ASM Sc. J., 12, Special Issue 1, 2019 for IQRAC2018, 182-190

# Extended Locality Preserving Partial Least Squares with Class Information

Muhammad Aminu*[1], Noor Atinah Ahmad[2], and Norhashidah Awang[3]

[1, 2, 3]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800, Penang, Malaysia*

**Corresponding author: muhammadaminu47@gmail.com*

Feature extraction techniques are methods widely used to reduce the dimensionality of a data while retaining most of the relevant information in the original data. Locality preserving partial least squares (LPPLS) is a recently developed feature extraction technique that aims to preserve the local structural information of data. LPPLS seeks to preserve local structure defined by nearest neighbors. However, the nearest neighbors may belong to different classes which might lead to the poor performance of LPPLS in discriminating the different classes in the data. In this paper, we propose an extension of LPPLS called extended locality preserving partial least squares which consider class label information. The binary (0-1) weighting technique together with label information is used to construct the similarity matrices that determine local projection of the data. Therefore, our extended LPPLS does not simply preserve local structure, but also has discriminating power to differentiate data from different classes. Experimental results on various data sets demonstrate the effectiveness of the proposed extended LPPLS. Two different evaluation metrics, normalized mutual information (NMI) and Fowlkes-Mallow index are used to measure the accuracy of methods used in the experiments.
**Keywords:** Class labels, feature extraction, local information, similarity matrix.

## I. Introduction

Most of the data obtained nowadays are high dimensional in nature which often leads to many challenging issues (Leskovec et al., 2014). Therefore, dimensionality reduction techniques are employed to determine a lower dimensional representation of the data such that most of the information in the data is preserved in the lower dimensional space. Partial least squares (PLS) (Rosipal and Krämer, 2006) and Principal component analysis (PCA) (Jolliffe, 2002) are the most commonly used dimension reduction techniques. PCA finds a projection of the data to lower dimensional subspace such that the variance of the data is maximized in the reduced space. While PLS finds a decomposition of two data matrices $X$ and $Y$ by maximizing the covariance between the two data matrices. These methods have been applied to many research fields such as machine learning, computer vision, pattern recognition etc. However, both PCA and PLS see only the global Euclidean structure of datasets which leads to the poor performance of these methods when the data lies on a nonlinear manifold.

In recent years, a lot of methods such as Locally linear embedding (LLE) (Roweis and Saul, 2000), Isomap (Tenenbaum et al., 2000) and Laplacian Eigenmap (Belkin and Niyogi, 2002, 2003) have been proposed to discover the manifold structure in which high dimensional data lies in low dimensional space. Such methods project the high dimensional data to a lower dimensional subspace such that the intrinsic structure of the data is well preserved. These methods has a lot of advantages which include few parameters, computational efficiency and ability to discover the intrinsic geometric structure of data manifolds. The Locally linear embedding (LLE) algorithm build local linear models of high dimensional datasets and then try to find a lower dimensional representation of the data such that the local distance relationships are preserved in the lower dimensional space. However, LLE is nonlinear and suffers from the out-of-sample problem (Shi et al., 2014). In order to extend LLE to linear case and also to solve the out-of-sample problem, a new dimension reduction method called Neighborhood preserving projections (NPP) (Pang et al., 2005) was proposed. NPP is derived from LLE and share the neighborhood preserving property of LLE. In NPP, a linear transformation matrix is obtain by solving an objective function similar to that of LLE. The transformation matrix is then use to project the high dimensional data to a lower dimensional subspace.

Locality preserving projections (LPP) (He and Niyogi, 2004) is a recently proposed dimensionality reduction method which is derived by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. LPP also has the ability to resolve the out-of-sample problem and can preserve the local neighborhood information of data points to a higher extent. These properties make LPP different and more efficient than other dimension reduction techniques such as PCA and PLS.

The PLS method in its traditional form does not consider the local information of data. To tackle this problem, a new dimension reduction method called Locality preserving partial least squares (LPPLS) (Zhang et al., 2016) was proposed. LPPLS is obtain by introducing local information into the objective function of PLS. Similarly, a modification of Fisher discriminant analysis (FDA) called Locality preserving fisher discriminant analysis (LPFDA) (Zhao and Tian, 2009) was also proposed to preserve local structure of data. This method combines the idea of FDA and LPP to perform dimension reduction. LPFDA has the discriminating ability of FDA and the locality preserving ability of LPP. This characteristics of LPFDA makes it more efficient than FDA.

In this paper, we address an extension of LPPLS which utilize label information. We used the binary weighting technique that is used in graph based feature extraction methods to construct the similarity matrices in LPPLS. This approach makes LPPLS more suitable for discrimination between classes in the data. The rest of the paper is organized as follows. In section 2, we first give a short review of LPP. In section 3, we give a brief review of LPPLS and the newly proposed extended LPPLS method is presented in section 4. Some experimental results were presented in Section 5 to demonstrate the effectiveness of our proposed method. Concluding remarks are given in section 6.

# II. Locality Preserving Projections

Locality preserving projections is a prominent dimensionality reduction technique derived based on the idea of Laplacian eigenmaps (Belkin and Niyogi, 2003). Given a dataset $X = [x_1, x_2, \dots, x_n] \in R^{N\times n}$, where each data point $x_i \in R^N$. LPP determines a low dimensional representation, $y_i \in R^d$ $(d \ll N)$ of the data points in $X$ by solving the following objective function:

$$\min \sum\nolimits_{ij} (y_i - y_j)^2 S_{ij} \qquad (1)$$

where $S_{ij}$ measures the similarity between $x_i$ and $x_j$. The way in which $S_{ij}$ is computed is given below:

$$S_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{t}\right); & \text{if } x_i \text{ and } x_j \text{ are neighbors} \\ 0; & \text{Otherwise} \end{cases}$$

where $t$ is a user specified parameter. $S_{ij}$ incurs a heavy penalty if points $x_i$ and $x_j$ are mapped far apart. The minimization problem (Eq. 1) can be reduced to the form:

$$\arg \min_{\substack{\text{u} \\ u^T XDX^T u = 1}} u^T XLX^T u \qquad (2)$$

where $X = [x_1, x_2, \dots, x_n]$, $D$ is a diagonal matrix whose entries are column (or row) sum of

$S$, i.e $D_{ii} = \sum S_{ij}$ and $L = D - S$ is the Laplacian matrix (Chung, 1996). The transformation vector $u$ that minimizes the objective function is obtained as the minimum eigenvalue solution to the following generalized eigenvalue problem:

$$XLX^T u = \lambda XDX^T u$$

The matrices $XLX^T$ and $XDX^T$ are symmetric and positive semidefinite. Let the vectors $u_1, u_2, \ldots, u_d$ be the solutions to the generalized eigenvalue problem above, arranged according to their eigenvalues $\lambda_1 < \lambda_2 < \ldots < \lambda_d$. Then, the transformation to lower dimensional space can be performed as follows:

$$x_i \rightarrow y_i = U^T x_i$$

where $U = (u_1, u_2, \ldots, u_d)$ is a $n \times d$ matrix and $y_i$ is a $d$-dimensional vector.

## III. Locality Preserving Partial Least Squares (LPPLS)

Given two data matrices $X = [x_1, x_2, \ldots, x_n] \in R^{N\times n}$ and $Y = [y_1, y_2, \ldots, y_n] \in R^{M\times n}$, LPPLS starts by finding the nearest neighbors of each $x_i$ and $y_i$ based on the Euclidean distance measure. The similarities of data points in each of the data matrices ($X$ and $Y$) is then determined. The two similarity matrices $S^x = [S^x_{ij}]_{n\times n}$ and $S^y = [S^y_{ij}]_{n\times n}$ are then computed as follows:

$$S^x_{ij} = \begin{cases} \exp^{(-\frac{\|x_i - x_j\|^2}{t_x})}; & \text{if } x_i \in N(x_j) \text{ or } x_j \in N(x_i) \\ 0; & \text{Otherwise} \end{cases}$$

$$S^y_{ij} = \begin{cases} \exp^{(-\frac{\|y_i - y_j\|^2}{t_y})}; & \text{if } y_i \in N(y_j) \text{ or } y_j \in N(y_i) \\ 0; & \text{Otherwise} \end{cases}$$

where $N(x_i)$ represents the $k$-nearest neighbors set of the variable $x_i$, and parameters $t_x$ and $t_y$ are defined as:

$$t_x = \textstyle\sum_{i=1}^n \sum_{i=1}^n \frac{2\|x_i - x_j\|^2}{n(n-1)}$$
$$t_y = \textstyle\sum_{i=1}^n \sum_{i=1}^n \frac{2\|y_i - y_j\|^2}{n(n-1)}$$

Adding these similarity measures of data points into the objective function of partial least squares, a new objective function with locality information is obtained as:

$$\max_{\substack{\text{u,v} \\ u^T u=1, v^T v=1}} u^T X(D^{xy} - S^{xy})Y^T v \qquad (3)$$

where

$$S^{xy} = S^x \circ S^y \qquad (4)$$

the operator $\circ$ denotes the element wise product of the two similarity matrices $S^x$ and $S^y$ and $D^{xy}$ is a diagonal matrix define as:

$$D^{xy} = diag(\textstyle\sum_j S^{xy}_{ij}) \text{ for } i = 1, 2, ..., n$$

The optimization problem (Eq. 3) is further reduced to:

$$\begin{aligned} MM^T u &= \lambda^2 u \\ M^T M v &= \lambda^2 v \end{aligned} \qquad (5)$$

where $M = X(D^{xy} - S^{xy})Y^T$. A solution to Eq. 5 is $M = u\lambda v^T$ where $u$ and $v$ can be determined as the right and left singular vectors of $M$ and $\lambda$ denotes the corresponding singular value. After computing the vectors $u$ and $v$ using SVD, power method is employed to determine the transformation matrices $U$ and $V$.

## IV. Extended locality preserving partial least squares

In classification tasks, the class labels of data points are given. The LPPLS method in its original form does not uses this class labels information, it constructs the adjacency graph based on nearest neighbors. In this section, we proposed a different way of computing the similarity matrices in LPPLS such that the class labels information is utilized. The class label information is used to guide the process of constructing the adjacency graph. We define the two similarity matrices $S^x = [S^x_{ij}]_{n\times n}$ and $S^y = [S^y_{ij}]_{n\times n}$ using the binary weighting technique as:

$$S^x_{ij} = \begin{cases} 1; & \text{if } x_i \text{ and } x_j \text{ both belong to the same class} \\ 0; & \text{Otherwise} \end{cases}$$

$$S^y_{ij} = \begin{cases} 1; & \text{if } y_i \text{ and } y_j \text{ both belong to the same class} \\ 0; & \text{Otherwise} \end{cases}$$

These new definitions of the similarity matrices takes into consideration the class label information of the data points which makes our proposed method more powerful most especially in classification tasks than LPPLS. The objective function of this new method remains the same with that of the original LPPLS

$$\max_{\substack{\text{u,v} \\ u^T u=1, v^T v=1}} \quad u^T X(D^{xy} - S^{xy})Y^T v \qquad (6)$$

except that now, the matrix $S^{xy}$ is obtained by taking the element wise product of the two similarity matrices $S^x$ and $S^y$ obtained using the binary weighting technique as defined in this section. We further solved the minimization problem (Eq. 6) using Lagrange multiplier method and obtained a generalized eigenvalue problem:

$$XLY^TYLX^Tu = \lambda^2 u$$

The vector $u$ is then computed as the eigenvector corresponding to the smallest eigenvalue and vector $v$ is computed from the relation:

$$v = \tfrac{1}{\lambda} YLX^T u$$

The vectors $u$ and $v$ can also be computed as the right and left singular vectors of the matrix $XLY^T$. Transformation of the data points into lower dimensional space can then be performed as:

$$x_i \to p_i = U^T x_i$$
$$y_i \to q_i = V^T y_i$$

where $U = (u_1, u_2, \ldots, u_d)$ and $V = (v_1, v_2, ..., v_d)$ are $n \times d$ matrices.

# V. Experimental Results

In this section, experiments with real-world data sets are carried out to demonstrate the effectiveness of our proposed extended LPPLS. More specifically, the extended LPPLS method was used in class discrimination. We present two metrics for measuring the accuracy of our proposed and the LPPLS methods in our experiments. The performance of our proposed method compared to that of the LPPLS method was evaluated using these metrics. Since the LPPLS and extended LPPLS methods are related to the output data matrix $Y$, we set $Y$ to be the same as the input data matrix $X$ in all our experiments.

## A. Evaluation metrics

The discriminating results of our experiments are evaluated by comparing the actual labels of the datasets with that obtained by the algorithms. After transforming the datasets into lower dimensional space using the LPPLS and Extended LPPLS, we used Kmeans to obtain labels for the data points in the reduced space. the performance of the methods is then evaluated by comparing the labels obtained in the reduced space using Kmeans and and the actual label of the data points. Two standard metrics, normalized mutual information (NMI) and Fowlkes-Mallow index (FMI) are used to determine the accuracies of the algorithms. The NMI measure the similarity between two matrices of partitions $A$ and $B$ as (Alexander-Bloch et al., 2012, Kuncheva and Hadjitodorov, 2004):

$$NMI(A,B) = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} N_{ij} log(\frac{N_{ij}N}{N_i N_j})}{\sum_{i=1}^{C_A} N_i log(\frac{N_i}{N}) + \sum_{j=1}^{C_B} N_j log(\frac{N_j}{N})}$$

where $C_A$ denotes the number of modules in partition $A$, $C_B$ denotes the number of modules in partition $B$, $N$ denotes the total number of nodes, $N_i$ denotes the number of nodes in $A$'s module $i$, $N_j$ denotes the number of nodes in $B$'s module $j$ and $N_{ij}$ denotes the number of nodes common in both $A$'s module $i$ and $B$'s module $j$. The $NMI(A, B)$ ranges from zero to one, where zero indicates that the partitions are totally independent and one indicates that the partitions are identical. Larger value of NMI (values close to one) denotes the good performance of a clustering technique.

Apart from NMI, we also used the Fowlkes-Mallows index (Fowlkes and Mallows, 1983) to evaluate the discriminating ability of our proposed method and that of the original LPPLS method. The Fowlkes-Mallows index is also a metric used to determine the similarity between two partitions. Like NMI, the Fowlkes-Mallows index ranges from zero to one, the higher the value is, the more similar the two partitions are. The formula for the Fowlkes-Mallows index is defined as:

$$FM = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}$$

where $TP$ denotes the number of true positives, $FP$ denotes the number of false positives and $FN$ denotes the number of false negatives.

## B. Real world datasets

The proposed extended LPPLS method is applied to class discrimination problem using real-world data sets. We compared the performance of LPPLS with that of the proposed extended LPPLS. The actual labels of the data points in all the experiments are given. After dimensionality reduction using the LPPLS and extended LPPLS methods, we used Kmeans to partition the data and obtain labels for the data points in the lower dimensional space. The obtained labels and the actual labels are then compared to determine the accuracy of the two methods in discriminating the different classes in the data.

### 1. The Iris plant dataset

An experiment was conducted with the Iris dataset (Dheeru and Karra Taniskidou, 2017). The dataset contains 3 classes with 50 samples each. Each class refers to a particular type of iris plant (Iris Setosa, Iris Versicolor and Iris Virginica). Four features, the length and width of the sepals and petals were measured in centimeters for each sample. The iris plant dataset is a good data for testing the performance of a newly developed classification algorithm. The data contains two clusters, one of the clusters contains samples for the iris setosa species while the other cluster contains samples from both iris versicolor and iris virginica species. The cluster containing the iris setosa specie is well separated from the other cluster. A good discrimination algorithm will try to separate the two classes (iris versicolor and iris virginica) from each other. We mapped the iris data to two-dimensional space using both LPPLS and the extended LPPLS. In the LPPLS method, the number of nearest neighbors used for constructing the graph is set to be 5 and the parameters $t_x$ and $t_y$ in the Gaussian heat kernel similarity are set to be 1. The projection results of LPPLS and the Extended LPPLS are shown in Figure 1 and Figure 2 respectively.

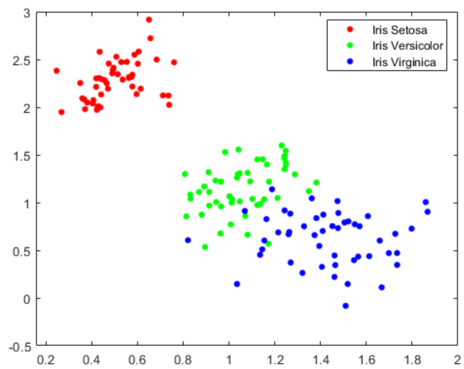


Figure 1: The LPPLS method applied to the iris plant data.

As can be seen from these figures, samples from different classes heavily overlapped in the projection result of LPPLS while samples belonging to the same class are mapped close to each other in the projection result of Extended LPPLS. The newly proposed extended LPPLS, performs much better than LPPLS in discriminating the classes in the data. This is confirmed by the NMI and Fowlkes-Mallows index obtained for the results of the two methods. the NMI value for the LPPLS method in this

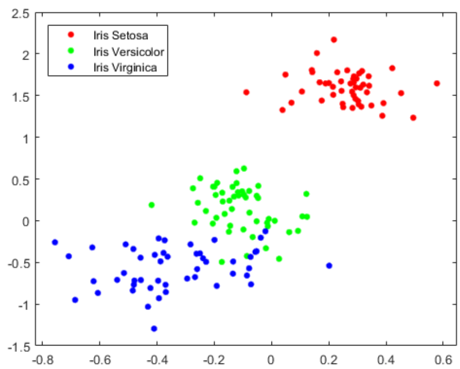


Figure 2: The extended LPPLS applied to the iris plant data.

experiment is 0.80, while the NMI value for the extended LPPLS method is 0.86. Also, the Fowlkes-Mallows value for the result obtained using LPPLS is 0.87 while the Fowlkes-Mallows value for the result obtained using the extended LPPLS is found to be 0.92. As can be seen from the NMI and the Fowlkes-Mallows values, the extended LPPLS outperforms the LPPLS method.

## 2. The banknote authentication dataset

Another experiment was conducted on the banknote authentication dataset (Dheeru and Karra Taniskidou, 2017). This data was extracted from images taken from genuine and forged banknote-like specimens. The images were taken for the evaluation of an authentication procedure for banknotes. The size of each image is $400 \times 400$ pixels. wavelet transformation tools are used to extract Features from the images. The extracted features are the variance, skewness, kurtosis and entropy of the images. The number of samples in this dataset is 1372. We used LPPLS and the extended LPPLS methods to project the data to two-dimensional space. The projection results of both the LPPLS and the extended LPPLS methods are shown in Figure 3 and Figure 4 respectively. Similar to the experiment conducted on the iris plant species, we set the number of nearest neighbor for the graph constructed in LPPLS to be 5 and the parameters $t_x$ and $t_y$ in the Gaussian heat kernel similarity are set to 1.

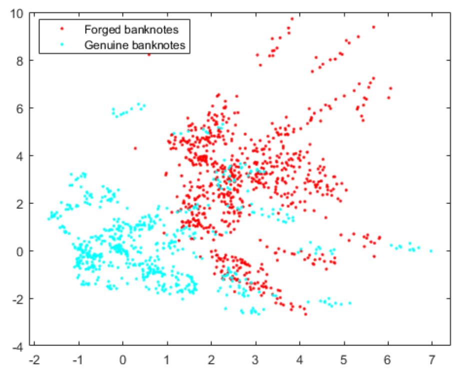


Figure 3: The projection result of the LPPLS method applied to the banknotes data.

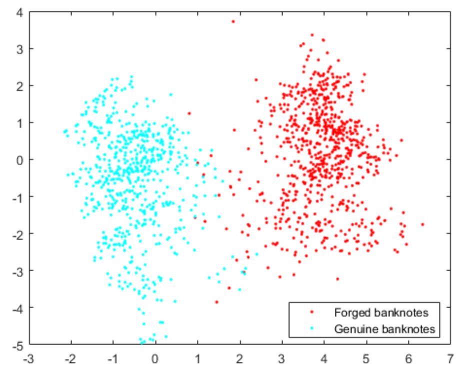


Figure 4: The projection result of the extended LPPLS method applied to the banknotes data.

As can be seen from these figures (Fig 3 and 4), the extended LPPLS perform much better than LPPLS. The two classes are well separated from each other in the result obtained using extended LPPLS, while samples from the two classes heavily overlapped in the LPPLS case. A lot of samples have been misclassified in the LPPLS result. To further show the better performance of our proposed method, we compute the NMI and Fowlkes-Mallow index for the partition results of the two methods. The NMI value for the result of LPPLS is 0.20

while that of the extended LPPLS is found to be 0.61. A Fowlkes-Mallow value of 0.66 was found for the LPPLS result while a much higher Fowlkes-Mallow value of 0.93 was obtained using the extended LPPLS method. These excellent results for the NMI and the Fowlkes-Mallow index shows the excellent performance of our proposed method.

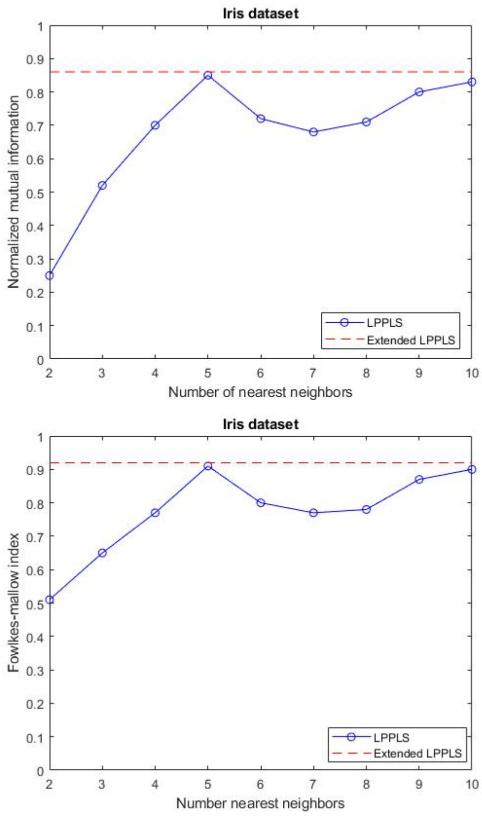


Figure 5: Performance of LPPLS with different neighborhood size on the Iris dataset

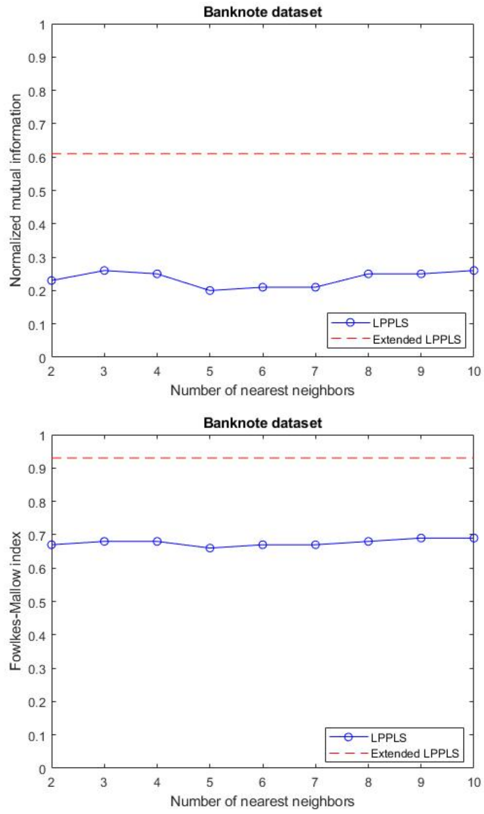


Figure 6: Performance of LPPLS with different neighborhood size on the Banknote dataset

## VI. Parameter selection

For the Extended LPPLS method, the only parameter is the subspace dimension. For LPPLS, there are two extra parameters, the neighborhood size $k$ and the the heat kernel width $t$. In our experiments, we set the value of these parameters by searching from a wide range of values and report the best results. In particular, for the Iris and Banknote datasets, we fix the value of the parameters $t_x$ and $t_y$ to $t_x = t_y = 1$ and search for the neighborhood size within the range $k = 2, 3, ..., 10$. We also fix the neighborhood size in the Iris and Banknote datasets to $k = 2$ and $k = 5$ respectively, and search for the heat kernel width $t_x$ and $t_y$ within the range $t_x = t_y = 0.1, 0.2, ..., 1$

As can be seen from Fig. 5 to 8, the performance of the Extended LPPLS is consistently better than the performance of the original LPPLS method. Our proposed method does not involve the use of any parameter while constrcting the adjacency graph. Thus, it can be seen from these figures that the performance of our method is fix.

## VII. Conclusion

This work introduces a new novel approach to LPPLS. When dataset contain class information, LPPLS does not make use of that class

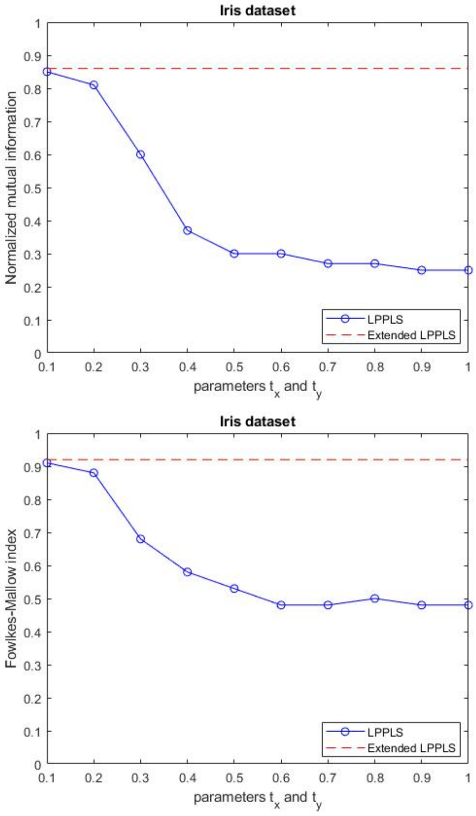


Figure 7: Performance of LPPLS with different heat kernel width on the Iris dataset

information. Studies have shown that, the class information of dataset can be used to improve the discriminating power of algorithms. Therefore, this study aims at finding the best way of improving the discriminating performance of LPPLS by making use of the class information of dataset if available. We used the simplest possible weighting technique (Binary weighting) to compute the similarity matrices in LPPLS. This new definition of the similarity matrices incorporates class information which helps in improving the discriminating ability of the LPPLS method. Experimental results on various datasets shows that the extended LPPLS method is more efficient, most especially in classification tasks than the original LPPLS method.

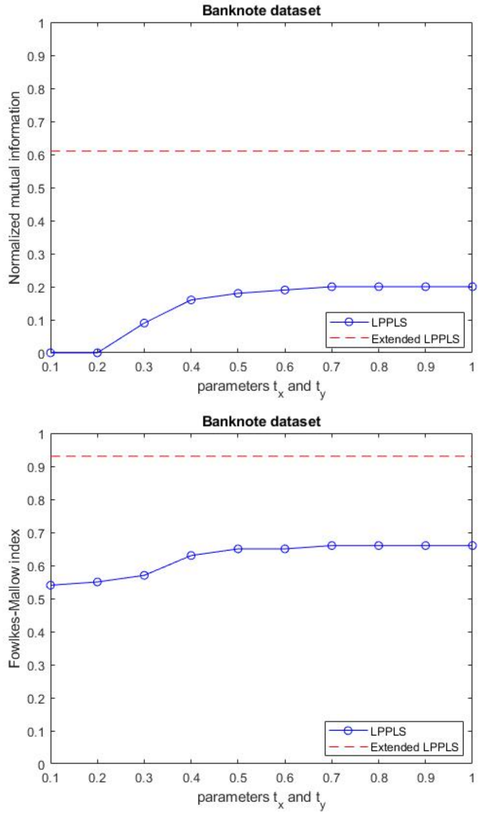


Figure 8: Performance of LPPLS with different heat kernel width on the Banknote dataset

# Acknowledgements

The work was supported by UNIVERSITI SAINS MALAYSIA through Research University Grant (RUI) (acc. No. 1001/PMATHS/8011040) and the USM Bridging Grant (acc. No. 304/PMATHS/6316049).

# References

[1] Aaron Alexander-Bloch, Renaud Lambiotte, Ben Roberts, Jay Giedd, Nitin Gogtay, and Ed Bullmore. The discovery of population differences in network community structure: new methods and applications to brain functional networks in schizophrenia. *Neuroimage*, 59(4):3889–3900, 2012.

[2] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques

for embedding and clustering. In *Advances in neural information processing systems*, pages 585–591, 2002.

[3] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[4] Fan RK Chung. Spectral graph theory (cbms regional conference series in mathematics, no. 92). 1996.

[5] Dua Dheeru and Efi Karra Taniskidou. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

[6] Edward B Fowlkes and Colin L Mallows. A method for comparing two hierarchical clusterings. *Journal of the American statistical association*, 78(383):553–569, 1983.

[7] Xiaofei He and Partha Niyogi. Locality preserving projections. In *Advances in neural information processing systems*, pages 153–160, 2004.

[8] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002. ISBN 9780387954424. URL https://books.google.com.my/books?id=_olByCrhjwIC.

[9] Ludmila I Kuncheva and Stefan Todorov Hadjitodorov. Using diversity in cluster ensembles. In *Systems, man and cybernetics, 2004 IEEE international conference on*, volume 2, pages 1214–1219. IEEE, 2004.

[10] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge university press, 2014.

[11] Yanwei Pang, Lei Zhang, Zhengkai Liu, Nenghai Yu, and Houqiang Li. Neighborhood preserving projections (npp): a novel linear dimension reduction method. In *International Conference on Intelligent Computing*, pages 117–125. Springer, 2005.

[12] Roman Rosipal and Nicole Krämer. Overview and recent advances in partial least squares. In *Subspace, latent structure and feature selection*, pages 34–51. Springer, 2006.

[13] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290 (5500):2323–2326, 2000.

[14] Kunju Shi, Shulin Liu, Hongli Zhang, and Bo Wang. Kernel local linear discriminate method for dimensionality reduction and its application in machinery fault diagnosis. *Shock and Vibration*, 2014, 2014.

[15] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[16] Zhaoqiang Zhang, Chun Qi, and Yuanhong Hao. Locality preserving partial least squares for neighbor embedding-based face hallucination. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 409–413. IEEE, 2016.

[17] Xu Zhao and Xiaoyan Tian. Locality preserving fisher discriminant analysis for face recognition. In *International Conference on Intelligent Computing*, pages 261–269. Springer, 2009.