:

$$
L(\mathbf{S}_1, \delta_1, \zeta_1) = \left(\mathbf{S}_1 + \frac{1}{2\gamma_1}\mathbf{Z}_1\right)^T\left(\mathbf{S}_1 + \frac{1}{2\gamma_1}\mathbf{Z}_1\right) - \delta_1(\mathbf{S_1e} - \mathbf{e}) - \zeta_1\mathbf{S}_1 \tag{15}
$$

According to the KKT condition [13], we derive the partial derivative of $L(\mathbf{S}_1, \delta_1, \zeta_1)$ with respect to the primal variables $\mathbf{S}_1$ and make it vanish, which results in

$$
\mathbf{S}_1 = \delta_1 + \zeta_1 - \frac{1}{2\gamma_1}\mathbf{Z}_1 \tag{16}
$$

Similarly, the similarity matrix $\mathbf{S}_2$ is achieved by

$$\mathbf{S}_2 = \delta_2 + \zeta_2 - \frac{1}{2\gamma_2}\mathbf{Z}_2 \tag{17}$$

where δ_2 and ζ_2 are positive Lagrange multipliers.

Since the constraint $\mathbf{Se} = \mathbf{e}$, we have

$$\delta_1 + \zeta_1 = \frac{1}{n} + \frac{1}{2n\gamma_1}\mathbf{Z}_1 \tag{18}$$

and

$$\delta_2 + \zeta_2 = \frac{1}{n} + \frac{1}{2n\gamma_2}\mathbf{Z}_2 \tag{19}$$

where the parameters γ_1 and γ_2 can be computed as follows [11]:

$$\gamma_1 = \frac{1}{n}\mathbf{e}^T\left(\frac{k}{2}\mathbf{Z}_1\mathbf{q}_k - \frac{1}{2}\mathbf{Z}_1\widetilde{\mathbf{q}}_{k-1}\right) \tag{20}$$

and

$$\gamma_2 = \frac{1}{n}\mathbf{e}^T\left(\frac{k}{2}\mathbf{Z}_1\mathbf{q}_k - \frac{1}{2}\mathbf{Z}_1\widetilde{\mathbf{q}}_{k-1}\right) \tag{21}$$

where k is the neighbor number in the graph, $\mathbf{q}_k$ and $\widetilde{\mathbf{q}}_{k-1}$ are indicator vectors. We set $\mathbf{q}_k = [0, 0, \cdots, 0, 1, 0, \cdots, 0, 0]^T \in \mathbb{R}^n$ in which the k-th element is one and the others are zero and $\widetilde{\mathbf{q}}_{k-1} = [1, 1, \cdots 1, 0, \cdots, 0, 0]^T \in \mathbb{R}^n$ in which the first $(k-1)$ elements are one and the others are zero.

2.3   Strategy of Classification

The pair of projection vectors $(\mathbf{v}_1, \mathbf{v}_2)$ can project data points into two different subspaces. The distance measurement is a reasonable way to estimate the class label of an unknown data point $\mathbf{x} \in \mathbb{R}^m$. Here, we define the strategy of classification using the minimum distance.

For an unknown point $\mathbf{x}$, we project it into two subspaces induced by $\mathbf{v}_1$ and $\mathbf{v}_2$. In the subspace induced by $\mathbf{v}_1$, the projection distance between $\mathbf{x}$ and positive samples is defined as:

$$d_1 = \min_{i=1,2,\cdots,\ell_1}\left(\mathbf{v}_1^T\mathbf{x} - \mathbf{v}_1^T\mathbf{x}_{1i}\right)^2 \tag{22}$$

In the subspace induced by $\mathbf{v}_2$, the projection distance between $\mathbf{x}$ and negative samples is computed as:

$$d_2 = \min_{i=1,2,\cdots,\ell_2}\left(\mathbf{v}_2^T\mathbf{x} - \mathbf{v}_2^T\mathbf{x}_{2i}\right)^2 \tag{23}$$

It is reasonable that $\mathbf{x}$ is taken as the positive point if $d_1 < d_2$, which is the minimum distance strategy. Thus, we assign a label to $\mathbf{x}$ by the following rule:

$$\hat{y} = \begin{cases} 1, & if \;\; d_1 \leq d_2 \\ -1, & Otherwise \end{cases} \tag{24}$$

2.4 Computational Complexity

Here, we analyze the computational complexity of ANG-LapPVP. Problems (4) and (5) are non-convex. The ultimate optimal pair-wise projection vectors are obtained by applying an iterative method. In iterations, the optimization problems of ANG-LapPVP can be decomposed into eigenvalue decomposition ones and quadratic programming ones with constraints.

The computational complexities of an eigenvalue decomposition problem and a quadratic programming one are $O\left(m^2\right)$ and $O\left(n^2\right)$, respectively, where m is the number of features and n is the number of samples. Let t be the iteration times. Then, the total computational complexity of ANG-LapPVP is $O\left(t\left(m^2 + n^2\right)\right)$. In the iteration process of ANG-LapPVP, the convergence condition is set as the difference between current and previous projection vectors, i.e., $||\mathbf{v}_1^t - \mathbf{v}_1^{t-1}|| \leq 0.001$ or $||\mathbf{v}_2^t - \mathbf{v}_2^{t-1}|| \leq 0.001$.

3 Experiments

We conduct experiments in this section. Firstly, we compare ANG-LapPVP with LapPVP on an artificial dataset to illustrate the improvement achieved by the ANG regularization. The comparison with other non-parallel planes algorithms on benchmark datasets is then implemented to analyze the performance of ANG-LapPVP.

3.1 Experiments on Artificial Dataset

An artificial dataset, called CrossPlane, is generated by perturbing points originally lying on two intersecting planes. CrossPlane contains 400 instances with only 2 labeled and 198 unlabeled ones for each class. The distribution of Cross-Plane is shown in Fig. 1. Obviously, some data points belonging to Class +1 are surrounded by the data points of Class −1 and vice versa.

Figure 2 plots the projection vectors learned by LapPVP and ANG-LapPVP. We can see that projection vectors learned by ANG-LapPVP are more suitable than LapPVP. The accuracy of LapPVP is 91.50%, and that of ANG-LapPVP is 97.00%. Clearly, ANG-LapPVP has a better classification performance on the CrossPlane dataset. In other words, ANG-LapPVP is robust to noise and outliers. In to all, the ANG regularization can improve the performance of LapPVP, which makes ANG-LapPVP better.
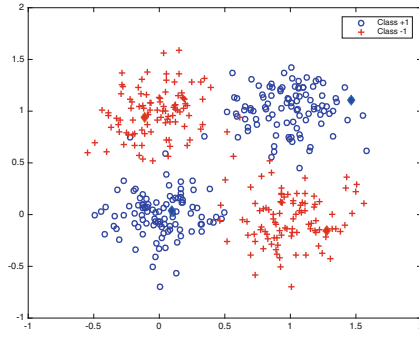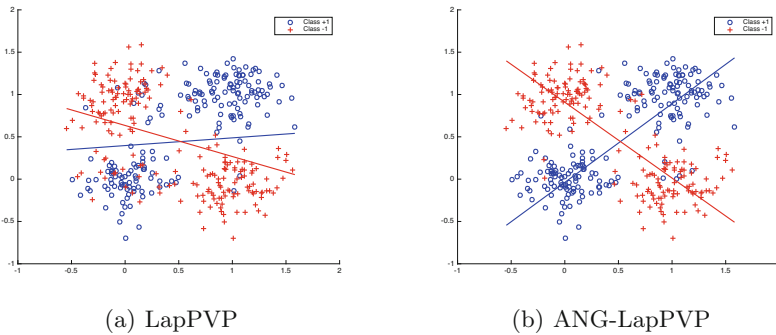
Fig. 1. Distribution of CrossPlane.

(a) LapPVP   (b) ANG-LapPVP

Fig. 2. Projection vectors obtained by LapPVP (a) and ANG-LapPVP (b).

3.2   Experiments on Benchmark Datasets

In the following experiments, we compare ANG-LapPVP with supervised algorithms, including GEPSVM, MVSVM, EMVSVM, TSVM and LSTSVM to evaluate the effectiveness of ANG-LapPVP, and compare it with semi-supervised algorithms (LapTSVM, LapLSTSVM and LapPVP) to verify the superiority of ANG-LapPVP on ten benchmark datasets. The benchmark datasets are collected from the UCI Machine Learning Repository [5]. We normalize the datasets so that all features range in the interval $[0, 1]$.

Each experiment is run 10 times with random 70% training data and the rest 30% test data. The average classification results are reported as the final ones. The grid search method is applied to finding the optimal hyper-parameters in each trial. Parameters β_1 and β_2 in both LapPVP and ANG-LapPVP are selected from $\{2^{-10}, 2^{-9}, \ldots, 2^0\}$, and other regularization parameters in all methods are selected from the set $\{2^{-5}, 2^{-4}, \ldots, 2^5\}$. In semi-supervised methods, the number of nearest neighbors is selected from the set $\{3, 5, 7, 9\}$.

Table 1. Mean accuracy and standard deviation (%) obtained by supervised algorithms with different scale of labeled data.

Dataset	Percent	GEPSVM	MVSVM	EMVSVM	TSVM	LSTSVM	ANG-LapPVP
Breast	10%	97.08 ± 1.15	95.92 ± 1.27	96.64 ± 0.91	96.02 ± 1.05	97.06 ± 0.66	**98.25 ± 1.18**
	30%	97.39 ± 0.98	96.73 ± 2.74	95.78 ± 2.08	**97.54 ± 1.16**	97.49 ± 0.87	**97.54 ± 0.80**
	50%	97.25 ± 0.66	96.58 ± 1.33	95.55 ± 1.82	97.87 ± 1.03	**97.91 ± 0.46**	97.44 ± 0.46
Check	10%	53.36 ± 1.82	54.68 ± 2.79	54.22 ± 2.43	61.13 ± 0.01	54.45 ± 3.30	**61.76 ± 3.15**
	30%	52.26 ± 1.50	51.46 ± 2.11	54.58 ± 3.76	59.80 ± 3.49	54.62 ± 3.58	**73.62 ± 3.82**
	50%	55.81 ± 2.07	50.53 ± 1.80	53.29 ± 2.70	55.61 ± 1.79	53.89 ± 1.79	**75.51 ± 2.34**
German	10%	71.13 ± 1.32	69.90 ± 1.21	60.00 ± 2.48	**73.80 ± 1.74**	73.60 ± 2.09	72.93 ± 1.72
	30%	74.63 ± 2.09	38.13 ± 8.83	61.87 ± 2.23	**76.10 ± 2.20**	75.97 ± 2.47	73.67 ± 1.04
	50%	75.00 ± 1.89	68.03 ± 5.16	61.13 ± 2.97	76.93 ± 1.00	**76.67 ± 1.72**	73.93 ± 0.90
Haberman	10%	66.88 ± 7.43	65.70 ± 7.12	58.82 ± 4.18	75.91 ± 3.17	76.02 ± 2.73	**77.67 ± 2.33**
	30%	71.72 ± 5.51	67.31 ± 6.04	55.70 ± 5.73	77.42 ± 2.27	78.17 ± 2.49	**78.71 ± 2.81**
	50%	75.81 ± 1.98	73.87 ± 3.51	55.91 ± 3.83	78.49 ± 2.03	77.96 ± 2.22	**79.14 ± 3.17**
Heart	10%	76.63 ± 7.69	73.04 ± 6.08	81.52 ± 3.90	67.61 ± 10.56	78.26 ± 4.38	**81.96 ± 3.77**
	30%	80.87 ± 2.36	60.33 ± 7.48	81.41 ± 3.18	80.89 ± 3.34	83.46 ± 3.40	**84.13 ± 2.77**
	50%	82.89 ± 2.08	82.39 ± 4.46	82.61 ± 1.85	82.65 ± 3.90	84.07 ± 3.60	**84.24 ± 2.47**
Hepatiti	10%	60.00 ± 8.79	67.23 ± 6.20	68.72 ± 3.62	60.43 ± 8.64	57.87 ± 8.26	**75.11 ± 2.00**
	30%	62.34 ± 7.31	64.47 ± 6.95	70.00 ± 5.16	62.77 ± 5.87	70.21 ± 4.91	**76.26 ± 3.81**
	50%	69.15 ± 1.81	67.23 ± 3.64	70.21 ± 4.60	68.09 ± 1.00	65.32 ± 3.02	**82.13 ± 2.69**
Liver	10%	58.56 ± 3.06	59.90 ± 5.07	57.60 ± 2.24	70.29 ± 1.15	68.65 ± 4.60	**71.35 ± 2.02**
	30%	57.50 ± 4.34	61.83 ± 2.87	55.10 ± 2.27	70.73 ± 2.87	69.52 ± 2.31	**71.06 ± 3.19**
	50%	61.83 ± 4.28	65.00 ± 2.80	52.02 ± 3.28	71.62 ± 1.82	70.29 ± 2.92	**72.40 ± 2.03**
Sonar	10%	62.03 ± 7.14	73.28 ± 7.81	61.41 ± 5.85	72.34 ± 3.04	67.81 ± 4.67	**74.84 ± 5.13**
	30%	73.13 ± 3.81	75.63 ± 2.11	62.34 ± 5.02	72.50 ± 3.70	63.28 ± 4.67	**79.22 ± 3.76**
	50%	80.47 ± 4.12	81.08 ± 5.37	59.22 ± 5.07	66.72 ± 6.08	72.50 ± 1.32	**81.41 ± 2.80**
Wdbc	10%	95.64 ± 2.16	94.77 ± 1.60	94.48 ± 1.00	84.48 ± 6.69	85.93 ± 4.49	**96.45 ± 0.89**
	30%	95.70 ± 1.26	79.07 ± 19.63	94.94 ± 0.99	96.51 ± 1.78	96.28 ± 1.51	**97.73 ± 0.70**
	50%	97.97 ± 0.88	96.86 ± 0.78	94.07 ± 2.14	97.38 ± 1.43	98.26 ± 0.27	**98.38 ± 0.92**
Wpbc	10%	76.23 ± 1.77	76.23 ± 4.84	66.07 ± 7.00	72.95 ± 4.59	68.20 ± 2.34	**78.69 ± 1.73**
	30%	75.08 ± 3.52	72.79 ± 3.39	63.93 ± 8.29	70.00 ± 8.07	71.31 ± 3.56	**80.49 ± 0.52**
	50%	77.70 ± 3.64	77.70 ± 2.07	63.93 ± 6.60	76.39 ± 1.15	79.67 ± 4.18	**80.67 ± 0.85**

Comparison with Supervised Algorithms. We first compare ANG-LapPVP with GEPSVM, MVSVM, EMVSVM, TSVM and LSTSVM to investigate the performance of adaptive neighbors graph. Specially, we discuss the impact of the different scale of labeled data on these algorithms. Additionally, ANG-LapPVP has 50% training samples as unlabeled ones.

Table 1 lists the results of supervised algorithms and ANG-LapPVP with 10%, 30% and 50% of training data as labeled samples, where the best results are highlighted. Experimental results in Table 1 show the effectiveness of ANG-LapPVP. With the increasing number of labeled data, the accuracy of ANG-LapPVP on most datasets goes up gradually, which indicates that the labeled data can provide more discriminant information. Moreover, we observe that ANG-LapPVP trained with unlabeled data has the best classification performance on all ten datasets except Breast with 50% labeled data and German with three situations, which fully demonstrates the significance of the adaptive

Table 2. Mean accuracy and standard deviation (%) obtained by semi-supervised algorithms on 30% unlabeled data.

Dataset	LapTSVM	LapLSTSVM	LapPVP	ANG-LapPVP
Breast	95.17 ± 1.58	97.30 ± 1.63	97.77 ± 0.95	**98.25 ± 1.18**
Check	50.90 ± 5.01	58.04 ± 3.36	60.00 ± 1.50	**75.78 ± 3.09**
German	70.17 ± 1.97	**73.87 ± 2.22**	70.93 ± 0.21	72.43 ± 1.16
Haberman	69.46 ± 3.59	74.52 ± 2.03	71.83 ± 7.07	**77.53 ± 2.51**
Heart	72.83 ± 6.87	81.41 ± 6.33	**84.02 ± 2.18**	81.96 ± 4.73
Hepatiti	57.87 ± 6.17	66.17 ± 6.22	73.40 ± 4.84	**75.11 ± 2.02**
Liver	59.23 ± 6.80	66.35 ± 3.42	60.19 ± 2.65	**70.67 ± 2.28**
Sonar	67.34 ± 4.51	74.22 ± 3.40	68.91 ± 5.28	**75.78 ± 4.12**
Wdbc	94.42 ± 1.82	96.51 ± 1.37	95.47 ± 0.77	**96.80 ± 0.63**
Wpbc	73.77 ± 5.01	79.02 ± 2.16	77.21 ± 0.52	**79.18 ± 1.74**

similarity matrices provided by labeled and unlabeled training data. Generally speaking, semi-supervised algorithms outperform the related supervised ones, and the proposed ANG-LapPVP gains the most promising classification performance.

Comparison with Semi-supervised Algorithms. To validate the superiority of ANG-LapPVP, we further analyze experimental results of LapTSVM, LapLSTSVM, LapPVP and ANG-LapPVP. Tables 2 and 3 list mean accuracy and standard deviation obtained by semi-supervised algorithms on 30% and 50% training samples as unlabeled ones, respectively, where the best results are in bold. Additionally, there are 20% training samples as labeled ones.

From the results in Tables 2 and 3, we can see that ANG-LapPVP has a higher accuracy than LapPVP on all ten datasets except Heart with 30% unlabeled data. The evidence further indicates that ANG-LapPVP with the ANG regularization well preserves the structure of training data and has a better classification performance than LapPVP. Moreover, compared with the other semi-supervised algorithms, ANG-LapPVP has the highest accuracy on eight datasets in Table 2 and on nine datasets in Table 3. That is to say, ANG-LapPVP has substantial advantages over LapTSVM and LapLSTSVM. On the whole, ANG-LapPVP has an excellent ability in binary classification tasks.

Table 3. Mean accuracy and standard deviation (%) obtained by semi-supervised algorithms on 50% unlabeled data.

Dataset	LapTSVM	LapLSTSVM	LapPVP	ANG-LapPVP
Breast	95.45 ± 2.00	97.77 ± 0.74	96.16 ± 1.54	**98.58 ± 0.55**
Check	50.86 ± 5.04	56.64 ± 2.64	59.63 ± 0.80	**61.76 ± 3.15**
German	71.17 ± 2.33	**74.10 ± 0.74**	72.67 ± 1.51	73.20 ± 1.00
Haberman	70.11 ± 4.35	66.45 ± 4.72	74.30 ± 2.12	**79.46 ± 2.89**
Heart	73.80 ± 7.02	77.93 ± 8.39	78.04 ± 5.22	**82.39 ± 2.28**
Hepatiti	59.79 ± 6.06	65.53 ± 5.92	67.66 ± 4.89	**71.06 ± 4.04**
Liver	59.62 ± 5.91	63.85 ± 5.48	62.21 ± 4.44	**71.35 ± 2.02**
Sonar	66.09 ± 3.76	72.03 ± 2.01	70.00 ± 1.77	**73.13 ± 5.25**
Wdbc	94.01 ± 1.40	94.88 ± 1.19	95.41 ± 0.75	**96.45 ± 0.89**
Wpbc	74.10 ± 3.26	90.68 ± 3.55	90.30 ± 1.28	**93.03 ± 3.27**

4   Conclusion

In this paper, we propose ANG-LapPVP for binary classification tasks. As the extension of LapPVP, ANG-LapPVP improves its classification performance by introducing the ANG regularization. The ANG regularization induces an adaptive neighbor graph where the similarity matrix is changed with iterations. Experimental results on the artificial and benchmark datasets validate that the effectiveness and superiority of the proposed algorithm. In a nutshell, ANG-LapPVP has a better classification performance than LapPVP and is a promising semi-supervised algorithm.

Although ANG-LapPVP achieves a good classification performance on datasets used here, the projection vectors obtained by ANG-LapPVP may be not enough when handling with a large scale dataset. In this case, we could consider projection matrices that may provide more discriminant information. Therefore, the dimensionality of projection matrices is a practical problem to be addressed in our following work. In addition, multi-class classification tasks in reality are also in consideration.

References

1. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. J. Mach. Learn. Res. **7**, 2399–2434 (2006)

2. Chapelle, O., Schölkopf, B., Zien, A.: Introduction to semi-supervised learning. In: Chapelle, O., Schölkopf, B., Zien, A. (eds.) Semi-Supervised Learning, pp. 1–12. The MIT Press, Cambridge (2006)
3. Chen, W., Shao, Y., Deng, N., Feng, Z.: Laplacian least squares twin support vector machine for semi-supervised classification. Neurocomputing **145**, 465–476 (2014)
4. Culp, M.V., Michailidis, G.: Graph-based semisupervised learning. IEEE Trans. Pattern Anal. Mach. Intell. **30**(1), 174–179 (2008)
5. Dua, D., Graff, C.: UCI machine learning repository (2017). https://archive.ics. uci.edu/ml
6. Fan, M., Gu, N., Qiao, H., Zhang, B.: Sparse regularization for semi-supervised classification. Pattern Recogn. **44**(8), 1777–1784 (2011)
7. Jayadeva, Khemchandani, R., Chandra, S.: Twin support vector machines for pattern classification. IEEE Trans. Pattern Anal. Mach. Intell. **29**(5), 905–910 (2007)
8. Kumar, M.A., Gopal, M.: Least squares twin support vector machines for pattern classification. Expert Syst. Appl. **36**(4), 7535–7543 (2009)
9. Mangasarian, O.L., Wild, E.W.: Multisurface proximal support vector machine classification via generalized eigenvalues. IEEE Trans. Pattern Anal. Mach. Intell. **28**(1), 69–74 (2006)
10. Nie, F., Dong, X., Li, X.: Unsupervised and semisupervised projection with graph optimization. IEEE Trans. Neural Netw. Learn. Syst. **32**(4), 1547–1559 (2021)
11. Nie, F., Wang, X., Huang, H.: Clustering and projected clustering with adaptive neighbors. In: Macskassy, S.A., Perlich, C., Leskovec, J., Wang, W., Ghani, R. (eds.) The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2014, 24–27 August 2014, pp. 977–986. ACM, New York (2014)
12. Qi, Z., Tian, Y., Shi, Y.: Laplacian twin support vector machine for semi-supervised classification. Neural Netw. **35**, 46–53 (2012)
13. Vapnik, V.: Statistical Learning Theory. Wiley, Hoboken (1998)
14. Xue, Y., Zhang, L.: Laplacian pair-weight vector projection for semi-supervised learning. Inf. Sci. **573**, 1–19 (2021)
15. Yang, Z., Xu, Y.: Laplacian twin parametric-margin support vector machine for semi-supervised classification. Neurocomputing **171**, 325–334 (2016)
16. Yang, Z.: Nonparallel hyperplanes proximal classifiers based on manifold regularization for labeled and unlabeled examples. Int. J. Pattern Recognit. Artif. Intell. **27**(5), 1350015 (2013)
17. Ye, Q., Ye, N., Yin, T.: Enhanced multi-weight vector projection support vector machine. Pattern Recogn. Lett. **42**, 91–100 (2014)
18. Ye, Q., Zhao, C., Ye, N., Chen, Y.: Multi-weight vector projection support vector machines. Pattern Recogn. Lett. **31**(13), 2006–2011 (2010)
19. Zhang, H., Wu, D., Nie, F., Wang, R., Li, X.: Multilevel projections with adaptive neighbor graph for unsupervised multi-view feature selection. Inf. Fusion **70**, 129–140 (2021)
20. Zhong, W., Chen, X., Nie, F., Huang, J.Z.: Adaptive discriminant analysis for semi-supervised feature selection. Inf. Sci. **566**, 178–194 (2021)