

# Implementation of Manifold Learning Algorithm Isometric Mapping

Huan Yang, Haiming Li\*

College of Science, Shanghai University of Electric Power, Shanghai, China

Email: \*zjxulhm@163.com

**How to cite this paper:** Yang, H. and Li, H.M. (2019) Implementation of Manifold Learning Algorithm Isometric Mapping. *Journal of Computer and Communications*, 7, 11-19.

<https://doi.org/10.4236/jcc.2019.712002>

**Received:** September 26, 2019

**Accepted:** December 6, 2019

**Published:** December 9, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In dealing with high-dimensional data, such as the global climate model, facial data analysis, human gene distribution and so on, the problem of dimensionality reduction is often encountered, that is, to find the low dimensional structure hidden in high-dimensional data. Nonlinear dimensionality reduction facilitates the discovery of the intrinsic structure and relevance of the data and can make the high-dimensional data visible in the low dimension. The isometric mapping algorithm (Isomap) is an important algorithm for nonlinear dimensionality reduction, which originates from the traditional dimensionality reduction algorithm MDS. The MDS algorithm is based on maintaining the distance between the samples in the original space and the distance between the samples in the lower dimensional space; the distance used here is Euclidean distance, and the Isomap algorithm discards the Euclidean distance, and calculates the shortest path between samples by Floyd algorithm to approximate the geodesic distance along the manifold surface. Compared with the previous nonlinear dimensionality reduction algorithm, the Isomap algorithm can effectively compute a global optimal solution, and it can ensure that the data manifold converges to the real structure asymptotically.

## Keywords

Manifold, Nonlinear Dimensionality Reduction, Isomap Algorithm, MDS Algorithm

---

## 1. Introduction

In the process of analyzing high-dimensional data, it faces the problem of “dimensionality disaster” [1]. Dimensionality reduction can reduce the complexity and space complexity of time, save the overhead of extracting unnecessary features, and remove the noise mixed in the data sets, and the simpler model has

stronger robustness on small data sets. The purpose of dimension reduction is to visualize the data, observe and study the data, and improve the efficiency of machine learning. Therefore, the problem of dimensionality reduction has received extensive attention in many fields such as pattern recognition, machine learning and computer vision; with the increasing number of high-dimensional data, the problem of dimensionality reduction has become a research hotspot. Manifold has caught everyone's attention. Dimensionality reduction in the field of machine learning and statistics refers to reducing the number of random variables that need to be considered.

The traditional dimensionality reduction techniques are divided into two types: linear methods and nonlinear methods. The nonlinear methods are divided into preserving local features and retaining global features. Retaining local features is based on reconstruction weights, adjacency graphs, and cut-based spaces. The retention of global features is based on retention distance, kernel-based, and neural network-based. Based on distance preservation, it is divided into multidimensional scaling (MDS) and isometric mapping (Isomap) based on geodesic distance.

The isometric mapping algorithm is a classical algorithm in manifold learning. The goal of manifold learning is to find low-dimensional structures embedded in high-dimensional data spaces and give an efficient low-dimensional representation. Because the manifold learning algorithm can utilize the local geometry of the dataset to reveal its intrinsic manifold structure, it can achieve efficient dimensionality reduction. In addition to the Isomap algorithm, well-known manifold learning algorithms include local linear embedding, Laplacian feature mapping, and local hold projection. These algorithms can keep the topology of the original data unchanged, and can better solve the "dimension disaster" problem in the data processing. This paper will introduce the Isomap algorithm and compare it with the MDS algorithm to compare the dimensionality reduction effects of the two algorithms. The content of this paper is as follows: The second chapter focuses on the Isomap algorithm and the principle of the MDS algorithm. The third chapter is the experimental comparison verification. The fourth chapter summarizes this article and its outlook for the future.

## **2. Isomap Algorithm**

### **2.1. Basic Principles of Isomap Algorithm**

The basic principle of the dimensionality reduction algorithm is to analyze and process high-dimensional data to find meaningful low-dimensional structures hidden in high-dimensional data. The main idea of the Isomap algorithm is to calculate the geodesic distance between data points by local neighborhood distance approximation, and complete the data dimensionality reduction by establishing the equivalence relationship between the geodesic distance of the original data and the distance between the data after dimension reduction. The Isomap algorithm is derived from the linear dimensionality reduction algorithm MDS, which has the main features of the MDS algorithm, namely the validity of the

calculation, the global optimization and the asymptotic convergence. At the same time, it can be more flexible to learn the nonlinear structure of data [2]. The key point of the Isomap algorithm is to replace the traditional Euclidean distance with the geodesic distance, so as to better find the internal structure of high-dimensional data. The geodesic distance [3] in Isomap is represented by the shortest path in the nearest neighbor graph. When calculating the geodesic distance, the distance between a sample and a point close to the sample is calculated by the Euclidean distance, and the distance between a sample and a point that is far from the sample is calculated by the shortest path calculated by Floyd algorithm to approximate the geodesic distance. This method can effectively express data in high-dimensional space in low-dimensional space and reduce the loss of data information after dimension reduction.

## 2.2. Isomap Algorithm Steps

Since the Isomap algorithm is based on the multidimensional scaling analysis MDS algorithm to improve the dimensionality reduction, before introducing the Isomap algorithm, it is very necessary to introduce the MDS algorithm, and understand the MDS algorithm can also understand the Isomap algorithm.

The MDS algorithm is a very traditional dimensionality reduction method. It is based on distance. The goal is to keep the distance between sample points in the original space and the distance between sample points in the low-dimensional space after dimension reduction equal [4]. Suppose there are  $m$  samples

$X = [x_1, x_2, \dots, x_m]$  in the original space, the distance matrix of these samples in the original space is  $D \in R^{m \times m}$ , the  $i$ -th row and  $j$ -th column elements  $dist_{ij}$  are the distances from the sample  $x_i$  to  $x_j$ , and the distance matrix  $D$  is expressed as

$$D = \begin{bmatrix} dist_{11} & \cdots & dist_{1m} \\ \vdots & \ddots & \vdots \\ dist_{m1} & \cdots & dist_{mm} \end{bmatrix} \quad (1)$$

The goal of the MDS algorithm is to obtain the representation of the  $m$  samples in the low-dimensional space  $Z = [z_1, z_2, \dots, z_m]$ , where  $Z$  is the corresponding sample point after the original sample point projection. MDS requires maintaining the Euclidean distance between sample points in the original space in low-dimensional space, so the Euclidean distance of any two samples in  $Z$  in low-dimensional space is equal to its distance in the original space, which is

$$\|z_i - z_j\| = dist_{ij} \quad (2)$$

Let the inner product matrix of the dimension-reduced samples be denoted by  $B$  and  $B = Z^T Z$ , for each element in matrix  $B$   $b_{ij} = z_i^T z_j$ , then matrix  $B$  can be expressed as

$$B = \begin{bmatrix} z_1^T z_1 & \cdots & z_1^T z_m \\ \vdots & \ddots & \vdots \\ z_m^T z_1 & \cdots & z_m^T z_m \end{bmatrix} = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mm} \end{bmatrix} \quad (3)$$

To square the two sides of the Equation (2), you can get

$$dist_{ij}^2 = \|z_i\|^2 + \|z_j\|^2 - 2z_i^T z_j = b_{ii} + b_{jj} - 2b_{ij} \tag{4}$$

Let the sample  $Z$  after dimension reduction be centered, which is  $\sum_{i=1}^m z_i = 0$ . Then the sum of the row and column of matrix  $B$  is 0, which is  $\sum_{i=1}^m b_{ij} = 0$ ,  $\sum_{j=1}^m b_{ij} = 0$ . Then, add different types of summation symbols to the left and right sides of Equation (4), and then simplify and merge the similar items to obtain

$$\begin{aligned} \sum_{i=1}^m dist_{ij}^2 &= \sum_{i=1}^m b_{ii} + \sum_{i=1}^m b_{jj} - \sum_{i=1}^m 2b_{ij} \\ &= tr(B) + mb_{jj} - 0 = tr(B) + mb_{jj} \end{aligned} \tag{5}$$

$$\begin{aligned} \sum_{j=1}^m dist_{ij}^2 &= \sum_{j=1}^m b_{ii} + \sum_{j=1}^m b_{jj} - \sum_{j=1}^m 2b_{ij} \\ &= mb_{ii} + tr(B) - 0 = mb_{ii} + tr(B) \end{aligned} \tag{6}$$

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^m dist_{ij}^2 &= \sum_{i=1}^m (mb_{ii} + tr(B)) = m \sum_{i=1}^m b_{ii} + mtr(B) \\ &= mtr(B) + mtr(B) = 2mtr(B) \end{aligned} \tag{7}$$

After the deformation of the formula (5), the formula (6), and the formula (7), it is obtained

$$b_{ii} = \frac{1}{m} \left( \sum_{j=1}^m dist_{ij}^2 - tr(B) \right) \tag{8}$$

$$b_{jj} = \frac{1}{m} \left( \sum_{i=1}^m dist_{ij}^2 - tr(B) \right) \tag{9}$$

$$2tr(B) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m dist_{ij}^2 \tag{10}$$

Then, since the expression with the summation symbol is long and cumbersome, use  $dist_{i.}^2$ ,  $dist_{.j}^2$  and  $dist_{..}^2$  instead of using the simple symbol and sum, the following expression can be obtained

$$dist_{i.}^2 = \frac{1}{m} \sum_{j=1}^m dist_{ij}^2 \tag{11}$$

$$dist_{.j}^2 = \frac{1}{m} \sum_{i=1}^m dist_{ij}^2 \tag{12}$$

$$dist_{..}^2 = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m dist_{ij}^2 \tag{13}$$

After the Equation (4) is deformed, the Equation (8) and Equation (9) are substituted into the transformed equation, and the Equations (11)-(13) are substituted into the Equation (8) and Equation (9). After the substitution, the formula obtained can be obtained

$$b_{ij} = -\frac{1}{2} \left( dist_{ij}^2 - dist_{i.}^2 - dist_{.j}^2 + dist_{..}^2 \right) \tag{14}$$

In this way, each element in the inner product matrix  $B$  can be calculated by the Euclidean distance matrix  $D$ , and each element in  $B$  is obtained. Then, the inner product matrix  $B$  is obtained. Since  $B = Z^T Z$ , eigenvalue decomposition

is performed on matrix  $B$ , and  $BV = \lambda V$  is obtained, then

$$B = Z^T Z = V \lambda V^T \quad (15)$$

among them,  $\lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}$  ( $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ ),  $V$  is a matrix composed of

eigenvectors corresponding to eigenvalues.

In order to make the dimension reduction effective, only the distance after dimension reduction is as close as possible to the distance in the original space, instead of being strictly equal. Assuming that the dimension to be finally reduced is  $d$ -dimensional, then the largest  $d$  (from large to small) of the eigenvalues constitutes a diagonal matrix  $\lambda_d$ , then the final output of the MDS algorithm, that is, the lower dimension of each sample in the original space is  $Z = \lambda_d^{\frac{1}{2}} V_d^T$ , where  $V_d^T$  is the eigenvector matrix composed of eigenvectors corresponding to  $d$  eigenvalues [5].

The input to the MDS algorithm is the Euclidean distance matrix  $D$ , but the Euclidean distance is not applicable to the manifold. For example, for the two-dimensional manifold of the Earth in three-dimensional space, suppose that the distance between the North Pole and the South Pole is calculated in three-dimensional space, which is the length of the line connecting the two points. However, this calculation is wrong. Because it is impossible to make a hole through the Arctic to the South Pole, it is necessary to walk along the surface of the earth. Of course, it is not acceptable to walk along any line, because there will be many different distances. Therefore, a new measure of the distance defined on the Earth's surface (manifold) is needed. In order to correspond to the European space, here is a general definition of the straight-line distance. In the European space, "between two points, the shortest line segment", the concept of the line segment is generalized to become "the shortest curve between two points is the line segment", this shortest curve is usually called "geodesic line" [6]. Therefore, the Isomap algorithm first obtains the geodesic distance matrix from the Euclidean distance matrix through the Floyd algorithm, and then puts the geodesic distance matrix into the MDS, thereby obtaining the final dimensionality reduction result.

The specific process of converting the Euclidean distance matrix into the geodesic distance matrix is as follows: first, the Euclidean distance matrix is known (the distance between all sample points is calculated by Euclidean distance), and the number of neighborhoods  $k$  is set, for each the sample point, the distance between a sample point and the  $k$  sample points that are closer to the sample point is calculated by the Euclidean distance, and the distance from the rest of the sample point (the point farther away from the sample point) is set to infinity; in the second step, the above matrix is updated to the shortest path matrix by the Floyd algorithm. The geodesic distance between sample points can be approximated by the shortest path, thus converting the Euclidean distance matrix to the

geodesic distance matrix. After the geodesic distance matrix is added, the geodesic distance matrix is placed into the MDS algorithm to obtain the final dimensionality reduction result.

The Isomap algorithm flow is summarized as follows: 1) Calculate the Euclidean distance between each sample point to obtain the Euclidean distance matrix; 2) Set the number of neighborhoods. In the Euclidean distance matrix, except for the neighborhood points, the remaining distances are set to infinity; 3) Update the above matrix to the shortest path matrix by Floyd algorithm; 4) Input the shortest path matrix into the MDS algorithm, and the result is the result of dimension reduction of Isomap algorithm.

### 2.3. Isomap Advantages

- 1) Capable of processing high dimensional data such as nonlinear manifolds;
- 2) Global optimization;
- 3) Whether the input space is highly folded, distorted, or curved, Isomap can still optimize the low-dimensional European representation globally;
- 4) Isomap can guarantee a gradual recovery to the real dimension.

### 2.4. Isomap Disadvantages

- 1) It may be unstable and dependent on the topological space of data;
- 2) When it is guaranteed to gradually recover to the geometry of the nonlinear manifold: when  $N$  is increased, the point provides a distance closer to the geodesic distance, but takes more calculation time; if  $N$  is small, geodesic the distance will be very imprecise.

## 3. Isomap Algorithm vs. MDS Algorithm

Both the Isomap algorithm and the MDS algorithm are dimension reduction algorithms. MDS algorithm is a linear dimension reduction algorithm, which is suitable for European space. Isomap algorithm is a nonlinear dimensionality reduction algorithm, which is suitable for manifolds, that is, high dimensional space.

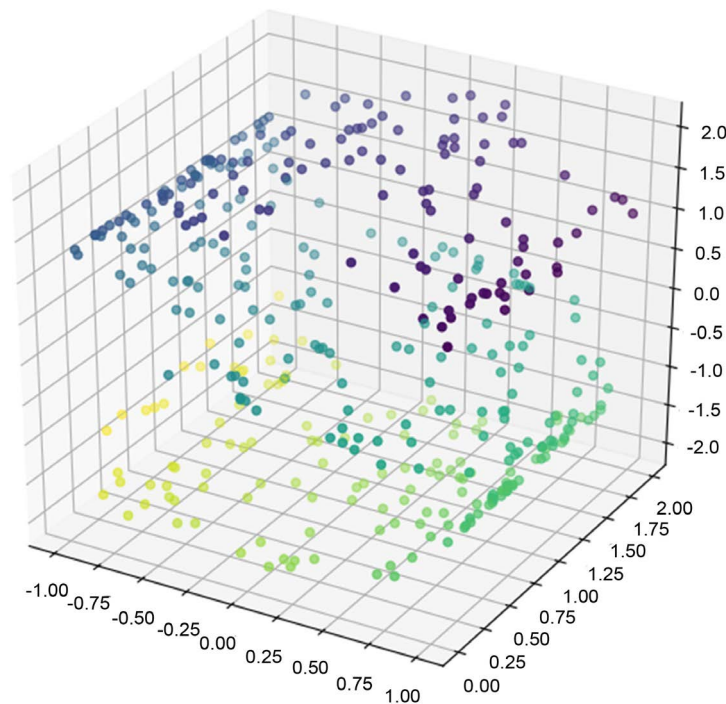
The Isomap algorithm achieves the goal of dimensionality reduction by modifying the algorithm MDS originally applied to the European space. The purpose of the MDS algorithm is to keep the distance between the sample points in the original space and the distance between the sample points in the low-dimensional space after dimensionality reduction. The MDS is designed for the European space with Euclidean distance, but if the data is distributed in a manifold, the Euclidean distance is not applicable, only the geodesic distance can be used. Therefore, the Isomap algorithm replaces the input of the MDS algorithm (Euclidean distance matrix) with the geodesic distance matrix obtained by the shortest path algorithm, thus solving the problem that the Euclidean distance is not applicable to the manifold, which is the biggest difference between Isomap algorithm and MDS algorithm.

In order to more intuitively compare the Isomap algorithm with the MDS al-

gorithm, an S-shaped surface as shown in **Figure 1** is randomly generated. This surface is composed of 400 sample points. The Isomap algorithm (the number of neighbors is set to 15) and the MDS algorithm are used to reduce the dimension and reduce it to 2 dimensions. The dimensionality reduction results are shown in **Figure 2** and **Figure 3**. As can be seen in **Figure 1**, the arrangement of the 400 sample points is not very tight, that is, there is a distance between the sample point and the sample point, but it can be seen from the results obtained by MDS algorithm after dimensionality reduction (**Figure 3**) that the sample points and the sample points are very close together, and some even overlap, which is far from the distance between the real sample points. The result obtained from the dimensional reduction of the Isomap algorithm (**Figure 2**) shows that the distance between the sample points obtained after dimensionality reduction is similar to the distance between the real sample points. Therefore, in the manifold, the Isomap algorithm has a better dimensionality reduction effect.

The above Isomap algorithm sets the neighborhood number  $k$  to 15 to obtain the dimensionality reduction result of **Figure 2**. Now the neighborhood number  $k$  is set to 400. The obtained result is shown in **Figure 4**. It can be seen that the Isomap algorithm and the MDS algorithm are dimensionally reduced. The result is the same. The reason is that  $k$  here takes 400, and the sample points are 400 in total. That is to say, the distance between sample points in the Isomap algorithm is calculated by Euclidean distance, and the geodesic distance is not used. At this point, we can also see the biggest difference between the Isomap algorithm and the MDS algorithm.

From the above experimental results, it can be seen that in the process of



**Figure 1.** S surface.

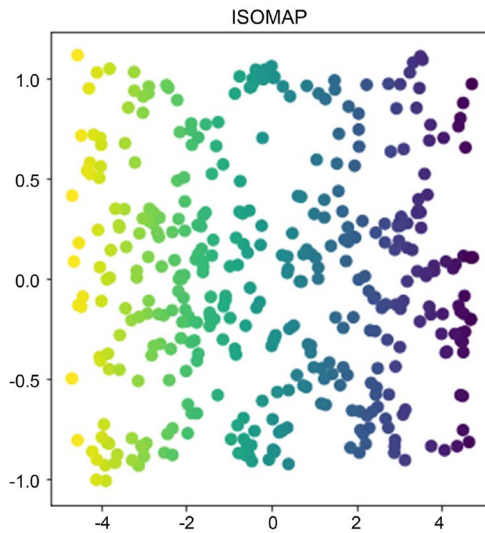


Figure 2. Isomap algorithm dimensionality reduction results.

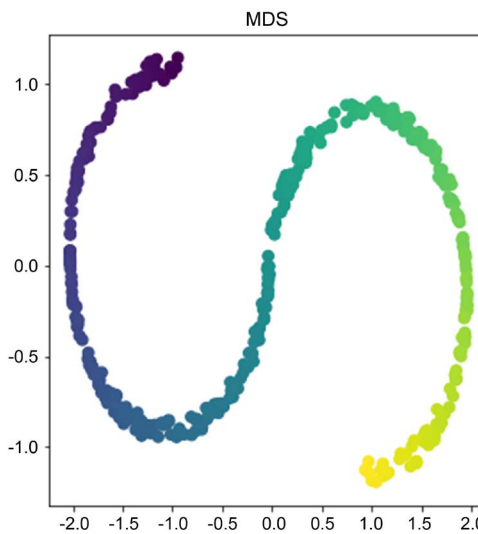


Figure 3. The result of dimensionality reduction with MDS algorithm.

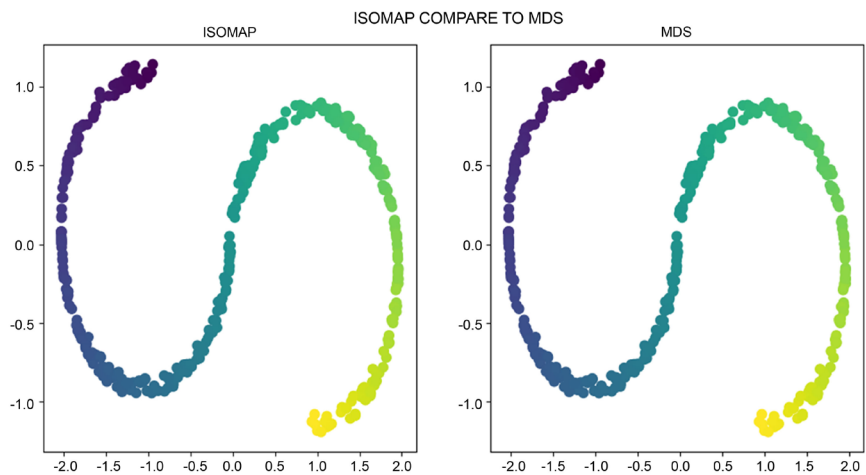


Figure 4. Isomap algorithm (k 400) and MDS algorithm comparison.



dimensionality reduction using the Isomap algorithm, the selection of the neighborhood number  $k$  plays a key role. If the value of  $k$  is too small, the graph will not be connected; and if the value of  $k$  is too large, it will make the Isomap algorithm tend to the MDS algorithm. Therefore, the choice of the number of neighborhoods  $k$  is crucial. For the selection problem of  $k$ , an adaptive method was proposed later [7].

#### 4. Conclusion

The dimensionality disaster caused by high-dimensional data has made the dimensionality reduction widely concerned. The traditional dimensionality reduction algorithm (such as MDS algorithm) applicable to European space has not been applied to high-dimensional space. Manifold learning is a new dimensionality reduction method, its main goal is to effectively discover the low-dimensional manifold structure inherent in high-dimensional data sets and give an effective low-dimensional representation. This paper mainly introduces the dimension reduction algorithm for manifolds, Isomap algorithm, which starts from the perspective of maintaining the global structure. In addition, this paper also compares it with the MDS algorithm through experiments, from the experimental results in addition to the difficulty of selecting the neighborhood number  $k$ ; it is proved that the dimensionality reduction is performed on the manifold. The Isomap algorithm can maintain the topology of the high-dimensional data more than the MDS algorithm; that is, the dimensionality reduction effect is better.

#### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

#### References

- [1] Yang, B., Xiang, M. and Zhang, Y. (2016) Multi-Manifold Discriminant Isomap for Visualization and Classification. *Pattern Recognition*, **55**, 215-230. <https://doi.org/10.1016/j.patcog.2016.02.001>
- [2] Wang, H.Y., Ding, X.J. and Cheng, Q.C. (2013) An Improved ISOMAP for Visualization and Classification of Multiple Manifolds. *International Conference on Neural Information Processing*, Daegu, 3-7 November 2013, 1-12. [https://doi.org/10.1007/978-3-642-42042-9\\_1](https://doi.org/10.1007/978-3-642-42042-9_1)
- [3] Zhao, L.W., Luo, S.W., Zhao, Y.C. and Liu, Y.H. (2005) Study on the Low-Dimensional Embedding and the Embedding Dimensionality of Manifold of High-Dimensional Data. *Journal of Software*, **16**, 1423-1430. <https://doi.org/10.1360/jos161423>
- [4] <https://wenku.baidu.com/view/1da1f06d27d3240c8447efd9.html>
- [5] [https://blog.csdn.net/zhangweiguo\\_717/article/details/69663452](https://blog.csdn.net/zhangweiguo_717/article/details/69663452)
- [6] Jenkins, O.C. (2004) A Spatio-Temporal Extension to Isomap Nonlinear Dimension Reduction. *International Conference on Machine Learning*, Banff, 4-8 July 2004, 56. <https://doi.org/10.1145/1015330.1015357>
- [7] Li, Y.Y., Xiao, D. and Wei, X.F. (2014) Isomap Algorithm for Adaptive Neighborhood Selection. *Journal of Hebei Institute of Architecture and Civil Engineering*, No. 2, 110-112.