# NEW SUPERVISED LOCALLY LINEAR EMBEDDING FOR DIMENSIONALITY REDUCTION USING DISTANCE METRIC LEARNING

*B. Yang[*], M. Xiang[*], Y.P. Zhang[*], S.M. Pang[†], X. Li[*]*

**Abstract:** Feature reduction is an important issue in pattern recognition. Lower feature dimensionality could reduce the complexity and enhance the generalization ability of classifiers. In this paper we propose a new supervised dimensionality reduction method based on Locally Linear Embedding and Distance Metric Learning. First, in order to increase the interclass separability, a linear discriminant transformation learnt from distance metric learning is used to map the original data points to a new space. Then Locally Linear Embedding is adopted to reduce the dimensionality of data points. This process extends the traditional unsupervised Locally Linear Embedding to supervised scenario in a clear and natural way. In addition, it can also be seen as a general framework for developing new supervised dimensionality reduction algorithms by utilizing corresponding unsupervised methods. Extensive classification experiments performed on some real-world and artificial datasets show that the proposed method can achieve comparable to or even better results over other state-of-the-art dimensionality reduction methods.

Key words: *dimensionality reduction, manifold learning, locally linear embedding, distance metric learning*

*Received: October 16, 2016* **DOI:** 10.14311/NNW.2016.26.026
*Revised and accepted: September 19, 2016*

## 1. Introduction

Dimensionality reduction plays an important role in fields of machine learning and pattern recognition. High feature dimensionality not only increases the complexity of classifier, but also increases the risk of overfitting in designing classifier. Many methods have been proposed to reduce the feature dimensionality, which can be roughly categorized to feature selection and feature extraction. Feature selection abandons redundant features and only keeps the most useful ones, whereas feature

[*]Bo Yang, Ming Xiang – Corresponding author, Yupei Zhang, Xin Li, Department of Computer Science and Technology, Xi'an Jiaotong University, No. 28, Xianning West Road, Xi'an, Shaanxi, 710049, China, E-mail: yangboo@stu.xjtu.edu.cn, mxiang@mail.xjtu.edu.cn, xjtuiezhyp@stu.xjtu.edu.cn, z295638718@stu.xjtu.edu.cn

[†]Shanmin Pang, School of Software Engineering, Xi'an Jiaotong University, Xi'an, China, E-mail: pangsm@xjtu.edu.cn

extraction creates new features from the original. Principal Component Analysis (PCA) [14] and Linear Discriminant Analysis (LDA) [16], due to their relative effectiveness and simplicity, have been the two classical feature extraction methods. Recently some manifold learning methods have attracted a lot of attention [24,25,28]. They assume that the real-world data observed in high dimensional space is usually generated by a process with relatively few degrees of freedom [30], and the data intended for embedding into a lower dimensional space should preserve the intrinsic geometrical structures [13]. In the past few years, many manifold learning-based algorithms have been proposed, e.g., Multi-Dimensional Scaling (MDS) [1], Locally Linear Embedding (LLE) [19], Isometric Maps (ISOMAP) [20], and Locality Preserving Projection (LPP) [7,8]. These methods have yielded very impressive results. Among them, LLE is one of the most widely used methods, since it requires only one free parameter to be determined and does not have the local extrema problem.

LLE is an unsupervised method, since it does not consider any class label information of data. In order to use it in supervised scenario, several supervised versions of LLE (SLLE) have been proposed, including $\alpha$-SLLE [9, 15, 18], ESLLE [27], PLLE [29], LDE [4], and LLDE [11]. Most of these methods try to modify the inter- and intraclass distances for achieving better classification results, basically following the idea of $\alpha$-SLLE. In $\alpha$-SLLE, the interclass distances are increased by adding a constant value; in contrast, the intraclass distances are kept unchanged. This idea is easily implemented in the training phase, since we have known all training data and its class labels. However when facing a testing data point, since its class label is unknown, the distances between this data point and its neighbors are unable to determine whether to increase or to keep. It is definitely a drawback of $\alpha$-SLLE, although it has been used successfully in many real world applications, such as handwritten digits classification [10], facial expression recognition [12], prediction of membrane protein types [21], gene expression data analysis [17], etc.. ESLLE [27] essentially has the same drawback as $\alpha$-SLLE. PLLE [29] attempts to overcome it via estimating roughly the class labels of testing points. LDE [4] and LLDE [11] do not have this drawback, since they use a linear transformation to achieve dimensionality reduction, therefore their effectiveness is also limited.

In this paper we propose a novel SLLE called Supervised Locally Linear Embedding Based on Distance Metric Learning (SLLE-DML). First, the original data points are mapped into a new space by a transformation learnt from DML, and then LLE is applied on the transformed data points to achieve dimensionality reduction. The class label information is utilized in the first phase, thus the above mentioned drawback of $\alpha$-SLLE is avoided spontaneously. In fact, the final mapping is nonlinear, since it is a composition of two transformations learnt via DML and LLE, respectively. Hence, the proposed method also overcomes the linear mapping limitation of LDE and LLDE.

The remainder of this paper is organized as follows. In Section 2, related works are briefly introduced. In Section 3, we present the Supervised Locally Linear Embedding based on Distance Metric Learning. In Section 4, the experimental results are presented. Finally, some concluding remarks are included in Section 5.

# 2. Reviews of LLE and SLLE

## 2.1 Locally Linear Embedding

LLE maps the original high dimensional data points into a low dimensional space with the local geometric structure preserved, via solving a certain constrained optimization problem. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N] \in R^{D\times N}$ be a dataset of $N$ points with dimension $D$, which lies on an underlying manifold, and whose intrinsic dimensionality is $d$, $d \ll D$. The corresponding low dimensional data is denoted by $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N] \in R^{d\times N}$. The procedure of LLE is given as follows. First, for each data point $\mathbf{x}_i$, find $k$ nearest neighbors ($k$-NN) according to Euclidean distance in $R^D$, and use its neighbors to reconstruct $\mathbf{x}_i$ via linear combination. The reconstruction weight $w_{ij}$ can be solved by minimizing the following quantity:

$$\left\| \mathbf{x}_i - \sum_{j=1}^{N} w_{ij}\mathbf{x}_j \right\|^2, \tag{1}$$

with satisfying the following constraints:

$$\begin{cases} \sum\limits_{j=1}^{N} w_{ij} = 1 \\ w_{ij} = 0 \quad \text{for } \mathbf{x}_j \notin \text{neig}(\mathbf{x}_i), \end{cases} \tag{2}$$

where $\text{neig}(\mathbf{x}_i)$ denotes the neighbors of point $\mathbf{x}_i$. Clearly, this is a constrained least squares problem, which can be easily solved by traditional optimization method. After computed the weights for each data point $\mathbf{x}_i$, we can obtain a $N \times N$ reconstruction weights matrix $\mathbf{W}$. Then the low-dimensional representation $\mathbf{Y}$ can be got by minimizing the following quantity:

$$\sum_{i=1}^{N} \left\| \mathbf{y}_i - \sum_{j=1}^{k} w_{ij}\mathbf{y}_j \right\|^2, \tag{3}$$

with two constraints:

$$\begin{cases} \sum\limits_{i=1}^{N} \mathbf{y}_i = 0 \\ \frac{1}{N} \sum\limits_{i=1}^{N} \mathbf{y}_i {\mathbf{y}_i}^{\mathrm{T}} = \mathbf{I}, \end{cases} \tag{4}$$

where $\mathbf{I}$ is a $d \times d$ identity matrix. In order to solve Eq.(4), we should rewrite the objective function as follows:

$$\min_{\mathbf{Y}} \text{tr}(\mathbf{Y}\mathbf{A}\mathbf{Y}^{\mathrm{T}}), \tag{5}$$

where $\mathbf{A} = (\mathbf{I} - \mathbf{W})^{\mathrm{T}}(\mathbf{I} - \mathbf{W})$. Then the solution $\mathbf{Y}$ can be achieved as the eigenvectors corresponding to the $d$ smallest eigenvalues of $\mathbf{A}$.

### 2.2 Supervised Locally Linear Embedding

LLE is an unsupervised method, since it without using any class label information. If the class label information were used, the recognition accuracy should be improved. The basic idea of mainstream SLLE methods, e.g. $\alpha$-SLLE, is to increase the distance between data pair from different classes, while to keep the distance unchanged for the pair from the same class. Let

$$\Delta' = \max \mathrm{dis}(\mathbf{x}_i, \mathbf{x}_j), i, j = 1, ..., N, \tag{6}$$

where dis(, ) is Euclidean distance in $R^D$. Then the new distance $\Delta_{ij}$ between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as

$$\Delta_{ij} = \mathrm{dis}(\mathbf{x}_i, \mathbf{x}_j) + \alpha\, \Delta'\, \Lambda_{ij}, \quad \alpha \in [0, 1], \tag{7}$$

where $\Lambda_{ij} = 1$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to different classes, and $\Lambda_{ij} = 0$ otherwise. Note that the new distance takes class information into account, and parameter $\alpha$ controls the amount of the class information used. For $\alpha = 0$ , SLLE degenerates to the original unsupervised LLE, for $\alpha = 1$, it is the fully supervised. Setting $\alpha$ to a value between 0 and 1 gives a partially supervised LLE (called $\alpha$-SLLE). In practice, the optimal value of $\alpha$ should be determined via validation sets. The new distance $\Delta_{ij}$ can be used in LLE instead of the Euclidean distance, so as to achieve supervised purpose.

## 3. Supervised Locally Linear Embedding based on Distance Metric Learning

LLE gives no direct mapping from the input space to the low dimensional embedding space. According to Eq.(3), the lower dimensional data points are obtained via solving the optimization problem. Thus, when we want to get the output $\mathbf{y}_0$ corresponding to a new input $\mathbf{x}_0$, we should in principle rerun the entire LLE algorithm on the original dataset augmented by $\mathbf{x}_0$. This property of LLE raises certain difficulties for the current SLLE methods, for a new data point $\mathbf{x}_0$ needed to be classified. Since the first step is to calculate the $k$ nearest neighbors of $\mathbf{x}_0$. However $\mathbf{x}_0$ has no label information, hence the distance $\Delta_{ij}$ in Eq.(7) is not available. This is just the main drawback of the methods in [9,15,18]. To overcome this shortcoming, we propose a new SLLE method based on distance metric learning, called SLLE-DML.

Before going into the details of our proposed method, let's introduce Distance Metric Learning (DML). The basic idea of DML is to learn a new distance metric from the given data [2,3,5,6,22,23,26]. Experimental results in the literatures have shown that using DML can significantly improve the performance of classification and clustering. The new distance metric can be learnt as follows. For any pair of data points $(\mathbf{x}_i, \mathbf{x}_j)$ of the same class, it is assigned to a similar set $SI$ , and for any pair of different classes, it is assigned to a dissimilar set $DS$. Then the new distance is given by

$$\mathrm{dis}_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{M}} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)}, \tag{8}$$

where the matrix $\mathbf{M}$ is learnt from the dataset $\mathbf{X}$, by solving the following constrained optimization problem [23]:

$$\begin{array}{ll} \max\limits_{\mathbf{M}} & \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in DS} \|\mathbf{x}_i-\mathbf{x}_j\|_{\mathbf{M}} \\ s.t. & \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in SI} \|\mathbf{x}_i-\mathbf{x}_j\|^2_{\mathbf{M}} \le 1 \\ & \mathbf{M}\succeq 0. \end{array} \tag{9}$$

Note that the optimization problem Eq.(9) is convex, so we can solve $\mathbf{M}$ by semidefinite programming (SDP) or gradient ascent method. The computational complexity for solving $\mathbf{M}$ is usually very high, so a simplified method is proposed in [22]. $\mathbf{M}$ can be written as $\mathbf{M}=\mathbf{L}^{\mathrm{T}}\mathbf{L}$ , so we can rewrite Eq.(8) as

$$\begin{aligned} \mathrm{dis}_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) &= \sqrt{(\mathbf{x}_i-\mathbf{x}_j)^{\mathrm{T}}\mathbf{M}(\mathbf{x}_i-\mathbf{x}_j)} \\ &= \sqrt{(\mathbf{x}_i-\mathbf{x}_j)^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}(\mathbf{x}_i-\mathbf{x}_j)} \\ &= \sqrt{(\mathbf{L}\mathbf{x}_i-\mathbf{L}\mathbf{x}_j)^{\mathrm{T}}(\mathbf{L}\mathbf{x}_i-\mathbf{L}\mathbf{x}_j)}. \end{aligned} \tag{10}$$

Let $\mathbf{z}_i$=$\mathbf{L}(\mathbf{x}_i)$ and $\mathbf{Z}=[\mathbf{z}_1,\mathbf{z}_2,...,\mathbf{z}_N]\in R^{D\times N}$, then we have:

$$\mathrm{dis}_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)=\mathrm{dis}(\mathbf{z}_i,\mathbf{z}_j). \tag{11}$$

The basic idea of SLLE-DML is via DML to realize the supervision and using LLE to achieve dimensionality reduction. Hence we first obtain the matrix $\mathbf{M}$ by solving the optimization problem Eq.(9), and consequently obtain the matrix $\mathbf{L}$. Then use matrix $\mathbf{L}$ to transform the original dataset $\mathbf{X}$ to $\mathbf{Z}$. Finally we run the LLE algorithm on $\mathbf{Z}$ using Euclidean distance to obtain the corresponding points $\mathbf{Y}$, and design a classifier using $\mathbf{Y}$ as the training set. According to Eq.(11), using Euclidean distance on $\mathbf{Z}$ is equivalent to using $\mathrm{dis}_{\mathbf{M}}$ on $\mathbf{X}$. Let $\mathbf{T}$ denote the transformation from $\mathbf{Z}$ to $\mathbf{Y}$, hence the feature reduction transformation $\mathbf{F}$ from $\mathbf{X}$ to $\mathbf{Z}$ is composed of $\mathbf{L}$ and $\mathbf{T}$. Since $\mathbf{T}$ is nonlinear, $\mathbf{F}$ also is nonlinear, which overcomes the shortcomings of LDE and LLDE as mentioned in Section 1. The class label information is incorporated into the transformation $\mathbf{L}$ in DML stage, so the drawback regarding testing new data points is also overcome. Specifically, for a new data point $\mathbf{x}_0$ to be classified, we first transform $\mathbf{x}_0$ to $\mathbf{z}_0=\mathbf{L}\mathbf{x}_0$, and run the LLE algorithm on the augmented dataset $\mathbf{Z}'=[\mathbf{z}_0,\mathbf{z}_1,...,\mathbf{z}_N]$ to get $\mathbf{Y}'=[\mathbf{y}_0,\mathbf{y}_1,...,\mathbf{y}_N]$, and then we can determine the class of $\mathbf{x}_0$ by classifying $\mathbf{y}_0$ using the designed classifier.

# 4. Experiments and results

Several experiments are presented in this section to investigate the performance of the proposed method for classification tasks. We compare SLLE-DML with some other feature reduction methods, including LLE, $\alpha$-SLLE, PCA, LDA and LLDE on 4 UCI datasets and the Yale face dataset. These UCI datasets are Wine, Vehicle, Multiple Feature, and Optical Digits, as described in Tab. I. In all classification experiments, the $k$-NN classifier with Euclidean distance is used after feature reduction to make the final classification decision.

| Abbr. | Name | Size | Dimension | Classes |
|---|---|---|---|---|
| Wine | Wine | 178 | 13 | 3 |
| Vehicle | Vehicle | 946 | 18 | 4 |
| Mul | Multiple features | 2000 | 647 | 10 |
| Opt | Optical digits | 5620 | 64 | 10 |

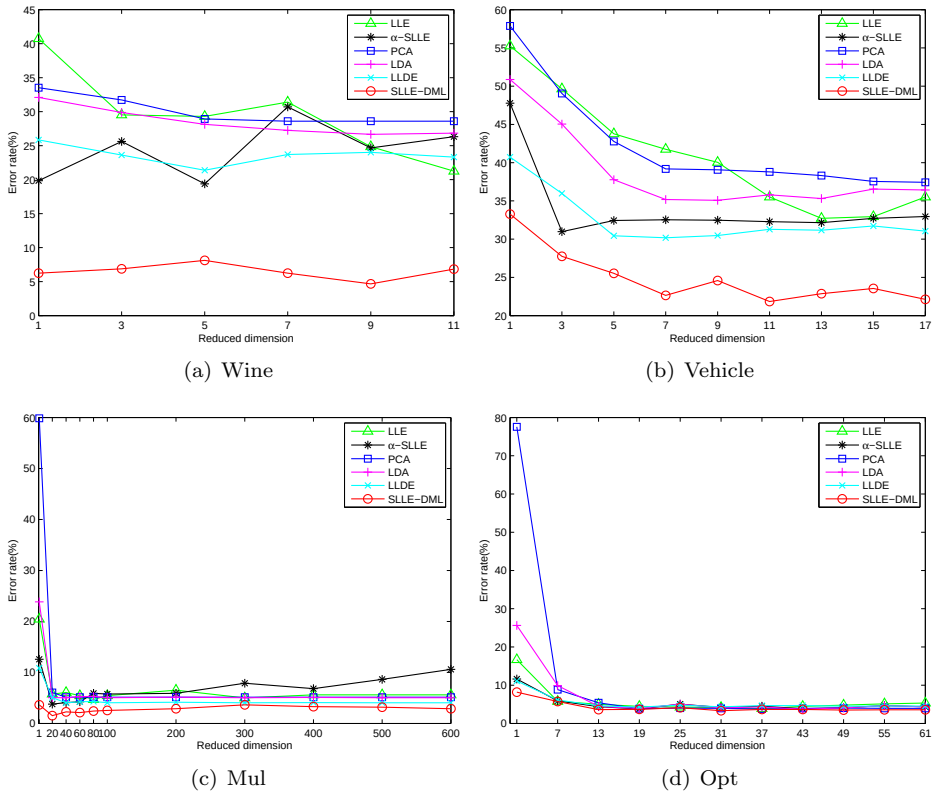
**Tab. I** *The UCI datasets description.*

The parameters for most of the methods are determined empirically through ten-fold cross validation. That is, for each parameter, several values are tested through ten-fold cross validation and the best one is selected. For $\alpha$-SLLE, the values of $\alpha$ is tested from 0.1 to 1 at intervals of 0.1. When applying $k$-NN classifier, several values of $k$ from 5 to 50 are tested. On each data set, ten times ten-fold cross validation is run. That is, in each time, the original data set is randomly divided into ten equal-sized subsets while keeping the proportion of the instances in different classes. Then, in each fold, one subset is used as testing set and the union of the remaining ones is used as training set. After ten folds, each subset has been used as testing set once.

The averaged misclassification rates of UCI datasets are reported in Figs. 1(a)-1(d). In them, the X-axis represents the number of dimensions, and the Y-axis represents the corresponding misclassification rates. From Fig. 1, we can see that our proposed method is generally superior to other feature reduction methods, and the advantage is more obvious when the datasets are reduced to very low dimensionality.

To further test the performance of SLLE-DML for higher dimensional data, Yale face dataset is adopted. It contains 165 images of 15 individuals and each individual has 11 images. A sample from the dataset is shown in Fig. 2. Before dimensionality reduction, all the original face images are normalized, so that the two eyes are aligned at the same position and cropped to the size of $100 \times 100$. In the following experiments, 6 randomly chosen image samples per person are used for training, and the remaining 5 images for testing. The experiment is repeated 10 times, and the misclassification rates are averaged. Fig. 3 describes the relationships between the average error rate and the reduced dimensionality for different algorithms. Form Fig. 3, it is clear that our method is much better than LLE, $\alpha$-SLLE, PCA, LDA and LLDE. Tab. II records the mean error rates and standard deviations of the best performance of different methods on different datasets. Comparing these results, we can clearly see that SLLE-DML is superior to other state-of-the-art methods.

In order to investigate the possible reasons for the good performance of our proposed method, we carry out an experiment using an artificial dataset with class labels. The dataset is sampled from the Swiss roll in $R^3$ as shown in Fig. 4(a), and the data set is shown in Fig. 4(b). The Swiss roll is a manifold in $R^3$ and its intrinsic dimensionality is two, since it is homeomorphic to a two dimensionality rectangle shown in Fig. 4(c). The different colors in these figures represent different classes, and the dataset contains 100 points for each class. To compare the performance of different methods for classification task, we reduce the dimensionality of the dataset




(a) Wine (b) Vehicle

(c) Mul (d) Opt

**Fig. 1** *Comparison of different dimensionality reduction methods for UCI datasets.*

**Fig. 2** *Sample images of one person in Yale dataset.*

to two, using LLE, $\alpha$-SLLE and SLLE-DML, respectively. The datasets with the reduced dimensionality using different methods are shown in Figs. 4(d)- 4(f). As shown in Fig. 4(f), the dataset obtained using SLLE-DML can be easily partitioned into four regions corresponding to four different classes. However, in Fig. 4(d) and 4(e), the results obtained by LLE and $\alpha$-SLLE, are severally overlapped. This clearly suggests that the result obtained with SLLE-DML is much easier to be classified than that obtained using LLE and $\alpha$-SLLE.

## 5. Conclusions

A new dimensionality reduction method is proposed in this paper. It combines DML and LLE to achieve efficient feature reduction for classification. In our

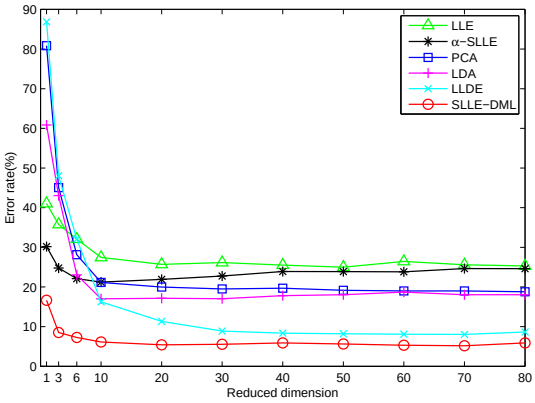
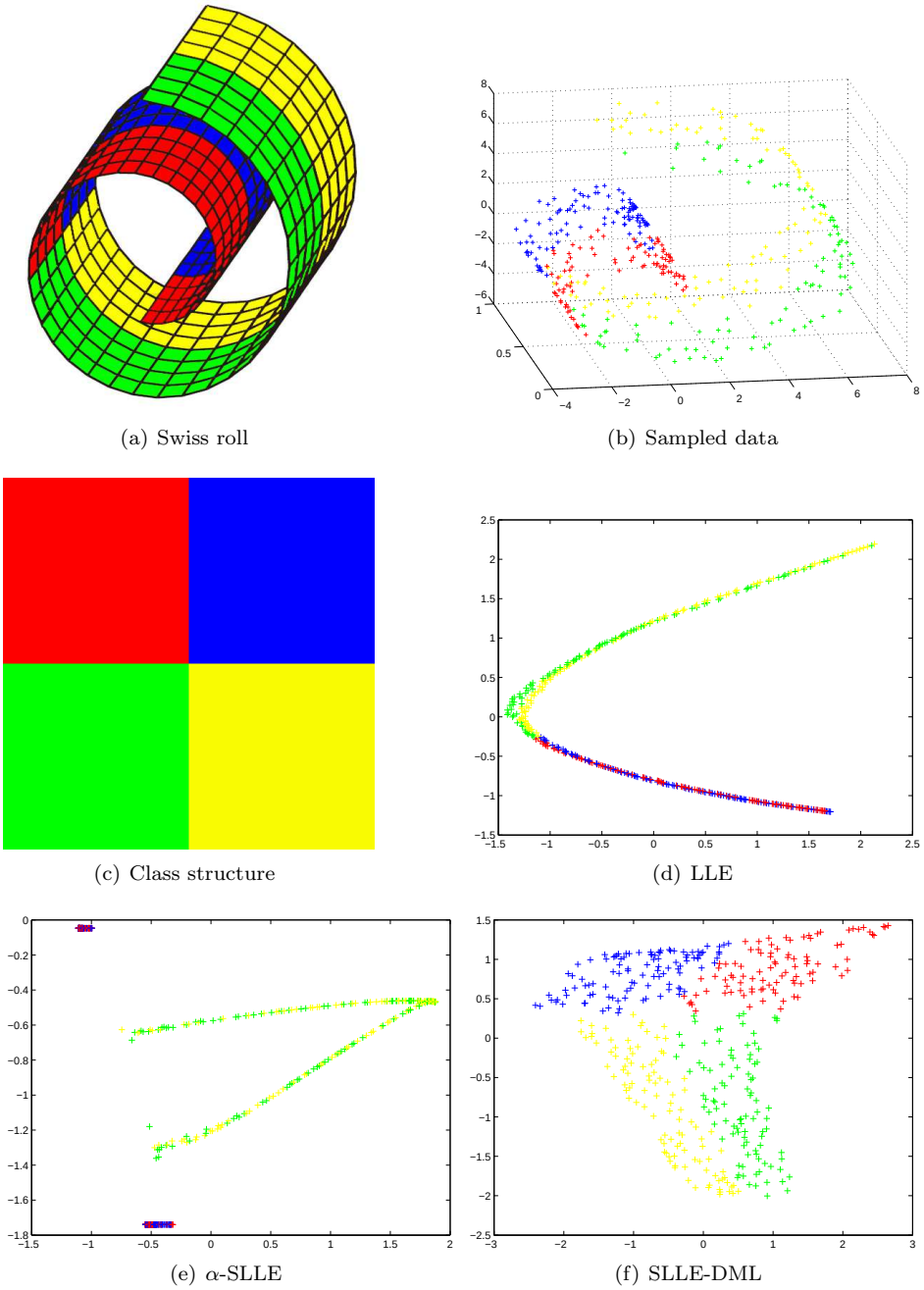



**Fig. 3** *Comparison of different dimensionality reduction methods for the Yale face dataset.*

| Dataset | LLE | $\alpha$-SLLE | PCA | LDA | LLDE | SLLE-DML |
|---|---|---|---|---|---|---|
| Wine | 0.2125 | 0.1938 | 0.2860 | 0.2684 | 0.2138 | **0.0467** |
| | ±0.0884 | ±0.0902 | ±0.0836 | ±0.0780 | ±0.0709 | ±0.0398 |
| Vehicle | 0.3272 | 0.3098 | 0.3743 | 0.3507 | 0.3018 | **0.2184** |
| | ±0.0682 | ±0.0728 | ±0.0609 | ±0.0454 | ±0.0559 | ±0.0370 |
| Mul | 0.0470 | 0.0379 | 0.0515 | 0.0503 | 0.0405 | **0.0152** |
| | ±0.0052 | ±0.0098 | ±0.0035 | ±0.0071 | ±0.0043 | ±0.0038 |
| Opt | 0.0390 | 0.0378 | 0.0378 | 0.0372 | 0.0428 | **0.0334** |
| | ±0.0193 | ±0.0238 | ±0.0207 | ±0.0186 | ±0.0210 | ±0.0159 |
| Yale | 0.2501 | 0.2124 | 0.1882 | 0.1698 | 0.0805 | **0.0518** |
| | ±0.0592 | ±0.0526 | ±0.0424 | ±0.0447 | ±0.0316 | ±0.0294 |

**Tab. II** *Best performance results of different methods on all datasets ( in A±B, A denotes the mean error rate, and B denotes the standard deviation).*

method, a distance metric learning procedure is first applied to increase the interclass separability, and then LLE is implemented to reduce the feature dimensionality. The final mapping is nonlinear, which is in fact the composition of the two transformations learnt via DML and LLE. This method not only overcomes the shortcomings in the current SLLE learning methods, but also is more effective. The experimental results show that the performance of the proposed SLLE-DML is better than many state-of-the-art dimensionality reduction techniques. In addition, experiments on an artificial dataset also clearly illustrate the effectiveness of the proposed method. Indeed, the proposed method can also be seen as a framework, which can be used in various tasks by adopting different algorithm combinations. One avenue for future work is to combine different DML and dimensionality reduction methods to further improve the performance of classification.



(a) Swiss roll (b) Sampled data

(c) Class structure (d) LLE

(e) $\alpha$-SLLE (f) SLLE-DML

**Fig. 4** *Comparison of different dimensionality reduction methods for the artificial dataset.*

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 61603289).

# References

[1] BORG I, GROENEN P.J.F. *Modern Multidimensional Scaling: Theory and applications*. USA: Springer, 2005, doi: 10.1007/978-1-4757-2711-1.

[2] CAO, Q., YING, Y., LI, P. Distance metric learning revisited. In: *Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2012*, Bristol, UK. Heidelberg, Berlin, GER: Springer, 2012, pp. 283–298.

[3] CAO, Q., YING, Y., LI, P. Similarity metric learning for face recognition. In: *Proceedings of International Conference on Computer Vision 2013*, Sydney, Australia. USA: IEEE, 2013, pp. 2408–2415.

[4] CHEN H.T., CHANG H.W., LIU T.L. Local discriminant embedding and its variants. In: *Proceedings of IEEE 2005 Computer Vision and Pattern Recognition*, San Diego, CA. USA: IEEE, 2005, pp. 846–853.

[5] DAVIS J.V., KULIS B., JAIN P., SRA S., DHILLON I.S. Information-theoretic metric learning. In: *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, USA. New York, NY, USA: ACM, 2007, pp. 209–216.

[6] GUILLAUMIN M., VERBEEK J., SCHMID C. Is that you? Metric learning approaches for face identification. In: *Proceedings of International Conference on Computer Vision 2009*, Kyoto, Japan. USA: IEEE, 2009, pp. 498–505.

[7] HE X.F., NIYOGI X. Locality preserving projections. In: *Proceedings of Neural Information Processing Systems 2004*, Vancouver, Canada. Cambridge, MA, USA: MIT, 2004, pp. 153–161.

[8] HE X.F., YAN S., HU Y., Niyogi P., ZHANG H. Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005, 27(3), pp. 328–340, doi: 10.1109/TPAMI.2005.55.

[9] KAYO. O. *Locally linear embedding algorithm extensions and applications* [online]. Faculty of Technology, University of Oulu, April 2006 [viewed 2016-05-26]. Research report. Available from: http://jultika.oulu.fi/files/isbn9514280415.pdf.

[10] KOUROPTEVA O., OKUN O., PIETIKAINEN M. Supervised locally linear embedding algorithm for pattern recognition. *Pattern Recognition and Image Analysis*. 2003, pp. 386–394, doi: 10.1007/978-3-540-44871-6_45.

[11] LI B., ZHENG C.H., HUANG D.S. Locally linear discriminant embedding: An efficient method for face recognition. *Pattern Recognition*. 2008, 41(12), pp. 3813–3821, doi: 10.1016/j.patcog.2008.05.027.

[12] LIANG D., YANG J., ZHENG Z., CHANG Y. A facial expression recognition system based on supervised locally linear embedding. *Pattern Recognition Letters*. 2005, 26(15), pp. 2374–2389, doi: 10.1016/j.patrec.2005.04.011.

[13] LIN T., ZHA H.B. Riemannian manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008, 30(5), pp. 796–809, doi: 10.1109/TPAMI.2007.70735.

[14] LIU H., MOTODA H. *Feature Selection for Knowledge Discovery and Data Mining*. USA: Springer, 2012, doi: 10.1007/978-1-4615-5689-3.

[15] MAATEN L.J.P., POSTMA E.O., HERIK H.J. *Dimensionality reduction: A comparative review* [online]. Tilburg Centre for Creative Computing, Tilburg University, Oct. 2009 [viewed 2016-05-26]. Research report. Available from: https://www.tilburguniversity.edu/upload/59afb3b8-21a5-4c78-8eb3-6510597382db_TR2009005.pdf

[16] NIXON S.M., AGUADO A.S. *Feature Extraction and Image Processing for Computer Vision*. USA: Academic Press, 2012, doi: 10.1016/B978-0-12-396549-3.00020-3.

[17] PILATI M., VIROLI C. Supervised locally linear embedding for classification: an application to gene expression data analysis. In: *Proceedings of 29th Annual Conference of the German Classification Society*, Magdeburg, German. Heidelberg, Berlin, GER: Springer, 2005, pp. 15–18.

[18] RIDDER D., KOUROPTEVA O., OKUN O., DICK D., PIETIKAINEN M., ROBERT P. Supervised locally linear embedding. In: *Proceedings of the 11st Artificial Neural Networks and Neural Information Processing*, Istanbul, Turkey. Heidelberg, Berlin, GER: Springer, 2003, pp. 333–341.

[19] ROWEIS S.T., SAUL L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000, 290(5500), pp. 2323–2326, doi: 10.1126/science.290.5500.2323.

[20] TENENBAUM J.B., DE SILVA V., LANGFORD J.C. A global geometric framework for nonlinear dimensionality reduction. *Science*. 2000, 290(5500), pp. 2319–2323, doi: 10.1126/science.290.5500.2319.

[21] WANG M., YANG J., XU Z.J., CHOU K. SLLE for predicting membrane protein types. *Journal of Theoretical Biology*. 2005, 232(1). pp. 7–15, doi: 10.1016/j.jtbi.2004.07.023.

[22] WEINBERGER K.Q., BLITZER J., SAUL L.K. Distance metric learning for large margin nearest neighbor classification. In: *Proceedings of Neural Information Processing Systems 2005*, Whistler, BC, CA. Cambridge, MA, USA: MIT, 2005, pp. 1473–1480.

[23] XING E.P., NG A.Y., JORDAN M.I., RUSSELL S. Distance metric learning, with application to clustering with side-information. In: *Proceedings of Neural Information Processing Systems 2003*, Whistler, BC, CA. Cambridge, MA, USA: MIT, 2003, pp. 505–512.

[24] YANG B., XIANG M., ZHANG Y.P. Learning discriminant Isomap for dimensionality reduction. In: *Proceedings of International Joint Conference on Neural Networks 2015*, Killarney, Ireland. USA: IEEE, 2015, pp. 1–8.

[25] YANG B., XIANG M., ZHANG Y.P. Multi-manifold discriminant Isomap for visualization and classification. *Pattern Recognition*. 2016, 55, pp. 215–230, doi: 10.1016/j.patcog.2016.02.001.

[26] YING Y., LI P. Distance metric learning with eigenvalue optimization. *Journal of Machine Learning Research*. 2012, 13(1), pp. 1–26. Available from: http://dl.acm.org/citation.cfm?id=2188386.

[27] ZHANG S.Q. Enhanced supervised locally linear embedding. *Pattern Recognition Letters*. 2009, 30, pp. 1208–1218, doi: 10.1016/j.patrec.2009.05.011.

[28] ZHANG Y.P., XIANG M., YANG B. Linear dimensionality reduction based on hybrid structure preserving projections. *Neurocomputing*. 2016, 173(3), pp. 518–529, doi: 10.1016/j.neucom.2015.07.011.

[29] ZHANG Z., ZHAO L. Probability-based locally linear embedding for classification. In: *Proceedings of the International Conference on Fuzzy Systems and Knowledge Discovery*, Hainan, China. New York, NY, USA: IEEE, 2007, pp. 243–247.

[30] ZHENG N.N., XUE J.R. *Statistical Learning and Pattern Analysis for Image and Video Processing*. USA: Springer, 2009, doi: 10.1007/978-1-84882-312-9.