Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 27 (2006) 968–979
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Selection of the optimal parameter value for the Isomap algorithm

O. Samko *, A.D. Marshall, P.L. Rosin

School of Computer Science, Cardiff University, 5 The Parade Roath, Cardiff CF24 3AA, United Kingdom

Received 12 October 2004; received in revised form 17 November 2005
Available online 17 February 2006

Communicated by M.A.T. Figueiredo

Abstract

The isometric feature mapping (Isomap) method has demonstrated promising results in finding low-dimensional manifolds from data points in high-dimensional input space. Isomap has one free parameter (number of nearest neighbours K or neighbourhood radius ϵ), which has to be specified manually. In this paper we present a new method for selecting the optimal parameter value for Isomap automatically. Numerous experiments on synthetic and real data sets show the effectiveness of our method.

Keywords: Nonlinear dimensionality reduction; Manifold learning; Isomap

1. Introduction

Usually, data from the real world is of a high-dimensional nature, and so is very difficult to understand and analyse. There are a number of known dimensionality reduction techniques, which attempt to solve this problem, letting users analyse or visualise complex data sets better. These techniques may be separated into two classes, linear and nonlinear. The examples of linear methods are principal component analysis (PCA, Jolliffe, 1986) or the original metric multidimensional scaling (MDS, Torgerson, 1952). Isomap (Tenenbaum et al., 2000), locally linear embedding (LLE) (Roweis and Saul, 2000), Hessian LLE (Donoho and Grimes, 2003), Laplacian eigenmaps (Belkin and Niyogi, 2003), geodesic nullspace analysis (GNA) (Brand, 2004) and nonlinear PCA (Karhunen and Joutsensalo, 1994) are in the second class. These algorithms are used to reveal low-dimensional manifolds that are not detected by classical linear methods, thus representing the original data more accurately.

All the above nonlinear algorithms share the same basic approach, consisting of three steps:

1. computing neighbourhoods in the input space;
2. constructing a square matrix with as many rows as elements in the input data set;
3. calculating spectral embedding using the eigenvectors of this matrix.

Here one usually has to manually specify the number of neighbours used in the first step. We call this number—the *free parameter*. There are a number of problems associated with the choice of this parameter. A large number of nearest neighbours causes smoothing or elimination of small-scale structures in the manifold. In contrast, too small a neighbourhood can falsely divide the continuous manifold into disjoint sub-manifolds.

Kouropteva et al. (2002) presented an automatic method used to detect the optimal value of this parameter for LLE. We have extended this idea to choose the optimal number of neighbours K and the neighbour region ϵ for Isomap.

* Corresponding author. Fax: +44 29 2087 4598.
E-mail addresses: O.Samko@cs.cf.ac.uk, oksana@samko.info (O. Samko), Dave.Marshall@cs.cf.ac.uk (A.D. Marshall), Paul.Rosin@cs.cf.ac.uk (P.L. Rosin).

0167-8655/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2005.11.017

In this paper we introduce the new method for choosing the neighbourhood size for Isomap and demonstrate how it works, providing several examples. We also compare the mapping produced by Isomap with the optimal parameter chosen by our algorithm with mappings produced by PCA, LLE and Isomap with different parameter value for the same data set. We use the following approaches to compare the results:

- Using various measures, such as residual variance, Spearman ρ, Pearson's correlation coefficient (see Friedrich, 2002; Navarro and Lee, 2001). For example, Navarro and Lee (2001) applied Isomap and MDS to the graphical depiction of the document set. They employed the variance and spatial visualisation to compare these techniques. Friedrich (2002) compared Isomap, LLE and PCA methods on different data sets, synthetic and real, using correlation coefficient and visualisation.
- Using classification (i.e. comparing how well these methods recognise images from the test data set). Since LLE and Isomap are unsupervised methods, we need to use their extensions to classify images. Yang (2002) has extended Isomap to improve classification with a Fisher linear discriminant (FLD) phase. de Ridder and Duin (2002) and de Ridder et al. (2003) presented supervised LLE (SLLE). Kouropteva et al. (2003) applied SLLE combined with support vector machines (SVM) for classifying handwritten digits.

The purpose of these comparisons is to prove that our algorithm can produce the reasonable result. Our experiments show that Isomap with the automatically selected optimal parameter (K or ϵ) performs better than PCA and LLE with various data sets, see Section 4.

2. Isomap algorithm

The Isomap technique is based on the idea of viewing the problem of creating a high dimension to low dimension transformation as a graph problem. The Isomap algorithm extends the classical techniques of principal component analysis (PCA) and multidimensional scaling (MDS) to a class of nonlinear manifolds.

On input the Isomap algorithm requires the distances $d_X(i,j)$ between all pairs i, j from N data points in the high-dimensional input space X, measured using either the standard Euclidean metric, or some domain-specific metric.

The algorithm outputs coordinate vectors Y_i in a (lower) d-dimensional Euclidean space Y that best represents the intrinsic geometry of the data.

The complete isometric feature mapping, or Isomap, algorithm has three steps:

1. The estimation of the neighbourhood graph. The first step determines which points are neighbours on the manifold M, based on the distances $d_X(i,j)$ between pairs of input points i, j in the input space X. Given the input points, we determine the set of neighbours for each point either by K nearest neighbours (K-Isomap) or all those within some fixed radius ϵ (ϵ-Isomap). These neighbourhood relations are represented as a weighted graph G over the data points, with edges of weight $d_X(i,j)$ between neighbouring points. Note, that in the case of K-Isomap, vertices in the graph may have degree greater than K since the K nearest neighbourhood relationship need not be symmetric.
2. Computing the shortest path graph given the neighbourhood graph. In its second step, Isomap estimates the geodesic distances $d_M(i,j)$ between all pairs of points on the manifold by computing the shortest path lengths $d_G(i,j)$ in the graph G. First we set $d_G(i,j) = d_X(i,j)$ if i, j are linked by an edge, and $d_G(i,j) = \infty$ otherwise. Then for each value of $k = 1, 2, \ldots, N$, we replace all entries $d_G(i,j)$ by $\min\{d_G(i,j), d_G(i,k) + d_G(k,j)\}$. The matrix of final values $D_G = \{d_G(i,j)\}$ will contain lengths of the shortest paths between all pairs of points in G.
3. Construction of lower dimensional embedding. The final step applies classical MDS to the matrix of graph distances $D_G = \{d_G(i,j)\}$, constructing an embedding of the data in a d-dimensional Euclidean space that best preserves the manifold's estimated intrinsic geometry. The coordinate vectors Y_i for points in Y are chosen to minimise the cost function

$$E = \|\tau(D_G) - \tau(D_Y)\|_{L^2} \tag{1}$$

where D_Y denotes the matrix of Euclidean distances $\{d_Y(i,j) = \|y_i - y_j\|\}$ and $\|A\|_{L^2}$ is the matrix norm $\sqrt{\sum_{i,j} A_{ij}^2}$. The τ operator converts distances to inner products, which uniquely characterise the geometry of the data in a form that supports efficient optimisation (Tenenbaum et al., 2000).

The only free parameter is the neighbourhood factor K or ϵ, which appears in the first step. Tenenbaum et al. (2002) pointed out that the success of Isomap depends on being able to choose an appropriate neighbourhood size. In the original Isomap paper (Tenenbaum et al., 2000) the parameter value was chosen manually.

3. Automatic selection of the optimal parameter value

How does one correctly choose the optimal parameter value for the Isomap? Neighbourhood selection is quite critical. If the neighbourhood is too large, the local neighbourhoods will include the data points from other branches of the manifold, shortcutting them, and leading to substantial errors in the final embedding. If it is too small, it will lead to discontinuities, causing the manifold to fragment into a large number of disconnected clusters.

Here, we present a method for determining the optimal neighbourhood size, inspired by the approach of Kouropteva et al. (2002) for a similar LLE parameter selection. By

Table 1
Algorithm for automatic selection of the optimal parameter value

1. Choose the interval of possible values of K, $K_{opt} \in [K_{min}, K_{max}]$
   - K_{min} is the minimal value of K, with which the neighbourhood graph (from the second step of Isomap) is connected
   - Maximal values:
     - choose K_{max} as the largest value of K in the equation

       $$\frac{2 * P}{N} \leqslant K + 2, \tag{2}$$

       here P is the number of edges and N is the number of nodes in the neighbourhood graph from the second step of the Isomap algorithm
     - choose ϵ_{max} as $\max(d_X(i,j))$, where $d_X(i,j)$ is the Isomap input matrix
2. Calculate the cost function $E(K)$ (see Eq. (1)) for each $K \in [K_{min}, K_{max}]$
3. Compute all minima of $E(K)$ and corresponding K which compose the set S_K of initial candidates for optimal value
4. For each $K \in S_K$ we run Isomap and determine K_{opt} using formula

   $$K_{opt} = \arg\min_K (1 - \rho^2_{D_x D_y}), \tag{3}$$

   where D_x and D_y are the matrices of Euclidean distances between pairs of points in input ant output spaces, respectively, and ρ is the standard linear correlation coefficient, taken over all entries of D_x and D_y

"optimal" here we mean that with this parameter Isomap will solve the dimensionality reduction problem most accurately (finding meaningful low-dimensional structures of data hidden in their high-dimensional observations).

The scale-invariant K parameter is typically easier to set than the neighbourhood radius ϵ, but Tenenbaum et al. (2000) noted that when the local dimensionality varies across the data set, the approach which uses K nearest neighbours, may yield misleading results.

We will describe our algorithm relative to K, making notes about ϵ if necessary. The algorithms for the automatic selection of optimal values of K and ϵ are similar, excluding the first step. Automatic selection of the optimal parameter value algorithm consists of four steps, which are detailed in Table 1.

During the first step, we choose the interval containing the possible optimal values of K or ϵ. The minimal values of K and ϵ are chosen using the same approach. We take the minimum possible K or ϵ such that the corresponding graph (produced by the second step of Isomap) is connected (with small parameter values this graph is disconnected).

The procedures we use to choose the upper bound are different for K and ϵ. We take the upper bound for ϵ to be the diameter of our input data set (we define this diameter as the maximum distance between any two points in the input space). In the case of K, after completing numerous experiments and analysing the resulting data, we noticed that in all the considered cases, if the average node degree of the graph (computed with some value of K) is greater than $K + 2$, one usually gets shortcuts which do not follow the surface of the manifold. Therefore we choose the upper bound of the minimum value of K to be the maximum value of K for which (2) holds. Although, this algorithm is heuristic (in that we cannot prove its correctness), our numerous experiments demonstrate that it leads to good results.

At the second step we count the values for the cost function E (see (1)) for all the parameters from the chosen interval. E is a function of the dimension of the embedding space, and we assume that the embedding dimension is known a priori. This dimension can be estimated using one of the numerous methods for the intrinsic dimension estimation (Kégl, 2002; Costa and Hero, 2004; Camastra and Vinciarelli, 2002). Note that in practice in all our experiments the best low-dimensional embedding always has the same dimension as the embedding with K_{min}.

Eq. (3) was used in (Kouropteva et al., 2002) to measure how well the high-dimensional structure is represented in the embedded space. Tenenbaum used the residual variance in the original Isomap paper to determine the dimension of the embedding space ("elbow" technique). The idea of the minima search in this equation is very similar to the Tenenbaum's idea about the "elbow".

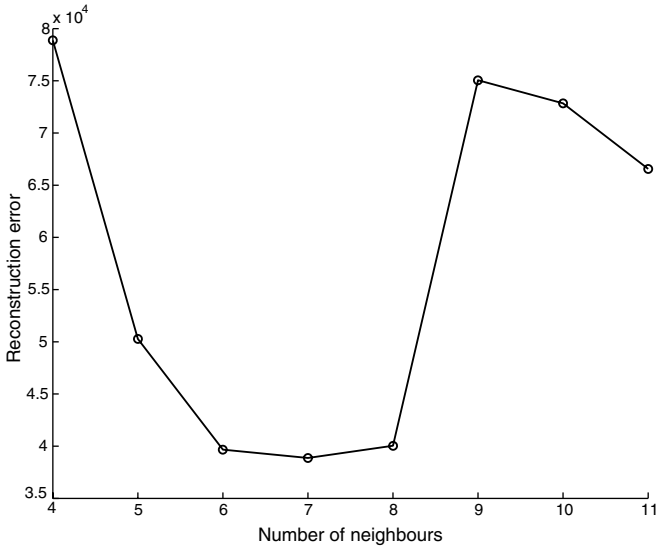


Fig. 1. Cost function for sculpture data set.

4. Experimental results

In this section we present the results of experiments with five different data sets. Here we evaluate how the proposed algorithm processes different data types, artificial and real. The first considered data set in Section 4.1 consists of artificial images of a face rendered with different poses and lighting directions which can be described by a limited number of features. In Section 4.3 we show how our algorithm works with a real data of articulated human motion.

Fig. 2. Two-dimensional Isomap representation of sculpture data set.

```
for angle:=3/2π to 6/2π
    for height:=0 to 21
        x:=angle·cos(angle)
        y:=height
        z:=angle·sin(angle)
    end
end
```

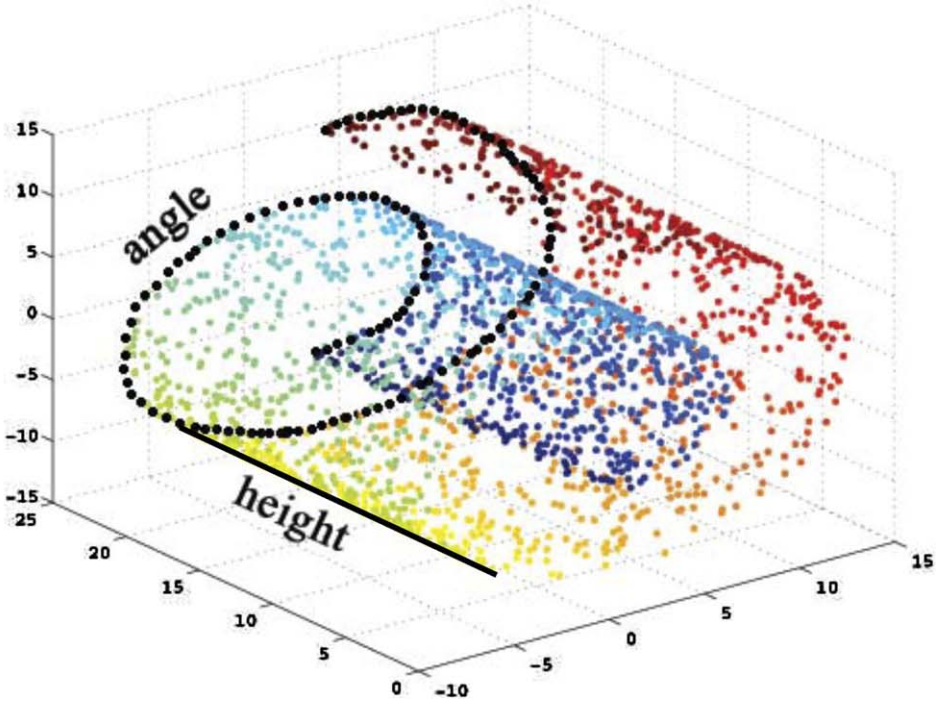


Fig. 3. Swissroll definition.

Section 4.2 utilise a synthetic two-dimensional manifold—Swissroll, lying in a three-dimensional space. In this data set the underlying structure is known and thus the quality of the embeddings, obtained by Isomap with different parameters and by PCA and LLE, can be measured by Pearson's correlation coefficient and Spearman's ρ of their resulting lower-dimensional coordinates with the actual coordinates. Section 4.4 utilises the Olivetti face database in which the faces vary slightly in pose as well as scale. Section 4.5 uses a set of handwritten digits from the MNIST database. Olivetti and MNIST databases were used for the classification experiments.

4.1. Sculpture face data set

First, we tried to select the parameter for the sculpture face data set automatically. This data set was used by Tenenbaum et al. (2000) in the original Isomap paper. The sculpture face data set consists of 698 gray-scale images, 64×64 pixels each. We represent each image by a raster scan vector and form Isomap input space X.

Following the first step of our algorithm, we choose the interval for K_{opt}, which according to our algorithm is Costa and Hero (2004) and Karhunen and Joutsensalo (1994). Then we reduce the number of possible optimal parameter values. In Fig. 1, one can see the plot of the cost function $E(K)$. The optimal value of K determined by our method for these data is 7.

Tenenbaum et al. (2000) used $K = 6$. Indeed, the cost function values for $K = 6$ and $K = 7$ are sufficiently close, as well as the residual variance values at these points. On the other hand, if we calculate the Spearman ρ between the input and output data with $K = 6$ and $K = 7$, then we get $\rho_{K=6} = 0.557$ and $\rho_{K=7} = 0.5645$. This means that the magnitude of the correlation between the input and output data is greater with $K = 7$.

Having applied Isomap to the sculpture face set, we obtained a three-dimensional embedding of initial data. Each dimension of embedding represents one degree of freedom of the underlying original data: left–right pose, up–down pose, and lighting direction. Fig. 2 illustrates the two-dimensional projection of the sculpture data set embedding, obtained by running Isomap with optimal parameter value of $K = 7$.

4.2. Swissroll

Next, we proceed applying different methods to the Swissroll data set, which is a synthetic example of a nonlinear manifold. The Swissroll was used by Tenenbaum et al.

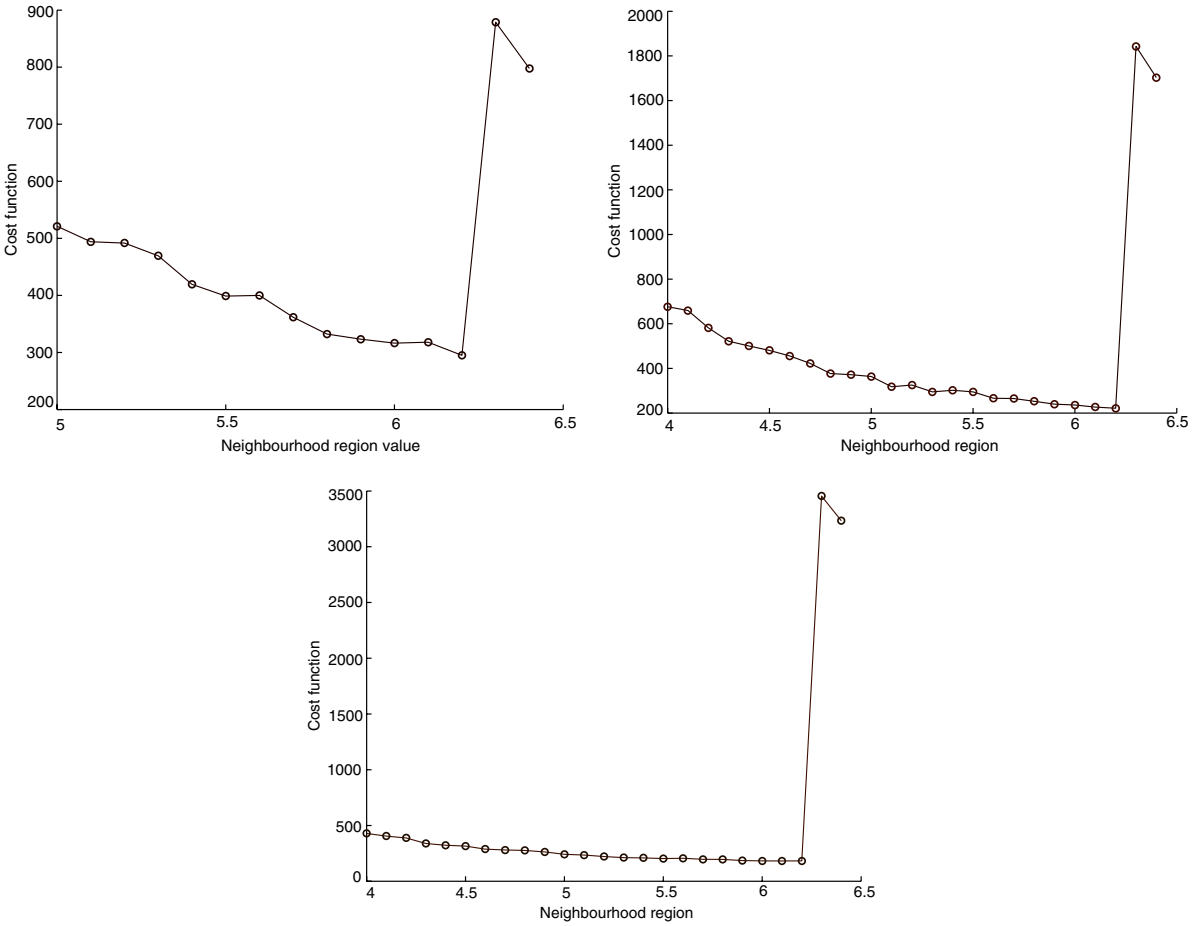


Fig. 4. Swissroll cost functions at 200, 1000 and 2000 points.

(2000) in the original Isomap paper to illustrate how the method works. The generating function is based on Roweis and Saul; Fig. 3 gives the definition of the Swissroll surface and shows its shape as well as 2000 points randomly sampled from the manifold.

We applied our algorithm to numerous Swissroll data sets, each containing from 200 to 2500 randomly chosen points. For each set we ran 50–100 experiments with different values of Isomap parameter ϵ. The cost function plot here is similar to the one for the Swissroll data sets with a different number of points (see Fig. 4), with minima in the same point $\epsilon = 6.2$. This ϵ value was chosen as optimal for all the Swissroll data sets.

To demonstrate the optimality of the chosen value of ϵ we calculate the following measures: residual variance, correlation coefficients and Spearman's ρ with angle and height compared to the discovered two-dimensional coordinates. First, we calculate these measures for Isomap with different parameter values, $\epsilon = 4$, 5, 6.2. For each parameter value we use Swissroll data sets which contain from 200 to 2500 points. Our experiments confirm that with small data sets (less than 700 points) the difference of measure values for different ϵ is quite large, whereas with large data sets this difference is small. Fig. 5 shows the residual variance and correlation coefficients of two-dimensional embeddings with the angle and the height of Swissroll. The plot of Spearman's ρ is similar to the correlation coefficient plot, so we do not include it.

At the same time, the cost function in the interval $\epsilon \in [4, 6.2]$ tends to get smoother as the number of points increases (see Fig. 4). So, we can make the conclusion: when we have a small number of points (less than 700), we will only get the best result with $\epsilon = 6.2$. But when number of points is growing, we can take ϵ from 4 to 6.2 without any significant loss in the quality of result. For example, Fig. 6 shows that the Isomap embeddings for $\epsilon = 4$ and $\epsilon = 6.2$ (the first two plots) look quite similar and good in contrast to the mapping obtained with $\epsilon = 6.3$ (the last plot).

Next, we compare Isomap (with selected optimal parameter) with other techniques: PCA and LLE. We chose LLE neighbourhood parameter value $K = 8$ because with this K we got the best result with LLE. Analogously, we measure residual variance, correlation coefficients and Spearman's ρ

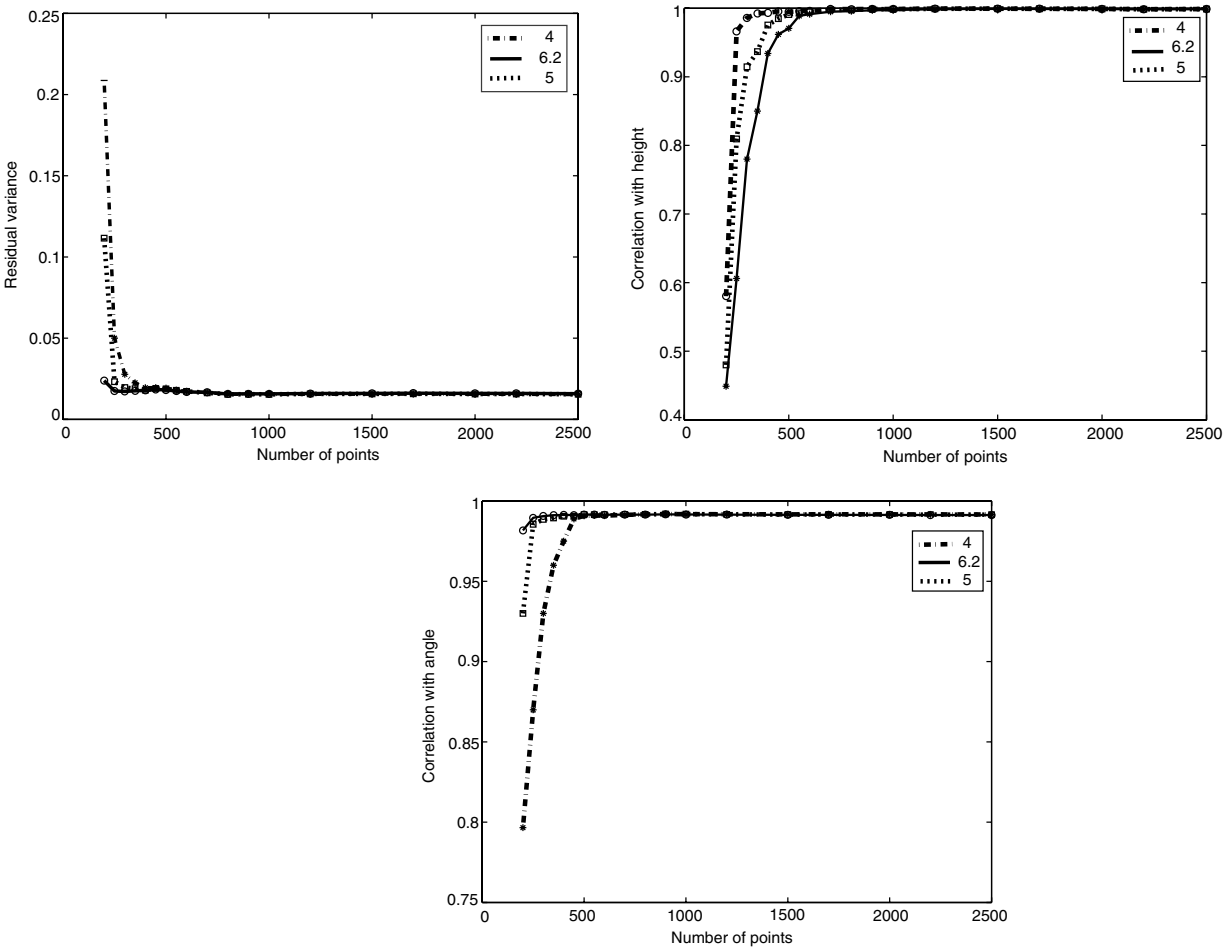


Fig. 5. Residual variance and correlation coefficients with height and angle (left to right). At each plot these values are shown for $\epsilon = 4$, 5, 6.2.

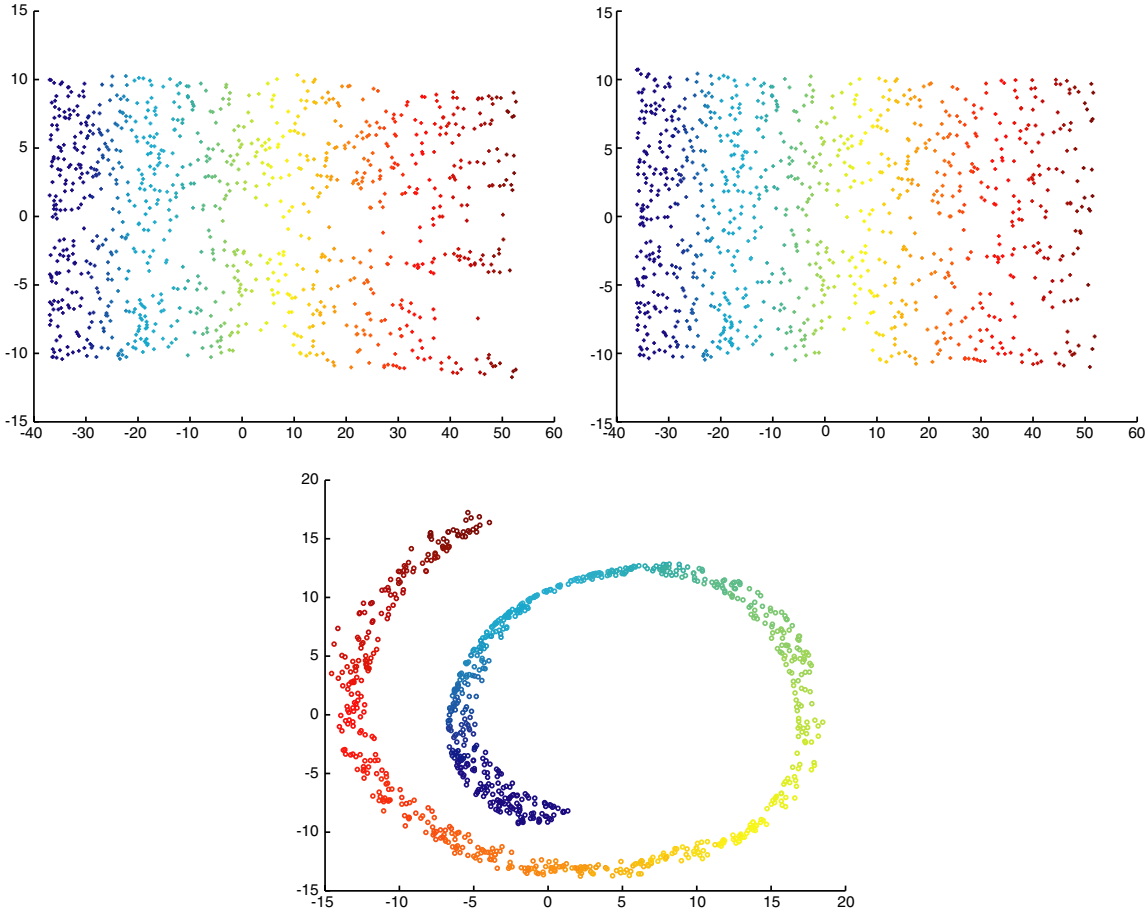


Fig. 6. Two-dimensional Isomap representation of Swissroll. Left to right: $\epsilon = 4$, $\epsilon = 6.2$ (the optimal), $\epsilon = 6.3$, 1000 points.

with the known angle and height of Swissroll. Our results show that Isomap performs much better than PCA and LLE, Fig. 7.

Finally, proceeding with our experiments we added zero-mean, normally distributed noise to the coordinates of the each Swissroll point. We consider Swissroll data sets consisting from 200 to 2500 randomly chosen points, and run Isomap 50–100 times. However, finding the optimal parameter for data with noise is a difficult task—at a given noise level our experiments failed to find a common optimal parameter value when the standard deviation of the noise was chosen to be more than 5% of smallest dimension of the bounding box enclosing the data. The problem is in the choice of the interval of the possible values—the lower bound of the interval is already too big (“short-circuit” problem). Fig. 8 shows the cost function for Swissroll with 4% and 5% noise added. Here we did 70 experiments for each noise level with Swissroll data sets containing 2000 points with zero-mean normally distributed noise added to the coordinates of each point, where the standard deviation of the noise was chosen to be 4% and 5%, respectively, of smallest dimension of the bounding box enclosing the data. The optimal parameter values selected by our method are $\epsilon = 3.3$ for 4% noise, and $\epsilon = 2.7$ for 5% noise. Comparing the last picture from Fig. 4 with the pictures from Fig. 8 one can see that the optimal parameter value decreases as the noise increases.

4.3. Human motion

This data represents motion of a walking person, consisting of two steps, right turn and one step. The initial feature parameters represent the coordinates of a human (arms, legs, torso) in the 3D space. Each pose is characterised by 17 points, and we have a set of 51 dimensional vectors as an input. The full description of how the data was captured can be found in (Karaulova et al., 2000).

First, we find the interval of possible neighbourhood values for the data set $K \in [3, 10]$. Fig. 9 shows the cost function plot. According to the algorithm, the optimal value for the neighbourhood parameter is 4.

Fig. 10 illustrates the Isomap space for the motion data. Hands and legs move in coordination with each other, i.e. right leg up correspond to left hand up during the step. Therefore the dimension of the feature space is limited and was determined by Isomap as 3. As one can see from the pictures, the first dimension corresponds to the human moving to the camera and away from the camera. The second dimension corresponds to the phase of the movement, it represents the step phases starting from when the left

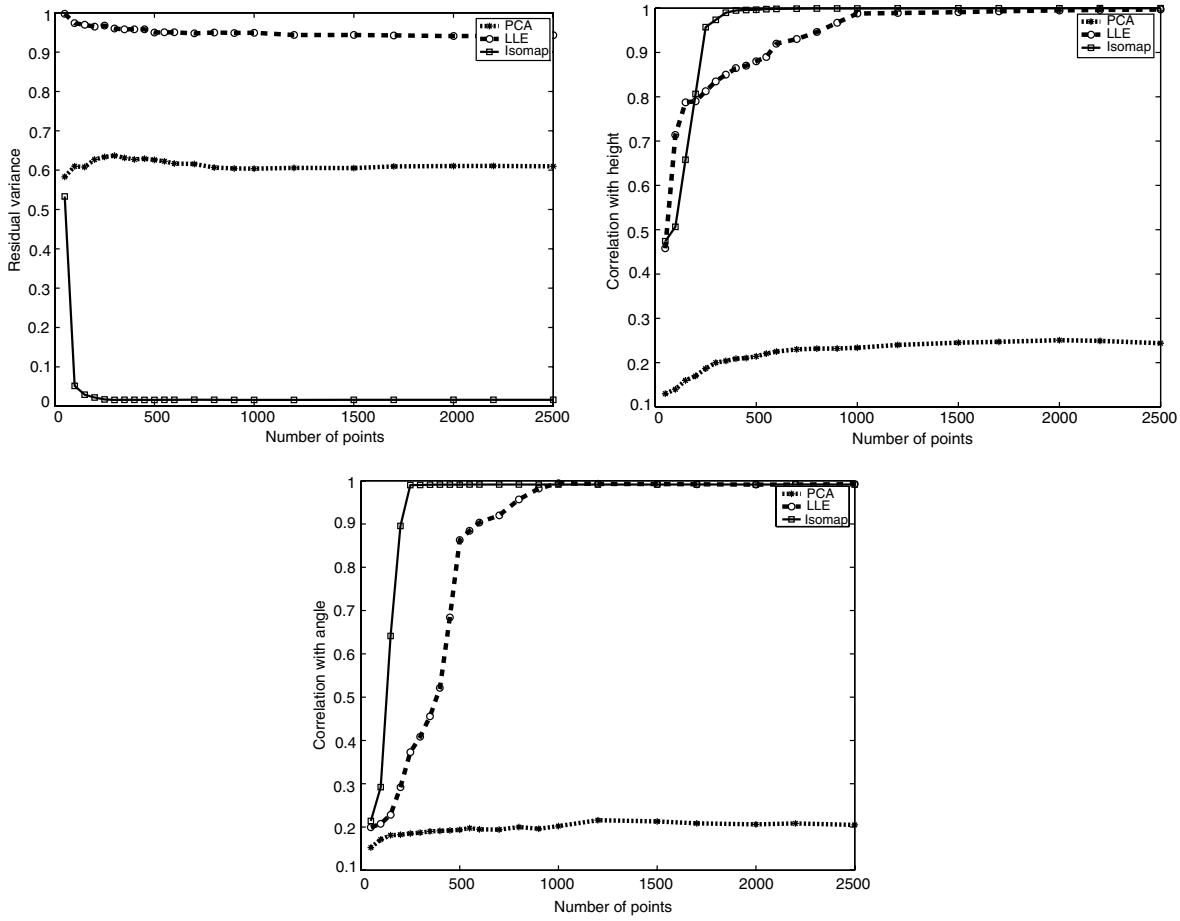


Fig. 7. Residual variance and correlation coefficients with height and angle (left to right) with PCA, LLE ($K = 8$) and Isomap ($\epsilon = 6.2$).

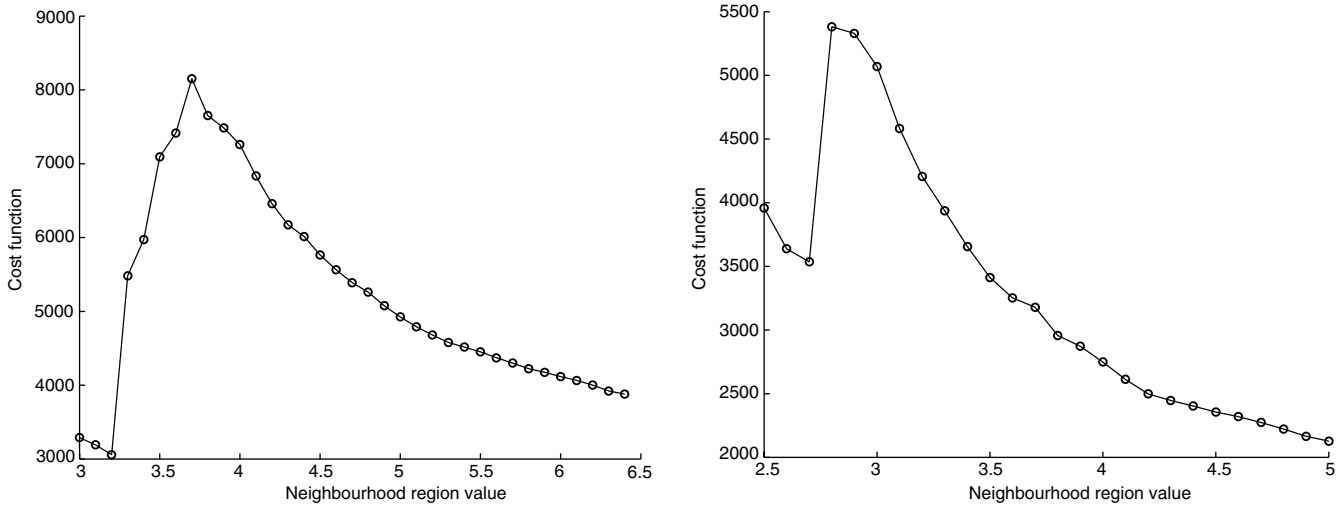


Fig. 8. Cost functions for Swissroll with 4% and 5% noise levels, 2000 points.

hand is bending and the right leg becomes straight, and up to when the right hand is bending and the left leg becomes straight. The third dimension represents the location of the human around the camera (closer-farther).

4.4. Classification experiments: Olivetti face database

The Olivetti face database is a widely used set of images, consisting of 400 images of 40 subjects, which can be

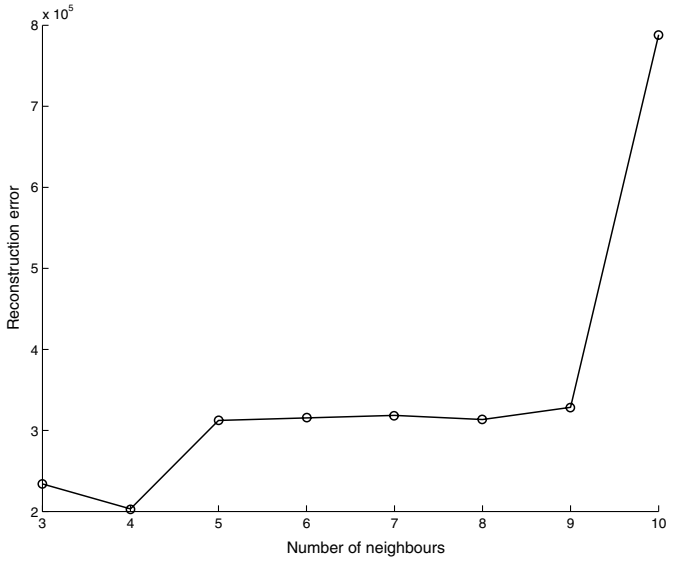


Fig. 9. Cost function for the human motion data.

downloaded from http://www.uk.research.att.com/face-database.html. All the images were taken against a dark homogeneous background with the subjects in an up-right, frontal position, with tolerance for some side movements. There are variations in facial expression (open/closed eyes, smiling/nonsmiling), and facial details (glasses/no glasses). The images are greyscale with a resolution of 92×112.

The basic Isomap method is not optimal from the classification viewpoint, so we use its extended version, discriminant isometric mapping (Yang, 2002, 2003). Applying our algorithm to find the optimal parameter for discriminant isometric mapping, we obtain the interval of parameter K values: (de Ridder et al., 2003; Donoho and Grimes, 2003), and choose $K_{\text{opt}} = 6$ because the cost function reaches its minimum at this point. Fig. 11 shows representation of the Olivetti data set in the first two dimensions. It can be seen that the illumination varies along the horizontal axis, and person's facial features change (from men with beard and without hair in the top to girl with long hair without beard in the bottom) are on the vertical axis.

We used the "leave-one-out" approach in our classification experiments. In this approach each image is left out from the training set of n images and the projection matrix is computed. Then all n images in the training set are projected into the embedded space and recognition of the left out image is performed using a nearest neighbour classifier in the embedded space.

To verify the classification results we calculated the error rate and its standard deviation for each method. To calculate the deviation of the error rate we divided all images into 10 groups and find the mean errors for the each group. Then we count the deviation over the mean error. The experimental results of classification are shown in Fig. 12. From this figure one can see that for each method there was no statistically significant difference between the error rates. Among all the methods, discriminant isometric mapping with optimal parameter value achieves the lowest error rate. The similar results were obtained by Yang (2003).

Overall, the extended Isomap with the optimal parameter outperformed PCA and LLE. It is resulted in a more accurate classification due to a better embedding.

4.5. Classification experiments: handwritten digits from the MNIST database

The MNIST database of handwritten digits was used for the second classification experiment (http://yann.lecun.com/exdb/mnist/index.html). It consists of 10 different classes of images from '0' to '9', 28×28 pixels each. In our experiments we used MNIST database subset, consisting of the 2000 images. The optimal neighbourhood region value for the extended Isomap, which was obtained using our algorithm is $\epsilon = 9.6$. The embedded space formed by the two bottom dimensions is presented in Fig. 13. Here the horizontal axis represents the digit shape, and the vertical axis—bottom loop.

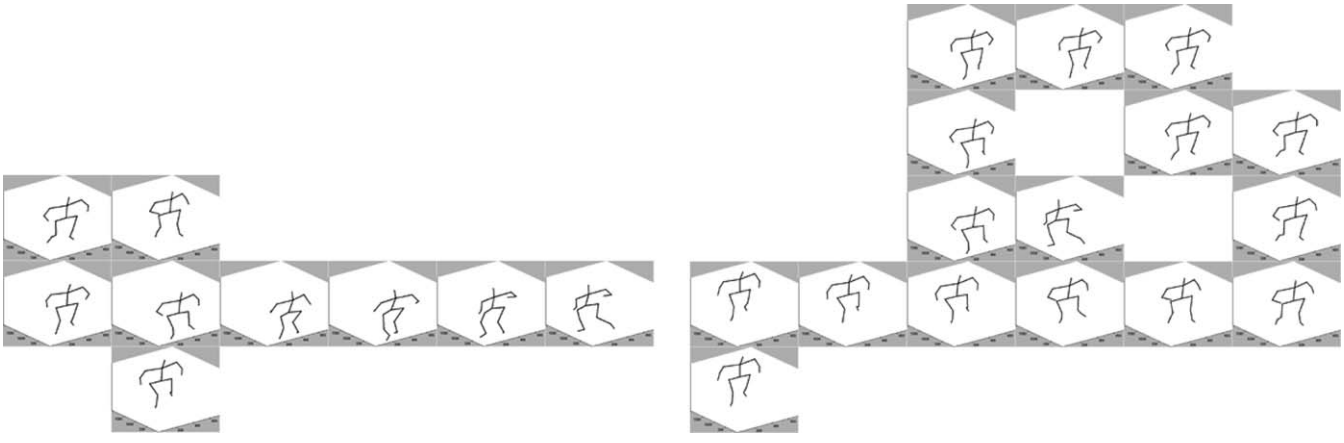
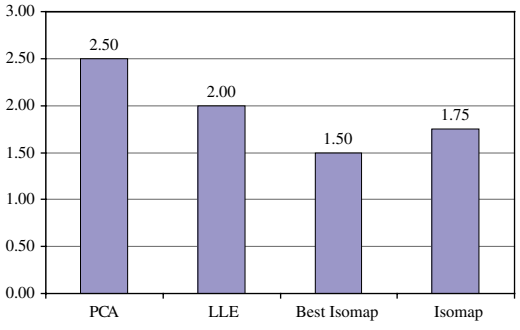


Fig. 10. Human motion data mapped into the embedding space described by first and second coordinates (right) and second and third coordinates (left).

Fig. 11. Two-dimensional discriminant isometric mapping ($K = 6$) of Olivetti data set.



Method	Error rate (%) mean	Std
PCA	2.5	0.0118
best LLE, $K = 14$	2	0.0105
best Ext. Isomap, $K = 6$	1.5	0.0129
Ext. Isomap, $K = 7$	1.75	0.0121

Fig. 12. Classification results with the Olivetti database: mean error and standard deviation.

Using the same approach as in the previous section, classification was performed on this data. The results of classification experiments are shown in Fig. 14. It is easily seen that the extended Isomap with the optimal parameter

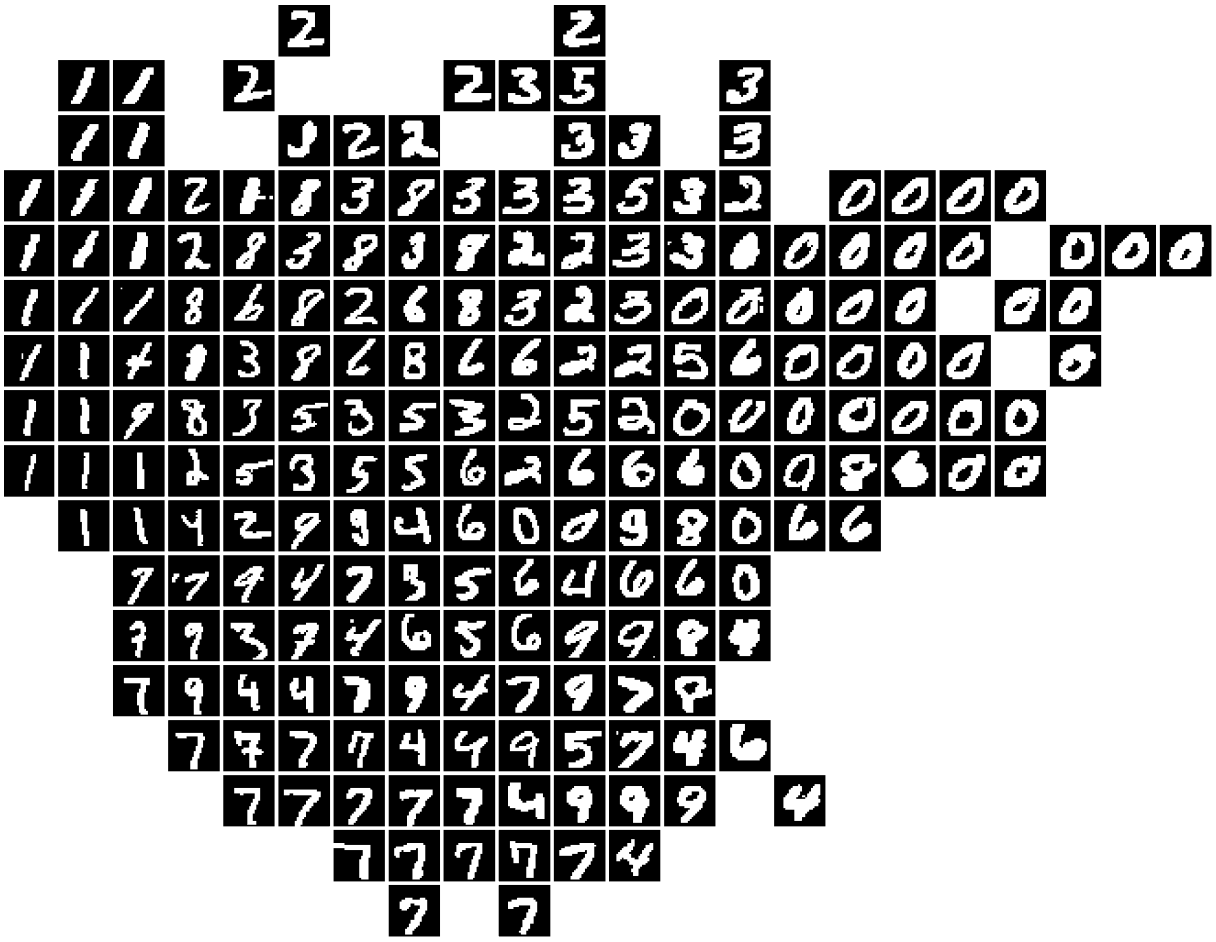
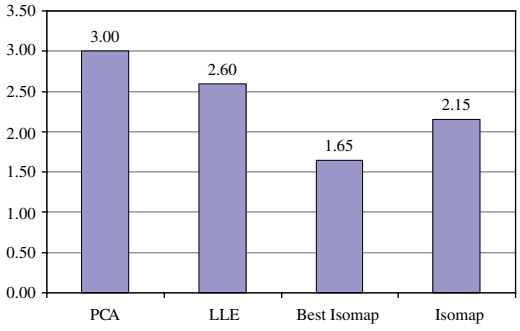


Fig. 13. Two-dimensional discriminant isometric mapping ($\epsilon = 9.6$) of MNIST data subset.



Method	Error rate (%) mean	Std
PCA	3	0.0033
best LLE, $K = 18$	2.6	0.0021
best Ext. Isomap, $\epsilon = 9.6$	1.65	0.0032
Ext. Isomap, $\epsilon = 10$	2.15	0.0034

Fig. 14. Classification results with the MNIST database: mean error and standard deviation.

outperformed PCA and LLE on this set. Note, that with this data set Isomap was less sensitive to the chosen ϵ parameter, comparing this experiment with using other data sets.

5. Conclusions

We have presented a method for automatic selection of optimal parameter value for Isomap. Our idea is based on minimising the Isomap cost function estimation in a carefully chosen parameter interval.

We tested the algorithm on various data sets. Results obtained from our experiments suggest a number of conclusions:

- Our method can be used for a wide class of input data, both real and synthetic.
- When a data set is too sparse (for example Swissroll built on 50 points), it is difficult to choose the optimal parameter. With $K = K_{\min}$ ($\epsilon = \epsilon_{\min}$), i.e. when the neighbourhood graph becomes connected, the graph already gets shortcuts and we obtain $K_{\max} \leqslant K_{\min}$

($\epsilon_{max} \leqslant \epsilon_{min}$). In this case it will be better to choose another dimensionality reduction technique, for example hessian LLE.
- Isomap performs better than PCA and LLE for nonlinear data sets.
- The optimal parameter for modified Isomap, discriminant isometric mapping, can be chosen using our method, in the same manner as for Isomap.

In future work we hope to extend our method to choose the optimal parameter values for other parametric methods for dimensionality reduction of nonlinear data, like Hessian LLE. We also plan to improve our algorithm by reducing the interval of possible parameter values.

Acknowledgment

We would like to thank Dr. Julia Hicks at the Cardiff University for the Human motion Database.

References

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15 (6), 1373–1396.

Brand, M., 2004. From subspace to submanifold methods. Technical Report TR2004-134, MERL Research Lab.

Camastra, F., Vinciarelli, A., 2002. Estimating the intrinsic dimension of data with a fractal-based method. IEEE Trans. Pattern Anal. Mach. Intell. 24 (10), 1404–1407.

Costa, J.A., Hero, A.O., 2004. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. IEEE Trans. Signal Process. 25 (8), 2210–2221.

de Ridder, D., Duin, R.P.W., 2002. Locally linear embedding for classification. Technical Report PH-2002-01, Pattern Recognition Group, Imaging Science and Technology Department, Faculty of Applied Science, Delft University of Technology.

de Ridder, D., Kouropteva, O., Okun, O., Pietikainen, M., Duin, R.P.W., 2003. Supervised locally linear embedding. In: Proc. ICANN/ICONIP 2003, Lecture Notes in Computer Science, 2714. Springer, pp. 333–341.

Donoho, D.L., Grimes, C.E., 2003. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. In: Proceedings of the National Academy of Arts and Sciences, vol. 100, pp. 5591–5596.

Friedrich, T., 2002. Nonlinear dimensionality reduction with locally linear embedding and isomap. Master's thesis, Department of Computer Science, The University of Sheffield, September.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer Verlag, New York.

Karaulova, J., Hall, P.M., Marshall, A.D., 2000. A hierarchical model for tracking people with a single video camera. In: Proc. British Machine Vision Conference (BMVC'2000), vol. 1, pp. 352–361.

Karhunen, J., Joutsensalo, J., 1994. Representation and separation of signals using nonlinear PCA type learning. Neural Networks 7, 113–127.

Kégl, B., 2002. Intrinsic dimension estimation using packing numbers. In: Neural Information Processing Systems, vol. 15.

Kouropteva, O., Okun, O., Pietikäinen, M., 2002. Selection of the optimal parameter value for the locally linear embedding algorithm. In: Proc. of the 1st International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'02), pp. 359–363.

Kouropteva, O., Okun, O., Pietikainen, M., 2003. Classification of handwritten digits using supervised locally linear embedding algorithm and support vector machine. In: Proc. of the 11th European Symposium on Artificial Neural Networks (ESANN'2003), pp. 229–234.

Navarro, D.J., Lee, M.D., 2001. Spatial visualisation of document similarity. In: Proceedings of the Defence Human Factors Special Interest Group Meeting, pp. 39–44.

Roweis, S.T., Saul, L.K. Locally linear embedding homepage. Available from: <http://www.cs.toronto.edu/~roweis/lle>.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290 (December), 2323–2326.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290 (December), 2319–2323.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2002. The isomap algorithm and topological stability—response. Science 295 (5552), 7.

Torgerson, W.S., 1952. Multidimensional scaling: Theory and method. Psychometrika 17, 401–419.

Yang, M.H., 2002. Extended isomap for pattern classification. In: Proceedings of the 18th National Conference on Artificial Intelligence (AAAI 2002), pp. 224–229.

Yang, M.H., 2003. Discriminant isometric mapping for face recognition. In: Computer Vision Systems, Third International Conference, ICVS 2003, vol. 2626, pp. 470–480.