doi:10.3969/j.issn.1673-4785.2010.01.015

Classification and recognition of image/non-image data based on multinomial logistic regression with nonlinear dimensionality reduction

Mudasser NASEER, QIN Shi-yin

(School of Automation Science and Electrical Engineering, Beihang University, Beijing 100037, China)

Abstract: In pattern classification and recognition oriented to massively complex data most classifiers suffer from the curse of dimensionality. Manifold learning based nonlinear dimensionality reduction (NLDR) methods provide a good preprocessing to reduce dimensionality before applying any classification method on high dimensional data. Multinomial logistic regression (MLR) can be used to predict the class membership of feature data. In this study several unsupervised NLDR methods are employed to reduce dimensions of the data and the MLR is used for class prediction of image/non-image data so that a new method of classification and recognition oriented to massively complex image/non-image data is proposed based on multinomial Logistic regression with nonlinear dimensionality reduction. Through a series of experiments and comparative analysis with supervised NLDR methods for a lot of typical test data the new proposed method is validated to outperform other supervised NLDR ones.

Keywords: nonlinear dimensionality reduction; data classification; multinomial logistic regression; image/non-image data

CLC Number: TP391　**Document code**: A　**Article ID**: 1673-4785(2010)01-0085-09

Many decision-making problems can be categorized as classification problems[1]. In a number of practical applications, digital images have been used to classifing objects; most examples have dealt with the classification of handwritten numbers, motor vehicle licence plates, images of human faces and so on. These images contain high dimensional data and require dimensionality reduction before they can be used for classification. Recently, a class of nonlinear dimensionality reduction methods, based on the concept of manifold learning, attracted researchers studying dimensionality reduction for nonlinear images and non-image data. Some conventional manifold based learning methods may be listed as locally linear embedding (LLE)[2], Isomap[3], Laplacian eigenmaps (LE)[4], diffusion maps[5], Hessian locally linear embedding (HLLE)[6], and local tangent space alignment (LTSA)[7]. These methods fall in the category of unsupervised learning and could not be directly used for the purpose of supervised learning of classification. For classification, supervised versions of most of the algorithms such as PCA-LLE and modified supervised LLE (MSLLE)[8], supervised LLE (SLLE)[9-10], WeightedIso[11], supervised Isomap[12] and supervised LTSA[13] were introduced from time to time.

For this paper, the idea of using unsupervised NLDR methods with MLR (U-NLDR + MLR) for image data classification was introduced and corresponding comparative analysis of their performance with some well known supervised versions of NLDR algorithms was carried out. For experimental and test purposes, some well known grey scale images of handwritten digits and human faces were used. The performance of various algorithms was evaluated by classification error rate (ER) for out-of-sample data points.

1 Classification

One problem in classification or supervised learning involves guessing or predicting unknown classes

Received Data: 2009-08-15.
Foundation Item: This work is supported by The Major Program of Hi-technology Research and Development (863) of China. (2008AA12A200) and Programs of National Natural Science Foundation of China (60875072).
Corresponding Author: QIN Shi-yin. E-mail: qsy@buaa.edu.cn.

based on observation. Typically speaking, a set of patterns of features along with a correct classification, known as training data, is available. Aftter training, the task is to classify a new set of patterns, known as test data. The training data set is assumed to be a random sample from some large population.

When the dimensionality of input data is relatively high, most classification methods, such as K-nearest neighbor (K-NN) classifier[14], will suffer from the curse of dimensionality and produce highly biased estimates. Most of the high dimensional data is intrinsically low dimensional. Thus, the problem of classification can be solved by firstly mapping the high dimensional data into low dimensional subspace by using a suitable NLDR method and then applying some classification method[9]. The above mentioned U-NLDR methods are not suitable for classification purpose. Some supervised versions of these algorithms were developed. In our study, we use four supervised NLDR methods—WeightedIso, supervised Isomap (S-Isomap), SLLE and SLTSA.

WeightedIso changed the first step of Isomap. It proceeded by first computing the K nearest neighbors of each data point x and denoted K_{same} as the set of nearest neighbors having the same class label as x. Then it "moved" each nearest neighbor in K_{same} closer to x by rescaling their Euclidean distance by a constant factor $1/\lambda(\lambda > 1)$. Remaining steps of the algorithm remain the same as of the unsupervised Isomap.

In S-Isomap Euclidean distance $d(x_i,x_j)$ is replaced by $D(x_i,x_j)$, where

$$D(x_i,x_j) = \begin{cases} \sqrt{1 - e^{\frac{-d^2(x_i,x_j)}{\beta}}}, & y_i = y_j; \\ \sqrt{e^{\frac{d^2(x_i,x_j)}{\beta}}} - \alpha, & y_i \neq y_j. \end{cases}$$

The parameter β is used to prevent $D(x_i,x_j)$ increasing too fast when $d(x_i,x_j)$ is relatively large. Usually, it is set to be the average Euclidean distance between all pairs of data points. The parameter α gives some opportunity to points in different classes to be "more similar".

Among different versions of SLLE, we choose one given in the reference[10], which used the idea of adding distance between samples in different classes as

$$\boldsymbol{D} = \boldsymbol{D} + \alpha\max(\boldsymbol{D})\Delta.$$

Where $\boldsymbol{D}$ is the pairwise distance matrix for combine data set, and Δ_{ij} equals to 1 if data points are from different classes, and 0 otherwise. Here $\alpha \in (0,1)$ controls the amount to which class information should be incorporated. This supervised version of LLE behaves as a nonlinear Fisher mapping which controls nonlinearity.

Supervised LTSA also used the idea of artificially increased shift distances between points belonging to different classes, but left them unchanged if samples were from the same class. The new pairwise distance matrix was given as $\boldsymbol{D}' = \boldsymbol{D} + \rho\Delta$, where the shift distance ρ is assigned a relatively large value in comparison with the distance between any pairs of points. Δ_i equals to 1 if data points are from different classes, and 0 otherwise.

We chose the above mentioned S-NLDR methods due to their similar approachs of using class information; they increased the distance between data points of different classes.

The S-NLDR methods do not explicitly provide any mapping function for out-of-sample data points, and it can be learnt by the estimation method[15] or by some nonlinear interpolation techniques, such as generalized regression networks (GRN)[16]. To summarize, the general classification procedure has three steps, as follows:

i. Map high dimensional data into a lower dimensional space using an S-NLDR method.

ii. Find mapping function for out-of-sample data points using an estimation method or GRN.

iii. Map the given query to low dimensional space using the mapping function and then predict its class label using K-NN.

2 Multinomial logistic regression

Multinomial logistic regression (MLR) is used to model the relationship between a multiple response variable and one or more predictor variables, which may be either discrete or continuous[17]. Let Y be a polychotomous random variable denoting the outcome of some experiment, and let $X = (x_1, x_2, \cdots, x_{p-1})$ be a collection of predictor variables. Then the conditional probability of outcome $P(Y=1/X)$ is presented by $\pi(x)$, where

$$\pi(x)=\frac{\exp(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_{p-1}x_{p-1})}{1+\exp(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_{p-1}x_{p-1})}.$$

If the x_j are varied and the n values $Y_1, Y_2, \cdots, Y_n$ of Y are observed, we write

$$\pi_i=\frac{\exp(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_{p-1}x_{ip-1})}{1+\exp(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_{p-1}x_{ip-1})}.$$

The problem is now to obtain an estimate of the vector $\hat{\boldsymbol{\beta}}=(\beta_0,\beta_1,\cdots,\beta_{p-1})$. As the number of significant predictor variables increases the predicted class probability becomes more accurate. MLR is a modal based approach and provides classification probability for individual objects as compared to the KNN method, which gives classification membership. We use U-NLDR methods to map the high dimensional image data into low dimensional subspace and then use an MLR model to estimate the classification probability for out-of-sample data points. The low dimensional embedded values are taken as an independent predictor variable. In contrast, class membership values are taken as values of the dependent variable Y. Since all the NLDR methods use eigen decomposition, the independence of predictor variables is obvious. We increase the number of predictor variables by increasing the dimensions of the embedded subspace to obtain more accurate estimates of $\hat{\boldsymbol{\beta}}$. An increase of predictor variables also give rise to the problem of including irrelevant or less important variables in the model. This problem is handled by checking the significance of the selected model and including the most important or significant predictor variables in the final model.

The procedure is summarized as follows:

i. Map high dimensional data into lower dimensional subspace using the U-NLDR method.

ii. Apply MLR to find a classification model.

iii. Find low dimensional mapping for out-of-sample data points using an estimation method or GRN.

iv. Find classification probability for out-of-sample data points by applying the MLR model obtained in (ii), and assign new data points to the class having maximum probability.

The flowchart is shown as Fig. 1.

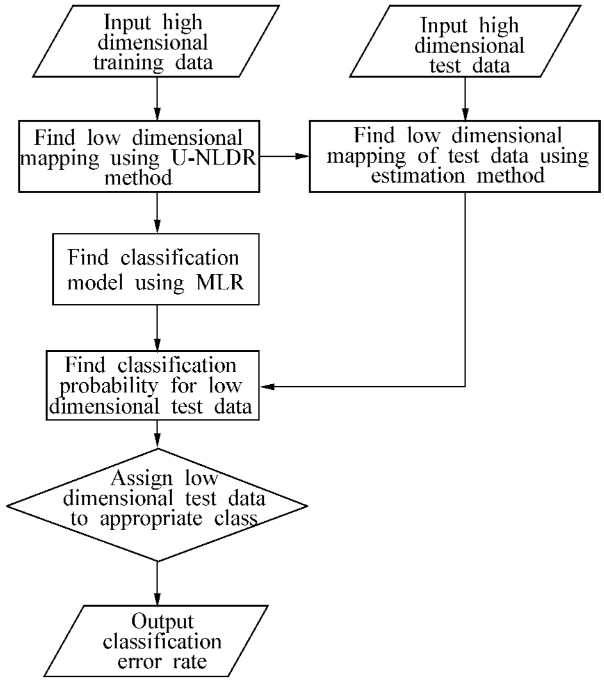


Fig. 1 Flowchart of U-NLDR + MLR algorithm

3 Performance evaluation

The most commonly used performance evaluation criteria for classification is the ER. If unlimited cases for training and testing are available, the error rate is simply the error rate on the test cases. The simplest technique for estimating error rates, the holdout method[18], is a single training and test experiment. The sample cases are broken into two groups of cases: a training group and a test group. The classifier is independently derived from the training cases, and the error estimate is the performance of the classifier on the test cases. A single random partition of training and test cases can be somewhat misleading. Random resampling can produce better estimates than a single train and test partition.

In this study, we used a 10-fold cross validation resampling method to find the error rate. That is, the original data set was randomly divided into ten equal-sized subsets. Then in each iteration, one subset was used as testing set and the union of the remaining ones was used as the training set. After ten iterations, the average result was taken as the final ER. For Olivitief-aces and the Lübeck University face images database, we used the leave-one-out resampling method due to the small number of images available for each person.

4 Test data

For the purpose of experimentation, we used seven different image and non-image datasets in this study. A brief description is given below:

i. The Yale face database-B: This dataset consists of images of 10 subjects, each seen under 576 viewing conditions (9 poses × 64 illumination conditions). For this study, we use 567 images of three persons (9 poses × 21 illumination conditions). The images are cropped and dimension reduced to 36 × 32. Fig. 2 shows some sample images.

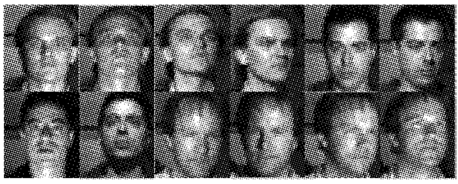


Fig. 2 Sample images from Yale-B database

ii. UMIST face images: This database consists of 564 images of 20 people. Each covers a range of poses from profile to frontal views. Again we use images of 3 persons. Dimension reduced to 45 × 37. Fig. 3 shows sample images of first subject.

Fig. 3 Sample images from UMIST database

iii. MNIST handwritten digits: MNIST handwritten digits consists of grayscale images of "0" through "9". The images are digitized as 28 × 28. We use 500 images of digits 1, 3 and 4. Fig. 4 shows sample images.

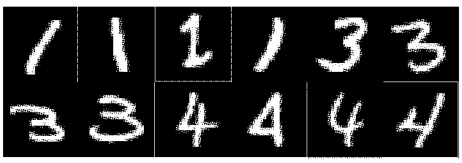


Fig. 4 Sample images from MNIST database

iv. Olivettiefaces database: The Olivettiefaces database consists of 400 images of 40 persons with different poses and expressions. We use images of 6 different persons. Image dimensions are 64 × 64. Fig. 5 shows some sample images.

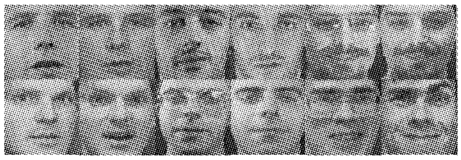


Fig. 5 Sample images from Olivettiefaces

v. Lübeck University face images: This database consists of face images of 31 persons, 13 images each, with different expression and light conditions. We use images of 5 persons. Images dimensions are 36 × 48. Fig. 6 shows some sample images.

Fig. 6 Sample images from Lübeck Uni. database

vi. UCI Isolet data set: This data set consists of 617 different features of spoken alphabets by 150 objects. Every subject speaks each alphabet twice. We use all instances of letters G, H, I, J, K and L.

vii. UCI optical digits: This data set consists of normalized bitmaps of handwritten digits (0 ~ 9) from a preprinted form. Total number of attributes are 64, whereas the total number of instances are 5 620.

5 Methodology

For a comparison between supervised versions of NLDR and U-NLDR + MLR, we ran six U-NLDR methods, namely Isomap, LLE, HLLE (results are not presented here due to very poor performance for all the data sets), LE, Diffusion Maps and LTSA. To find low dimensional embedding for out-of-sample test data, both estimation and GRN methods were used. In order to select significant predictor variables in the MLR model, we used the Wald test [15].

Under supervised NLDR we use four methods, namely WeightedIso, S-Isomap, SLLE and SLTSA. In the following experiments, the number of neighborhood points K was taken as 8 and 10, for both supervised and unsupervised methods. These values of K are most commonly used by different researchers. In our experiments these values were optimal ones for given data sets. Target dimension d varied from 2 to K. The parameter λ in WeightedIso was set to 10, the parameter α in S-Isomap was set to 0.5 and β was set to average the Euclidean distance of all pair-wise distances. These optimal parameter values we those suggested by other authors. The value of α in SLLE was 1 to make full use of class information, whereas the value of ρ in SLTSA was set to $\max(\boldsymbol{D})$. To predict class labels, the K-NN method was used.

Table 1 shows details about the number of categories, the number of images and the sampling methods used for different datasets.

Table 1　Summary information of datasets used

Dataset	No. of categories used	No. of samples	Sampling method
Yale-B	3	550	10-fold
UMIST	3	100	10-fold
MNIST	3	500	10-fold
Olivettiefaces	6	60	Leave-one-out
Lübeck Uni.	5	65	Leave-one-out
UCI Isolet	6	1 620	10-fold
UCI Optdigit	6	1 000	10-fold

6　Experimental results and analysis

We ran all supervised and unsupervised algorithms for the seven above mentioned data sets. Results for $K=10$ were marginally better than $K=8$, especially for supervised classification methods. As an out-of-sample mapping function, estimation method performed slightly better than GRN for both unsupervised and supervised methods. Keeping the above in mind, we presented results only for $K=10$, and with the estimation method as the mapping function. On a few occasions, we also presented results for $K=8$, but only when it gave better results.

Table 2 shows percentage mean error rate (PMER) for 10-fold resampling and $d=2$, for selected S-NLDR methods. S-Isomap produced disconnected geodesic distance graphs for all the data sets except Olivettiefaces. Table 3 ~ 9 show PMER for 10-fold or leave-one-out resampling for U-NLDR + MLR. Target dimensions varied from 2 to K. For every algorithm, the best performance is in bold.

Table 2　PMER for supervised NLDR Methods ($d=2$) %

Datasets	WeightedIso	S-Isomap	SLLE	SLTSA
Yale faces ($K=10$)	0.0	–	0.0	0.0
UMIST faces ($K=10$)	26.0	–	26.0	26.0
MNIST Digits ($K=10$)	1.8	–	0.0	1.6
Olivetttiefaces ($K=8$)	1.7	0.0	26.0	15.0
Lübeck Uni. ($K=8$)	7.7	–	1.6	0.0
UCI Isolet ($K=10$)	18.1	–	21.7	21.4
UCI Optdigit ($K=10$)	1.8	–	15.0	6.1

Table 3　PMER for Yale-B database　%

Algorithm	Target dimension (d)								
	2	3	4	5	6	7	8	9	10
Isomap	3.2	**0.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LLE	9.6	7.4	0.2	**0.0**	0.0	0.0	0.0	0.0	0.0
Laplacian	1.4	0.8	**0.0**	0.0	0.0	0.0	0.0	0.0	0.0
Diff. Maps	31.2	32.8	9.4	0.6	0.2	**0.0**	0.0	0.0	0.0
LTSA	14.6	14.8	6.6	1.6	0.8	**0.2**	10.8	60.8	64.6

Table 4　PMER for UMIST database　%

Algorithm	Target dimension (d)								
	2	3	4	5	6	7	8	9	10
Isomap	2.0	**1.0**	1.0	1.0	1.0	2.0	2.0	2.0	1.0
LLE	8.0	11.0	13.0	**1.0**	1.0	1.0	1.0	2.0	2.0
Laplacian	**1.0**	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0
Diff. Maps	9.0	**1.0**	6.0	1.0	1.0	1.0	1.0	1.0	1.0
LTSA	1.0	1.0	**0.0**	0.0	8.0	18.0	29.0	46.0	60.0

Table 5　PMER for MNIST database　%

Algorithm	Target dimension (d)								
	2	3	4	5	6	7	8	9	10
Isomap	11.0	8.6	7.2	7.6	7.4	6.6	5.4	5.6	**5.2**
LLE	8.4	5.0	**4.8**	4.8	5.4	5.4	5.0	5.2	5.0
Laplacian	10.6	6.0	5.6	5.6	5.6	5.6	5.4	**5.4**	5.6
Diff. Maps	5.4	5.6	5.6	5.2	5.2	5.2	4.8	**4.4**	4.4
LTSA	30.6	20.2	8.4	9.8	**8.2**	21.8	29.8	60.8	59.0

Table 6　PMER for Olivettiefaces　%

Algorithm	Target dimension (d)								
	2	3	4	5	6	7	8	9	10
Isomap	16.7	10.0	8.3	8.3	13.7	8.5	5.0	4.4	**4.0**
LLE	28.3	18.3	18.3	30.0	35.0	5.7	2.5	**1.1**	4.0
Laplac. ($K=8$)	20.0	13.3	40.0	68.3	46.7	25.7	**10.0**	–	–
Diff. Maps	48.3	36.7	35.0	6.67	3.33	1.43	1.25	7.78	**1.0**
LTSA ($K=8$)	25.0	**6.7**	40.0	20.0	40.0	48.0	57.5	–	–

Table 7　PMER for Lübeck University dataset　%

Algorithm	Target dimension (d)								
	2	3	4	5	6	7	8	9	10
Isomap	21.5	3.1	1.5	**0.0**	0.0	0.0	0.0	0.0	0.0
LLE	26.2	21.5	9.2	**0.0**	0.0	0.0	0.0	0.0	0.0
Laplacian	20.0	10.8	4.6	**0.0**	0.0	0.0	0.0	0.0	0.0
Diff. Maps	32.3	4.6	3.1	0.0	0.0	0.0	0.0	0.0	0.0
LTSA	32.3	26.2	23.1	16.9	12.8	**11.0**	18.3	29.9	27.7

Table 8　PMER for UCI Isolet dataset　%

Algorithm	Target dimension (d)								
	2	3	4	5	6	7	8	9	10
Isomap	17.4	15.6	14.7	9.9	7.9	8.3	7.6	**7.2**	6.7
LLE (k = 8)	56.2	49.4	44.4	37.7	**28.4**	33.2	43.8	-	-
Laplac. (k = 8)	31.8	27.8	24.4	**18.0**	25.9	26.1	34.6	-	-
Diff. Maps	17.0	11.1	11.2	8.5	8.9	8.1	**7.7**	8.3	-
LTSA	**30.3**	32.2	33.7	38.7	38.2	46.6	38.9	68.4	86.1

Table 9　PMER for UCI optical digits dataset　%

Algorithm	Target dimension (d)								
	2	3	4	5	6	7	8	9	10
Isomap	25.0	7.3	3.9	2.5	2.3	2.2	2.1	**1.7**	1.9
LLE	16.8	11.6	9.2	3.9	3.8	5.8	3.8	**2.3**	3.0
Laplacian	24.5	6.9	10.3	14.3	2.9	3.1	**2.6**	3.2	2.7
Diff. Maps	62.2	46.7	38.5	44.3	42.2	37.2	35.7	**17.0**	19.7
LTSA	20.8	17.8	17.9	**7.5**	9.9	12.1	42.6	72.2	82.1

In order to find out significant predictor variables (target dimensions) for a given method, we applyied the Wald test and then selected the most significant subset of predictor variables. Tables from 10 to 16 present PMER for different algorithms, along with reduced dimensions, which are in parenthesis below the PMER values. A quick overview of these tables reveals that for most of the datasets, the value of PMER very quickly reaches its minimum value along with the optimal number of significant predictor variables. The value of PMER fluctuates around this minimum value even if we continue to include additional predictor variables (dimensions) in the MLR model. For LTSA, however, the value of PMER increases with increased target dimensions after it touches the minimum value. It should be kept in mind that reduced dimensions do not represent first d dimensions, rather these are best subsets selected from actual full dimensions. Table 17 provides a summary comparison between PMER under full and reduced MLR models. In most cases, the reduced MLR model provides same accuracy, even better in some cases, with decreased dimensions compared to the full MLR model.

Table 10　PMER for Yale-B DB under reduced dimensions　%

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Isomap	20.4	2.6	**0.0**	0.0	0.0	0.0	0.0	0.0	0.0
	(1)	(2)	**(3)**	(3)	(3)	(3)	(3)	(3)	(3)
LLE	14.2	13.2	3.6	**0.0**	0.0	0.0	0.0	0.0	0.0
	(1)	(2)	(2)	**(3)**	(3)	(3)	(3)	(3)	(3)
Laplacian	21.2	5.0	**1.0**	1.0	1.0	1.0	1.0	1.0	1.0
	(1)	(2)	**(2)**	(2)	(2)	(2)	(3)	(3)	(3)
Diff. Maps	30.2	32.8	8.8	2.0	1.2	**0.0**	0.0	0.0	0.0
	(1)	(2)	(2)	(4)	(5)	**(5)**	(5)	(5)	(5)
LTSA	27.0	22.2	14.6	3.2	0.6	**0.0**	12.6	72.4	72.4
	(1)	(2)	(2)	(4)	(4)	**(5)**	(6)	(0)	(0)

Table 11　PMER for UMIST DB under reduced dimensions　%

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Isomap	11.0	**1.0**	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	(1)	**(2)**	(3)	(4)	(2)	(2)	(3)	(3)	(3)
LLE	24.0	8.0	2.9	1.7	9.5	6.3	1.7	**1.0**	1.0
	(1)	(2)	(1)	(2)	(1)	(1)	(2)	**(2)**	(2)
Laplacian	4.0	**3.0**	3.0	4.0	4.0	3.0	3.0	3.0	3.0
	(1)	**(1)**	(1)	(1)	(1)	(1)	(1)	(1)	(1)
Diff. Maps	28.0	9.0	30.0	**1.0**	1.0	1.0	1.0	1.0	1.0
	(1)	(2)	(2)	**(3)**	(3)	(3)	(3)	(3)	(3)
LTSA	10.0	10.0	**1.0**	10.0	5.0	26.0	34.0	59.0	59.0
	(1)	(1)	**(3)**	(2)	(2)	(4)	(6)	(0)	(0)

Table 12　PMER for MNIST DB under reduced dimensions %

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Isomap	21.4	10.4	8.2	7.8	7.0	5.8	**5.6**	6.0	5.6
	(1)	(2)	(3)	(4)	(5)	(5)	**(5)**	(6)	(5)
LLE	10.6	7.8	5.2	5.2	5.2	5.4	**5.0**	5.2	5.0
	(1)	(2)	(1)	(3)	(4)	(4)	**(5)**	(5)	(6)
Laplacian	13.8	8.4	6.0	6.0	5.6	5.4	5.4	**5.2**	5.2
	(1)	(2)	(3)	(3)	(3)	(5)	(5)	**(6)**	(6)
Diff. Maps	13.0	13.0	6.8	5.4	5.4	5.0	5.2	**4.6**	4.8
	(1)	(1)	(2)	(3)	(2)	(4)	(4)	**(6)**	(7)
LTSA	40.0	45.0	**11.8**	14.0	13.4	28.6	33.0	63.0	63.4
	(1)	(1)	**(3)**	(4)	(5)	(5)	(7)	(1)	(0)

Table 13　PMER for Olivettiefaces under reduced dimensions %

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Isomap	33.3	18.3	8.3	6.7	6.7	5.7	6.2	4.4	**4.0**
	(1)	(2)	(3)	(4)	(5)	(6)	(6)	(5)	**(5)**
LLE	43.3	16.7	11.7	15.0	3.3	8.5	2.5	**2.0**	4.4
	(1)	(2)	(3)	(3)	(4)	(4)	(5)	**(5)**	(5)
Laplacian	40.0	15.0	20.0	10.0	21.7	10.0	**7.5**	–	–
(K = 8)	(1)	(2)	(2)	(3)	(5)	(6)	**(6)**		
Diff. Maps	63.3	40.0	15.0	5.0	3.3	4.3	2.5	**2.2**	3.0
	(1)	(2)	(3)	(4)	(4)	(4)	(4)	**(4)**	(5)
LTSA	51.7	30.0	38.3	15.0	16.7	**4.3**	17.5	–	–
(K = 8)	(1)	(2)	(3)	(4)	(5)	**(5)**	(7)		

Table 14　PMER for Lübeck Uni. under reduced dimensions %

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Isomap	44.6	33.8	3.1	**0.0**	0.0	0.0	0.0	0.0	0.0
	(1)	(2)	(3)	**(4)**	(4)	(4)	(4)	(4)	(4)
LLE	47.7	35.4	13.8	4.6	6.3	**0.0**	0.0	0.0	0.0
	(1)	(2)	(3)	(4)	(4)	**(5)**	(5)	(5)	(5)
Laplacian	27.7	16.9	6.2	**0.0**	0.0	0.0	0.0	0.0	0.0
	(1)	(2)	(3)	**(4)**	(4)	(4)	(4)	(4)	(4)

After table 14

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Diff. Maps	41.5	29.2	12.3	6.2	2.6	**1.1**	1.9	1.7	1.5
	(1)	(2)	(3)	(4)	(4)	**(4)**	(4)	(4)	(4)
LTSA	52.3	41.5	32.3	27.7	19.2	**12.1**	19.2	40.2	40.0
	(1)	(2)	(3)	(3)	(4)	**(5)**	(7)	(0)	(0)

Table 15　PMER for UCI Isolet DB under reduced dimensions %

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Isomap	26.1	19.8	15.6	11.2	10.1	8.5	7.3	7.3	**6.9**
	(1)	(2)	(3)	(4)	(5)	(6)	(6)	(7)	**(8)**
LLE	70.3	53.4	36.4	20.1	**18.9**	20.2	24.6	–	–
(k = 8)	(1)	(2)	(3)	(4)	**(5)**	(5)	(6)		
Laplacian	51.9	30.7	28.0	**15.0**	16.4	15.2	15.0	–	–
(k = 8)	(1)	(2)	(2)	**(4)**	(4)	(4)	(4)		
Diff. Maps	41.0	15.4	13.1	10.0	8.9	8.6	**7.8**	8.7	7.8
	(1)	(2)	(3)	(4)	(5)	(6)	**(7)**	(8)	(9)
LTSA	47.5	34.0	35.6	33.7	**32.8**	33.5	41.4	68.9	87.1
	(1)	(2)	(2)	(3)	**(3)**	(4)	(5)	(6)	(0)

Table 16　PMER for UCI Optdigit under reduced dimensions %

Algorithm	Actual dimension								
	2	3	4	5	6	7	8	9	10
Isomap	25.0	7.3	3.9	2.5	2.3	2.2	2.1	**1.7**	1.9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	**(7)**	(8)
LLE	16.8	19.7	11.0	4.1	4.5	5.8	3.0	**2.8**	3.0
	(1)	(2)	(3)	(4)	(5)	(6)	(6)	**(7)**	(8)
HLLE	83.1	80.3	79.4	80.3	80.4	79.3	80.0	**80.9**	78.5
	(1)	(2)	(3)	(4)	(4)	(5)	(6)	**(8)**	(8)
Laplacian	24.5	6.9	10.3	14.3	2.9	3.1	**2.6**	3.2	2.7
	(1)	(2)	(2)	(3)	(5)	(6)	**(6)**	(6)	(6)
Diff. Maps	62.2	46.7	38.5	44.3	42.2	37.2	35.7	**17.0**	18.3
	(0)	(0)	(2)	(4)	(5)	(5)	(6)	**(7)**	(8)
LTSA	20.8	17.8	17.9	7.49	**9.9**	12.1	42.6	72.2	82.1
	(1)	(2)	(2)	(4)	**(4)**	(6)	(7)	(6)	(0)

Table 17 Summary comparison between PMER under full and reduced MLR models %

Datasets	Isomap		LLE		Laplacian		Diff Maps		LTSA	
	Actual dim.	Reduced dim.	Actual dim.	Reduced dim.	Actual dim.	Reduced dim.	Actual dim.	Reduced dim.	Actual dim.	Reduced dim.
Yale - B	0.0 (3)	0.0 (3)	0.0 (5)	0.0 (3)	0.0 (4)	1.0 (2)	0.0 (7)	0.0 (5)	0.2 (7)	0.0 (5)
UMIST	1.0 (3)	1.0 (2)	1.0 (5)	1.0 (2)	1.0 (2)	3.0 (1)	1.0 (3)	1.0 (3)	0.0 (4)	1.0 (3)
MNIST	5.2 (10)	5.6 (5)	4.8 (4)	5.0 (5)	5.4 (9)	5.2 (6)	4.4 (9)	4.6 (6)	8.2 (6)	11.8 (3)
Olivettiefaces	4.0 (10)	4.0 (5)	1.1(9)	2.0 (5)	10.0 (8)	7.5 (6)	1.0 (10)	2.2 (4)	6.7 (3)	4.3 (5)
Lübeck Uni.	0.0 (5)	0.0 (4)	0.0 (5)	0.0 (5)	0.0 (5)	0.0 (4)	0.0 (5)	1.1 (4)	11.0 (7)	12.1 (5)
UCI Isolet	7.2 (9)	6.9 (8)	28.4 (6)	18.9 (5)	18.0 (5)	15.0 (4)	7.7 (8)	7.8 (7)	30.3 (2)	32.8 (3)
UCI Optdigit	1.7 (9)	1.7 (7)	2.3 (9)	2.8 (7)	2.6 (8)	2.6 (6)	17.0 (9)	17.0 (7)	7.5 (5)	9.9 (4)

7 Conclusion

An overview of the above tables reveals that no single algorithm, using supervised or unsupervised methods, performed best for all datasets. Among S-NLDR methods, WeightedIso has a slight edge on other methods. In contrast, S-Isomap failed to produce meaningful low dimensional embedding in most cases. Among unsupervised NLDR methods, Isomap and LLE performed better than other methods.

While comparing U-NLDR + MLR with S-NLDR we observed that for Isomap and LLE, U-NLDR + MLR performed far better than S-NLDR, whereas for LTSA it was other way round. The U-NLDR + MLR framework also produced promising results for Laplacian and diffusion maps by providing comparable results with other supervised and unsupervised methods.

The overall performance of U-NLDR algorithms in conjunction with MLR leads us to the following general conclusions:

i. The use of MLR in conjunction with unsupervised NLDR methods can be used as a general framework for classification purposes, especially for LLE, Isomap, LE and Diffusion Maps. It overcomes the need for developing different supervised versions for different NLDR methods.

ii. For all the algorithms, except LTSA, error rates decreased as the target dimensions approached K.

iii. Suitable number of target dimensions depends on the properties of datasets and may vary from 3 to K. Best subsets can be obtained by selecting significant variables for the MLR model.

References:

[1] JAMES M. Classification algorithms[M]. New York: John Wiley & Sons Inc, 1985.

[2] ROWEIS S T, SAUL L K, Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290: 2323-2326.

[3] TENENBAUM J B, SILVA V D, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290: 2319-2323.

[4] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering[J]. Neural Computations, 2003, 15(6): 1373-1396.

[5] COIFMAN R R, LAFON S. Diffusion maps[J]. Applied and Computational Harmonic Analysis, 2006, 21(1): 5-30.

[6] DONOHO D L, GRIMES C E. Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data [J]. Proceedings of the National Academy of Sciences of the USA, 2003, 100(10): 5591-5596.

[7] ZHANG Zhenyue, ZHA Hongyuan. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment[J]. SIAM Journal on Scientific Computing, 2004, 26 (1): 313-338.

[8] BUSA F K, KOCSOR A. Locally linear embedding and its variants for feature extraction [C]//IEEE International Workshop on Soft Computing Applications (SOFA). Szeged, Hungary and Arad, Romania, 2005: 216-222.

[9] RIDDER D, DUIN R P W. Locally linear embedding for classification. Delft: PH-2002-01[R]. Pattern Recognition Group, Dept of Image Science & Technology, Delft University of Technology. Delft, Netherlands, 2002.

[10] RIDDER D, KOUROPTEVA O, OKUN O, et al. Supervised locally linear embedding[C]//Proceedings of Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP). Istanbul, Turkey, 2003: 333-341.

[11] VLACHOS M, DOMENICONI C, GUNOPULOS D, et al. Non-linear dimensionality reduction techniques for classification and visualization [C]//Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton, Canada, 2002: 645-651.

[12] GENG Xin, ZHAN Dechuan, ZHOU Zhihua. Supervised nonlinear dimensionality reduction for visualization and

classification[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 2005, 35(6): 1098-1107.

[13]LI Hongyu, CHEN Wenbin, SHEN Ifan. Supervised local tangent space alignment for classification[C]//Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence. Edinburgh, Scotland, 2005: 1620-1621.

[14]ROSS D A, LIM J, LIN R S, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1): 125-141.

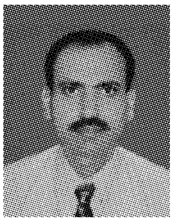
[15]BENGIO Y, PAIEMENT J-F, VINCENT P, et al. Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering[J]. Advances in Neural Information Processing Systems, 2004(16): 177-184.

[16]HO T K. Nearest neighbors in random subspaces[J]. LNCS: Advances in Pattern Recognition, 1998, 1451: 640-648.

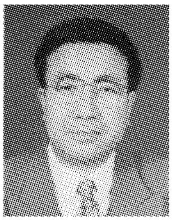
[17]HOSMER D W, LEMESHOW S. Applied logistic regersion[M]. 2nd ed. New York: John Wiley & Sons Inc, 2000: 128-135.

[18]FUKUNAGA K. Introduction to statistical pattern recognition[M]. 2nd ed. San Diago: Academic Press, 1990: 219-237.

About the authors:



Mudasser NASEER received his Master's and M. Phil degrees in Statistics from Govt. College University Lahore, Pakistan, in 1990 and 2001. He completed his MS in CS from LUMS Lahore in 2004. Presently he is pursuing his PhD in Pattern Recognition from Beihang University, Beijing, China. His field of interest is pattern recognition, machine learning and heuristic optimization techniques. So far he has published two research papers.



QIN Shi-yin, received his Bachelor Degree and Master's Degree for Engineering Science in Automatic Controls and Industrial Systems Engineering from Lanzhou Jiaotong University in 1978 and 1984 respectively, and his PhD Degree in Industrial Control Engineering and Intelligent Automation from Zhejiang University in 1990. He is now a professor at the School of Automation Science and Electrical Engineering in Beihang University(Beijing University of Aeronautics and Astronautics). He is a standing member of the council and the secretary-general of the Chinese Association for Artificial Intelligence (CAAI), and the vice-chairman of the Society of Intelligent Control and Intelligent Management in CAAI, and is also a member of the Committee of Intelligent Automation of the Chinese Association of Automation (CAA). He has also been a professor in Xi'an Jiaotong University and Beijing University of Technology. His current research interests include intelligent autonomous controls of complex spacecraft; autonomous intelligent control of formation processes for multi-robot systems; theory of hybrid control systems with applications; fault diagnosis and fault-tolerant controls; image processing and pattern recognition; intelligent systems and artificial life; the modeling and optimizing decision of open complex giant systems; computational intelligence and entropy optimization. He is the author of 3 monographs. He was awarded the First Level Prize of "'1999' National Excellent Books of Science and Technology and the Progress of Science and Technology", and the Gold Medal Prize for the Excellent Software in the "5th National Engineering Design" (1999). He has published more than 150 papers in journals and proceedings in the fields of intelligent controls for complex spacecrafts and large scale systems, hybrid control systems, systems engineering, artificial intelligence, neural networks, fuzzy control systems, expert systems, evolutionary computation and entropy optimization, and so on.

基于非线性降维多项式逻辑斯蒂回归的图像/非图像数据的分类与识别

Mudasser NASEER, 秦世引

(北京航空航天大学 自动化科学与电气工程学院,北京 100037)

摘　要:在面向大规模复杂数据的模式分类和识别问题中,绝大多数的分类器都遇到了维数灾难这一棘手的问题.在进行高维数据分类之前,基于监督流形学习的非线性降维方法可提供一种有效的解决方法.利用多项式逻辑斯蒂回归方法进行分类预测,并结合基于非线性降维的非监督流形学习方法解决图像以及非图像数据的分类问题,因而形成了一种新的分类识别方法.大量的实验测试和比较分析验证了本文所提方法的优越性.

关键词:非线性降维;数据分类;多项式逻辑斯蒂回归;图像/非图像数据

中图分类号:TP391　**文献标识码:**A　**文章编号:**1673-4785(2010)01-0085-09