The Journal of Supercomputing (2023) 79:18648–18666
https://doi.org/10.1007/s11227-023-05354-5

Isometric projection with reconstruction

Ruisheng Ran[1] · Qianghui Zeng[1] · Xiaopeng Jiang[1] · Bin Fang[2]

Accepted: 26 April 2023 / Published online: 17 May 2023

Abstract

Isometric Projection (IsoP) is a linear dimensionality reduction method, which provides the best linear approximation to the true isometric embedding of data. However, IsoP and all its variants only consider the one-way mapping from high-dimensional space to low-dimensional space. The projected low-dimensional data may not "represent" the original sample accurately and effectively. In this paper, based on the "encoding-decoding" mechanism, a new IsoP method called IsoP-R (Isometric Projection with Reconstruction) has been proposed. In this method, the conventional projection of IsoP is viewed as the encoding stage, and the decoder is used to reconstruct the original high-dimensional data from the projected low-dimensional data. In this way, our algorithm makes the low-dimensional embedding data "represent" the original data more accurately and effectively. Experiment results on Handwritten Alphadigits, COIL-100, Olivetti Research Laboratory and Georgia Tech face datasets show that the proposed IsoP-R approach better represents the data and achieves much higher recognition accuracy.

Keywords Isometric projection · Autoencoder · Dimensionality reduction · Manifold learning

✉ Bin Fang
fb@cqu.edu.cn

Ruisheng Ran
rshran@cqnu.edu.cn

Qianghui Zeng
2021210516092@stu.cqnu.edu.cn

Xiaopeng Jiang
2021210516042@stu.cqnu.edu.cn

[1] The College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China

[2] The College of Computer Science, Chongqing University, Chongqing 400044, China

1 Introduction

Curse of dimensionality [1, 2] was first proposed by mathematician Richard Bellman when he studied dynamic programming problems. In Machine Learning (ML) [3], the curse of dimensionality often refers to that as dimensionality increases, in order to obtain accurate and reliable results in statistics, the amount of data required to support this result usually increases exponentially with the increase in dimension.

Dimensionality reduction (DR) [4, 5] is one of the effective ways to solve the curse of dimensionality. DR methods are generally divided into linear and nonlinear [6]. Linear dimensionality reduction techniques assume that the data structure is linear. It uses a simple linear function to project high-dimensional data to low-dimensional data to obtain low-dimensional features of the data. The representative algorithms of linear dimensionality reduction include Principal Component Analysis (PCA) [7] and Linear Discriminant Analysis (LDA) [8, 9]. Their commonality is that they all assume that the original dataset is a global, linear structure. Therefore, using them on nonlinear data often leads to weak performance.

As a result, some nonlinear methods are proposed for nonlinear data. And the nonlinear methods have two different solutions: kernel-based [10] and manifold-based [11] dimensionality reduction methods. The kernel-based method will project the original data to a higher dimensional space to make it possible to be linearly separable, but the selection of the most critical kernel function is more difficult and can only be judged empirically. In addition to the kernel-based methods, the manifold-based methods are also popular nonlinear dimensionality reduction techniques. And its representative method is Locally Linear Embedding (LLE) [12] and Isometric Mapping (Isomap) [13]. However, the disadvantage of nonlinear methods is that they are only defined on the training set and cannot be mapped on the test set. Therefore, the linearization of the manifold-based nonlinear methods are proposed, such as Locality Preserving Projections (LPP) [14] is the linearization of Laplacian Eigenmap (LE) [15, 16] and Isometric Projection (IsoP) [17] is the linearization of Isomap.

The IsoP algorithm first constructs the nearest neighbor graph of data and then computes the shortest paths for all pairs of data points in the graph. Through this process, an estimate of the global structure of the data is obtained. Then the Multi-dimensional Scaling (MDS) [18, 19] technology is used, the mapping function is required to be linear, and the objective function of IsoP is obtained. IsoP retains the advantages of Isomap while overcoming the disadvantage of Isomap only providing embeddings for training data.

IsoP has many improvement methods that possess better performance. However, in practical applications, such as face recognition, IsoP methods face the small-sample-size (SSS) [20] problem. This is because after the image matrix is stretched to an image vector, the dimensionality of the vector is much larger than the number of samples, resulting in that the feature matrix may be singular, and then affects the performance of IsoP. To solve this problem, PCA or Singular Value Decomposition (SVD) [21, 22] is usually used to preprocess the original data, which avoids the SSS problem but also inherits the shortcomings of PCA [23].

Some researchers have proposed many variants of IsoP, such as Tensor-based Isometric Projection (TIsoP) [24] and Isometric Projection based on Maximal Margin Criterion (IsoP-MMC) [25]. TIsoP uses the two-dimensional image matrix instead of the traditional one-dimensional vector, and performs SVD in the tensor space, thereby avoiding the SSS problem. Other improved methods of IsoP include Orthogonal Isometric Projection (OIsoP) [26] and Uncorrelated Discriminant Isometric Projections (UDIsoP) [27], of which OIsoP can be regarded as an extension of IsoP, and UDIsoP is a feature extraction method based on face recognition. The idea of OIsoP is the same as IsoP but further requires that the projection matrix is orthogonal, and its constraints are different from the orthogonal projection of Cai's projection. According to the regularization method given in Ref. [28], it can be applied to the IsoP method, that is, the Regularized Isometric Projection (RIsoP) is obtained, and the Exponential Isometric Projection (EIsoP) can be obtained from the exponential embedding using matrix exponential given in Ref. [29].

However, the IsoP method and its variants only consider one-way mapping from high-dimensional manifold space to low-dimensional space. This mapping only enables the embedded low-dimensional data points to preserve the intrinsic geometry of the original sample, but it may not "represent" the original sample very accurately and efficiently.

In this work, based on the "encoding-decoding" mechanism, a new IsoP method called IsoP-R (Isometric Projection with Reconstruction) has been proposed. Specifically, under the condition of maintaining the geodesic distance information of the sample, the data points in high-dimensional manifold space are encoded into data points in low-dimensional space by using the conventional IsoP projection model. However, we also consider using the decoder to reconstruct the original high-dimensional data points from the embedded low-dimensional data points. The innovation of IsoP-R lies in:

1. Bi-directional mapping of IsoP-R

IsoP only considers one-way mapping from high-dimensional space to low-dimensional space, and other improved IsoP methods focus on optimizing this one-way mapping from different approaches. The newly proposed IsoP-R method adds a reconstruction phase to the IsoP method by reconstructing the data in the low-dimensional space. The above reconstruction stage enables a more accurate and efficient representation of the original data for low-dimensional embedded data.

2. Improved classification capability of Isometric Projection

Based on the encoder-decoder paradigm, the first stage uses the traditional IsoP model as an encoder, which maps the high-dimensional data to the low-dimensional data. The second stage reconstructs the low-dimensional embedding points into the original space using a linear self-encoder. The classification capability of IsoP is significantly enhanced by the reconstruction process in the second stage.

The rest of this paper proceeds as follows: Sect. 2, we review the Isomap method, IsoP method and Autoencoder. Section 3, we propose the novel IsoP method with the encoder-decoder paradigm. Section 4, we make experiments to evaluate the new methods and finally concludes this paper in the last section.

2 Related works

2.1 Isometric mapping

IsoP is a linear approximation of Isomap, so a brief introduction to Isomap is first presented. Isomap was proposed by Tenenbaum et al. and is a nonlinear dimensionality reduction algorithm. It is a global optimization algorithm based on the MDS. Isomap algorithm replaces the traditional Euclidean distance with geodesic distance between data. Then MDS algorithm is used to map data points from the high-dimensional space to the low-dimensional space to obtain the low-dimensional manifold that keeps the internal structure between samples unchanged, namely the low-dimensional embedding coordinates of each sample. The mathematical formulation is

$$\min \sum_{i,j} \left(d_{\mathrm{G}}(\boldsymbol{x}_i, \boldsymbol{x}_j) - d_{\mathrm{E}}(\boldsymbol{y}_i, \boldsymbol{y}_j)\right)^2 \tag{1}$$

where d_{G} is the geodesic distance, which is defined locally to be the shortest path on the manifold, d_{E} is the Euclidean distance, and in a matrix form:

$$\min \|\tau(D_{\mathrm{G}}) - \tau(D_{\mathrm{E}})\|_{\mathrm{L}^2} \tag{2}$$

where D_{G} is the geodesic distance matrix, D_{E} is the Euclidean distance matrix, and the equation for the inner product matrix $\tau(D)$ can be simplified as follows: Let $S_{ij} = {D_{ij}}^2$ be the squared distance matrix between vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Let $H = I - (1/m)ee^{\mathrm{T}}$ where I is the identity matrix and e is a vector of all ones. Then, $\tau(D) = -\mathrm{HSH}/2$. Eq. (2) is solved under the MDS framework.

2.2 Isometric projection

Cai et al. [30] extended Isomap algorithm to learn a linear projection using a spectral graph optimization problem and then proposed the IsoP method. Let $\boldsymbol{X} = \left[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\right]$ be a set of original data in the high-dimensional space R^{D}. The purpose of IsoP is to find a transformation matrix $\boldsymbol{V}$ to map data in high-dimensional space into low-dimensional space in $R^{\mathrm{d}}(d \ll D)$, i.e.,

$$\boldsymbol{Y} = \left[\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N\right] = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{X} \tag{3}$$

In order to preserve the intrinsic geometrical structure of the samples, the objective function of IsoP is:

$$\min \|\tau(D_{\mathrm{G}}) - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\|^2 \tag{4}$$

To make the problem tractable, they imposed a constraint $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I}$, and rewrote the minimization problem as

$$\underset{\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{I}}{\arg\max}\ tr(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\tau(D_{\mathrm{G}})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}) \tag{5}$$

The objective function Eq. (5) can be solved with Lagrange multiplier method, which reduce Eq. (5) to a generalized eigenvalue problem:

$$\boldsymbol{X}\left[\tau(D_{\mathrm{G}})\right]\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v} = \lambda\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v} \tag{6}$$

Let $\boldsymbol{V} = \left[\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_d\right]$ is the matrix composed of the vector $\boldsymbol{v}_i$ from Eq. (6), the d-dimensional representation $\boldsymbol{y}_i$ of each data point can be obtained using the linear embedding formula given in Eq. (7),

$$\boldsymbol{x}_i \rightarrow \boldsymbol{y}_i = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i \tag{7}$$

IsoP is a linearization version of Isomap. Unlike Isomap, which is difficult in learning the new samples, IsoP can be defined anywhere in the ambient space, so IsoP is easier to use.

2.3 Autoencoder

The standard autoencoder consists of three parts: input layer, hidden layer, and output layer, and it is usually the fully connected neural layer. These three layers make up two parts: encoder and decoder. The encoder consists of input layer and hidden layer, and the decoder consists of hidden layer and output layer. The encoding stage can be regarded as converting the input data to the hidden layer expression through deterministic mapping, and the decoding stage tries to remap the hidden layer expression to the input data. The autoencoder trains the weight parameters by minimizing the error between the reconstruction and the original input data, to get the optimal feature representation of the input data.

In the past 10 years, many improved versions of autoencoder have been proposed one after another, and they have been widely used in various research fields. These improvements include Stacked Autoencoder [31, 32], Sparse Autoencoder [33], Convolutional Autoencoder [34], Variational Autoencoder [35], Denoising Autoencoder [36], Contractive Autoencoder [37], etc.

For autoencoder, when the number of neurons in the input layer is larger than that in the hidden layer, it is called the compressed structure or undercomplete model; otherwise, it is called the sparse structure or overcomplete model. However, if the number of neurons in the input layer and the hidden layer are equal, it is called the

equal-dimensional structure. If the activation function of the hidden layer is linear, the autoencoder is called linear autoencoder.

In this paper, the proposed method is based on the linear, undercomplete autoencoder with only one hidden layer.

3 Isometric projection with reconstruction

3.1 Framework

In this Section, we present the framework of the new IsoP method. Based on the encoder-decoder paradigm, the new IsoP can be divided into two stages.

The first stage is the conventional projection model of IsoP. The model maps the high-dimension data point $\boldsymbol{x}_i$ to the low-dimensional data point $\boldsymbol{y}_i$ with the linear mapping $\boldsymbol{y}_i = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i$. From the perspective of linear autoencoder, this mapping can be regarded as encoding each high-dimensional data point into a low-dimensional data point. Moreover, the mapping simultaneously preserves the intrinsic geometrical structure of the original samples and provides a more accurate representation of the true global structure of the data.

The second stage is to reconstruct the low-dimensional embedded points $\boldsymbol{y}_i$ to the original space with the linear autoencoder. Let $\hat{\boldsymbol{x}}_i$ as the reconstruction of the original data point $\boldsymbol{x}_i$, $\boldsymbol{V}^*$ as the weight matrix of the decoder. Mathematically, the process of decoding may be formulated as follows:

$$\hat{\boldsymbol{x}}_i = \boldsymbol{V}^*\boldsymbol{y}_i \tag{8}$$

To simplify the model, the weights of encoder and decoder in autoencoder can apply the method in Ref. [38], i.e., $\boldsymbol{V}^* = (\boldsymbol{V}^{\mathbf{T}})^{\mathbf{T}} = \boldsymbol{V}$. This shows that the forced symmetry is achieved between the encoder and decoder of the autoencoder by sharing the weights. Much recent works such as [39–41] show that this structure can work. Therefore, in this paper, we also refer to this structure. Thus, Eq. (8) can be rewritten as:

$$\hat{\boldsymbol{x}}_i = \boldsymbol{V}\boldsymbol{y}_i \tag{9}$$

And then, we also want to minimize the error between the input data $\boldsymbol{x}_i$ and its reconstruction $\hat{\boldsymbol{x}}_i$, so that the low-dimensional embedded point $\boldsymbol{y}_i$ "represents" the original sample $\boldsymbol{x}_i$ more accurately and effectively. The framework of the proposed IsoP-R method is shown in Fig. 1:

Thus, the new IsoP method has two advantages. One is that it still preserves the intrinsic geometrical structure of the original samples. The other is that the low-dimensional embedding data "represent" the original samples more accurately and effectively due to the additional reconstruction.

Since the new IsoP method adds a reconstruction phase to IsoP based on the encoder-decoder paradigm, it is named IsoP-R (Isometric Projection with Reconstruction).

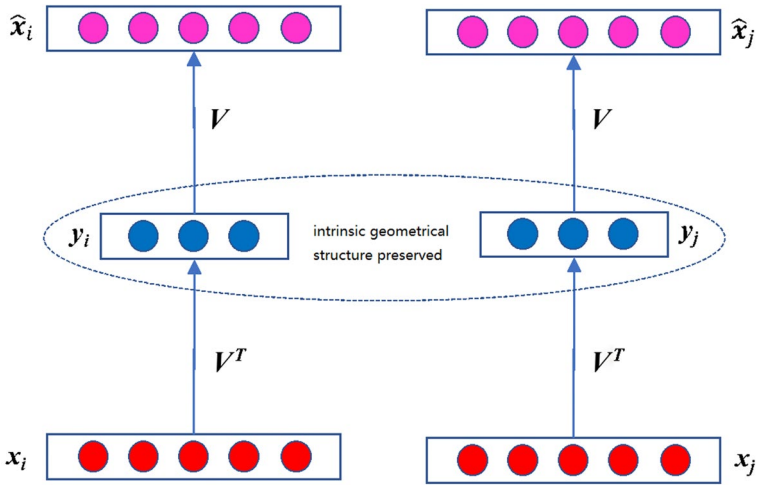

Fig. 1 The framework of the proposed IsoP-R method

3.2 The objective function

In the above framework, the objective function of the IsoP-R method is presented in this Section. The first stage of IsoP-R is the conventional projection model of IsoP, which preserves the intrinsic geometrical structure of the samples. Equation (5) is the original of IsoP method and finds the maximum. In this paper, Eq. (5) is regarded as the first item of the objective function of IsoP-R, and its minimum value is obtained, i.e.,

$$\mathcal{L}_{1\text{st}} = -tr\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\tau(D_{\mathrm{G}})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\right) \tag{10}$$

And the constraint imposed on Eq. (5) can be relaxed as:

$$\mathcal{L}_{2\text{nd}} = tr\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\right) - d \tag{11}$$

The parameter d is the dimension of low-dimensional space, which is described in Sect. 2.1. Equation (11) may be taken as the second item of the objective function of IsoP-R.

The second stage of IsoP-R is to reconstruct the low-dimensional embedded points $\boldsymbol{y}_i$ and minimize the error between the input data $\boldsymbol{x}_i$ and its reconstruction $\hat{\boldsymbol{x}}_i$. It can be formulated as:

$$\mathcal{L}_{3\text{rd}} = \sum_{i=1}^{N} \left\|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\right\|_2^2 \tag{12}$$

And take it as the third item of the objective function of IsoP-R. By bringing the linear mapping of IsoP into the reconstructed data, so one has:

$$\hat{\boldsymbol{x}}_i = \boldsymbol{V}\boldsymbol{y}_i = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i \tag{13}$$

Thus, substituting Eq. (13) into Eq. (12) can be rewritten as:

$$\begin{aligned}\mathcal{L}_{3\mathrm{rd}} &= \sum_{i=1}^{N}\left\|\boldsymbol{x}_i - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_i\right\|_2^2 \\ &= \sum_{i=1}^{N}\left\|\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right)\boldsymbol{x}_i\right\|_2^2 \\ &= tr\left(\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}}\right)\end{aligned} \tag{14}$$

Combine Eqs. (10), (11) and (14), the new method is to find such a projection matrix $\boldsymbol{V}$ and minimize the following objective function:

$$\begin{aligned}\mathcal{L}(\boldsymbol{V},\lambda,\gamma) &= \mathcal{L}_{1\mathrm{st}} + \lambda\mathcal{L}_{2\mathrm{nd}} + \gamma\mathcal{L}_{3\mathrm{rd}} \\ &= -tr\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\tau(D_{\mathrm{G}})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\right) + \lambda\left(tr\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\right) - d\right) \\ &\quad + \gamma tr\left(\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}}\right)\end{aligned} \tag{15}$$

where λ and γ are the balance parameters.

In Eq. (15), if $\gamma = 0$, Eq. (15) is reduced to

$$\mathcal{L}(\boldsymbol{V},\lambda) = -tr\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\tau(D_{\mathrm{G}})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\right) + \lambda\left(tr\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\right) - d\right) \tag{16}$$

Equation (16) is the expression of applying the Lagrange multiplier method to Eq. (5). According to Eqs. (5) and (6), IsoP is to seek the eigenvectors corresponding to the maximum eigenvalues. Therefore, in Eq. (16), it is reasonable to take smaller values for the parameter λ.

In Eq. (15), the first item $\mathcal{L}_{1\mathrm{st}}$ can be regarded as intrinsic geometrical structure preserving item and its coefficient can be regarded as 1. The third item $\mathcal{L}_{3\mathrm{rd}}$ can be regarded as the reconstruction item and the parameter γ reflects the proportion of reconstruction. If $0 < \gamma < 1$, the proportion of reconstruction item is less than that of intrinsic geometrical structure preserving item. If $\gamma \geq 1$, the proportion of reconstruction item is equal to or greater than that of intrinsic geometrical structure preserving item.

It is worth mentioning that $\|\boldsymbol{V}\|_F^2$ regularization has not been considered in our model. It is unnecessary because the weights of the encoder and decoder are tied, i.e., $\boldsymbol{V}^* = \boldsymbol{V}$. If the norm $\|\boldsymbol{V}\|_F^2$ is large, the low-dimensional projection produced by the encoder will have large values; and then, in the decoding stage, after the low-dimensional projection is multiplied by the matrix $\boldsymbol{V}$, bad reconstruction will be produced. That is, the $\|\boldsymbol{V}\|_F^2$ regularization has been automatically handled by the reconstruction constraints [38].

3.3 Optimization

The IsoP method is linear, so it is easy to solve. However, the objective function of IsoP-R, Eq. (15), is a nonlinear function with the matrix $\boldsymbol{V}$, the parameter λ and γ. Although it is difficult to solve directly, we can use the multivariate gradient descent method to obtain the optimal projection matrix $\boldsymbol{V}$ and the parameter λ and γ.

The main process of multivariate gradient descent consists of the following two steps:

Firstly, calculate the gradient of the objective function (15). The mathematical form of the gradient is as follows:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \boldsymbol{V}} &= \frac{\mathrm{d}\mathcal{L}_{1\mathrm{st}}}{\mathrm{d}\boldsymbol{V}} + \lambda \frac{\mathrm{d}\mathcal{L}_{2\mathrm{nd}}}{\mathrm{d}\boldsymbol{V}} + \gamma \frac{\mathrm{d}\mathcal{L}_{3\mathrm{rd}}}{\mathrm{d}\boldsymbol{V}} \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= tr\left(\boldsymbol{V}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{V}\right) - d \\
\frac{\partial \mathcal{L}}{\partial \gamma} &= tr\left(\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right) \boldsymbol{X} \boldsymbol{X}^{\mathrm{T}} \left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}}\right)
\end{aligned} \tag{17}$$

Where,

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{L}_{1\mathrm{st}}}{\mathrm{d}\boldsymbol{V}} &= -2\boldsymbol{X}\tau(D_{\mathrm{G}})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V} \\
\frac{\mathrm{d}\mathcal{L}_{2\mathrm{nd}}}{\mathrm{d}\boldsymbol{V}} &= 2\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V} \\
\frac{\mathrm{d}\mathcal{L}_{3\mathrm{rd}}}{\mathrm{d}\boldsymbol{V}} &= -4\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}\right)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}
\end{aligned} \tag{18}$$

Secondly, using the gradient (17), update the matrix $\boldsymbol{V}$, the parameter λ and γ with the following formula until the optimal projection matrix $\boldsymbol{V}$, the parameter λ and γ are found:

$$\begin{aligned}
\boldsymbol{V}_{t+1} &= \boldsymbol{V}_t - \alpha \frac{\partial \mathcal{L}}{\partial \boldsymbol{V}_t} \\
\lambda_{t+1} &= \lambda_t - \alpha \frac{\partial \mathcal{L}}{\partial \lambda_t} \\
\gamma_{t+1} &= \gamma_t - \alpha \frac{\partial \mathcal{L}}{\partial \gamma_t}
\end{aligned} \tag{19}$$

Where the parameter α represents the search step in the gradient direction, t represents the t-th step. The algorithm of IsoP-R can be expressed as Table 1.

Through the above algorithm process, we obtain the optimal projection matrix $\boldsymbol{V}$, the optimal parameter λ and γ. Through the linear mapping Eq. (7), the data points $\boldsymbol{x}_i$ in the high-dimensional space are embedded into low-dimensional space as the data point $\boldsymbol{y}_i$.

Table 1 The algorithm of IsoP-R

Algorithm: multivariate gradient descent method for training the optimal matrix $\boldsymbol{V}$

Input: the data matrix $\boldsymbol{X}$

Initialize: initialize $\boldsymbol{V}$, parameter λ and γ randomly, set a step size α of gradient descent,and set a threshold ϵ for the change of loss function.

Pre-calculation: calculate the distance matrix D_{G}, the inner product matrix $\tau(D_{\mathrm{G}})$, and the objective function $\mathcal{L}$ using Eq. (15).

Repeat:

1. Calculate the gradient $\partial\mathcal{L}/\partial\boldsymbol{V}$, $\partial\mathcal{L}/\partial\lambda$ and $\partial\mathcal{L}/\partial\gamma$ using Eq. (17)
2. Update the matrix $\boldsymbol{V}$, λ and γ using Eq. (19)
3. Calculate the objective function $\mathcal{L}$ using Eq. (15).

Until the change of objective function $< \varepsilon$

Output: the projection matrix $\boldsymbol{V}$, λ and γ

4 Experimental results

In this Section, we make experiments on the Handwritten Alphadigits dataset, COIL-100 object dataset, Olivetti Research Laboratory(ORL) and Georgia Tech face dataset. The proposed IsoP-R is compared with LPP, Neighborhood Preserving Embedding (NPE) [42], IsoP, OIsoP, EIsoP, RIsoP, IsoP-MMC, Supervised Isometric Mapping (S-Isomap) [43, 44] and Wasserstein Isometric Mapping (Wassmap) [45], respectively. Besides, principal component analysis (PCA) is also used to make experiments and compare.

Compared with IsoP, OIsoP, etc., LPP and NPE aim to preserve the local structure of the data. The core of LPP is to make points that are close in higher dimensional space also close in lower dimensional space. The basic idea of NPE is to maintain the manifold structure of the whole data by keeping the nearest neighbor reconstruction coefficient of each point in the low-dimensional embedding space. IsoP-MMC is an application of the Maximum Margin Criterion (MMC) [46] to IsoP. The MMC maximizes the average margin between classes and makes the data separable. While S-Isomap and Wassmap are variants of Isomap, where S-Isomap replaces the Euclidean distance in a high-dimensional manifold with the dissimilarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ to ensure its classification ability. Wassmap uses Wasserstein distance instead of Euclidean distance.

For the IsoP, NPE and LPP methods, PCA needs to be used to reduce the dimension of the original sample. For all data sets, 98% of the principal components are retained in this step.

4.1 Experiments datasets

Handwritten Alphadigits dataset. The dataset includes a total of 36 classes composed of letters and numbers, with 39 samples per class. In the experiment, each image is a hand-written image with 20×16 pixels.

COIL-100 is a dataset composed of images of different objects at different angles in a $360^{c}irc$ rotation. It contains 100 objects and is photographed every $5^{c}irc$, so each object has a total of 72 images. In the experiment, the image is 32×32 pixels.

ORL face dataset contains a total of 400 images of 40 different people, and the images of some volunteers include changes in poses, expressions and facial accessories. The facial expressions and details of the collected objects all change, such as smiling or not, eyes open or closed, and wearing or not wearing glasses, etc. In the experiment, the image size is adjusted to 32×32 pixels.

Georgia Tech face dataset contains 50 distinct individuals, each with 15 distinct images. These images show frontal or oblique faces with different facial expressions, lighting conditions, and scales. Each image was manually labeled to determine the location of the face in the image. The size of each image is 32×32 pixels in the experiments.

Some sample images of the Handwritten Alphadigits, COIL-100, ORL and Georgia Tech are shown in the following Fig. 2.

4.2 Experimental setting

First, we randomly select p images from the dataset as the training set and the remaining images as the test set. Second, for a given p, the dimension of the low-dimensional subspace increases from 10 to 100 in step 5. Third, for a fixed p and subspace dimension, increase the neighborhood size k from 5 to 25 in step 5. Fourth, the best recognition rate of the best k value is used as the recognition rate of the current p samples and subspaces. Then, for the dimension of each subspace, we can get the corresponding recognition rate.

We regard the above process as a cycle. For a given p, we calculate 10 cycles. In this way, there are 10 recognition rates for each subspace dimension, and then we take their average value as the recognition rate for the current p and subspace

Fig. 2 Some samples in the experiments. The first line is Handwritten Alphadigits database. The second line is COIL-100 database. The third line is ORL face dataset. The fourth line is Georgia Tech face dataset

Table 2 Best average recognition accuracy (in percent), standard deviation, and the optimal dimension (in parentheses) on Handwritten Alphadigts dataset

Method	5 trains	7 trains	9 trains
LPP	42.65 ± 1.38 (20)	46.51 ± 1.30 (25)	49.52 ± 1.15 (25)
NPE	36.62 ± 2.16 (35)	38.33 ± 2.71 (30)	42.59 ± 2.13 (35)
IsoP	48.26 ± 2.94 (35)	50.66 ± 1.24 (30)	51.87 ± 1.57 (25)
OIsoP	54.88 ± 1.95 (65)	58.32 ± 1.02 (100)	61.09 ± 1.61 (95)
EIsoP	49.24 ± 3.44 (90)	53.05 ± 1.13 (90)	52.40 ± 3.37 (90)
RIsoP	50.62 ± 2.24 (35)	52.94 ± 1.54 (30)	54.13 ± 1.76 (60)
IsoP-MMC	51.17 ± 0.89 (30)	54.63 ± 1.46 (50)	57.00 ± 1.43 (35)
S-Isomap	55.19 ± 2.66 (100)	57.44 ± 0.77 (100)	58.37 ± 0.99 (100)
Wassmap	52.39 ± 2.05 (35)	56.37 ± 1.41 (10)	59.16 ± 1.38 (95)
IsoP-R	**56.05 ± 1.42 (5)**	**60.88 ± 1.35 (40)**	**63.21 ± 1.16 (85)**

Table 3 Best average recognition accuracy (in percent), standard deviation, and the optimal dimension (in parentheses) on COIL-100 dataset

Method	6 trains	8 trains	10 trains
LPP	67.41 ± 2.10 (45)	73.25 ± 2.61 (50)	77.77 ± 1.08 (55)
NPE	67.38 ± 2.14 (100)	71.67 ± 2.64 (50)	76.62 ± 0.97 (55)
IsoP	68.73 ± 2.13 (45)	73.90 ± 2.78 (50)	78.12 ± 1.84 (55)
OIsoP	70.92 ± 2.87 (100)	73.93 ± 3.32 (100)	78.86 ± 4.19 (95)
EIsoP	72.24 ± 1.79 (80)	74.71 ± 1.83 (95)	**88.58 ± 0.97 (70)**
RIsoP	72.08 ± 3.39 (100)	76.89 ± 3.02 (50)	80.79 ± 2.09 (55)
IsoP-MMC	76.79 ± 2.14 (45)	81.15 ± 1.48 (80)	84.24 ± 1.45 (100)
S-Isomap	76.28 ± 2.85 (100)	80.69 ± 1.41 (85)	86.31 ± 1.56 (100)
Wassmap	71.80 ± 3.21 (100)	74.92 ± 2.92 (50)	77.56 ± 1.93 (55)
IsoP-R	**76.92 ± 3.29 (85)**	**81.17 ± 2.46 (80)**	84.80 ± 4.55 (85)

Table 4 Best average recognition accuracy (in percent), standard deviation, and the optimal dimension (in parentheses) on the ORL face dataset

Method	6 trains	7 trains	8 trains
LPP	88.00 ± 1.52 (85)	91.50 ± 2.75 (95)	93.00 ± 3.50 (100)
NPE	91.00 ± 2.03 (100)	92.91 ± 3.35 (100)	93.87 ± 3.23 (60)
IsoP	91.44 ± 2.09 (100)	93.91 ± 1.58 (70)	95.75 ± 2.57 (30)
OIsoP	94.75 ± 1.51 (100)	96.33 ± 1.45 (100)	97.50 ± 1.36 (80)
EIsoP	94.75 ± 1.72 (100)	96.41 ± 1.62 (70)	97.00 ± 1.39 (90)
RIsoP	93.31 ± 1.97 (100)	95.66 ± 1.81 (70)	96.12 ± 1.52 (80)
IsoP-MMC	92.06 ± 1.48 (15)	93.50 ± 2.32 (30)	95.13 ± 1.63 (20)
S-Isomap	94.19 ± 1.40 (95)	95.42 ± 1.80 (95)	96.00 ± 1.35 (100)
Wassmap	94.44 ± 1.26 (75)	95.75 ± 1.08 (50)	97.00 ± 1.78 (30)
IsoP-R	**95.56 ± 1.80 (95)**	**97.08 ± 1.59 (100)**	**98.25 ± 1.15 (65)**

Table 5 Best average recognition accuracy (in percent), standard deviation, and the optimal dimension (in parentheses) on the Georgia Tech face dataset

Method	7 trains	8 trains	9 trains
LPP	49.25 ± 1.40 (75)	51.77 ± 2.27 (45)	54.63 ± 1.44 (55)
NPE	52.47 ± 2.48 (100)	53.65 ± 1.72 (90)	54.63 ± 2.81 (85)
IsoP	58.67 ± 1.78 (75)	60.80 ± 1.76 (85)	61.50 ± 1.71 (85)
OIsoP	67.35 ± 1.88 (75)	69.85 ± 1.62 (95)	71.70 ± 2.23 (85)
EIsoP	64.90 ± 1.83 (100)	65.45 ± 3.71 (90)	67.46 ± 1.84 (100)
RIsoP	62.92 ± 2.54 (100)	64.08 ± 2.13 (95)	65.03 ± 3.07 (95)
IsoP-MMC	59.03 ± 1.51 (45)	60.66 ± 1.04 (45)	62.50 ± 0.85 (55)
S-Isomap	63.43 ± 2.13 (100)	64.92 ± 1.42 (95)	65.87 ± 2.50 (100)
Wassmap	62.33 ± 1.54 (95)	64.23 ± 1.46 (70)	65.93 ± 1.58 (85)
IsoP-R	**67.70 ± 1.63 (100)**	69.37 ± 1.85 (95)	70.83 ± 1.69 (100)

dimension. Finally, we take the best recognition rate from the best subspace dimension as the result in the case that the training sample is p.

4.3 Experimental results

The best recognition rates, standard deviation, and the optimal subspace dimension are reported in Tables 2, 3, 4 and 5 for the above ten methods based on the

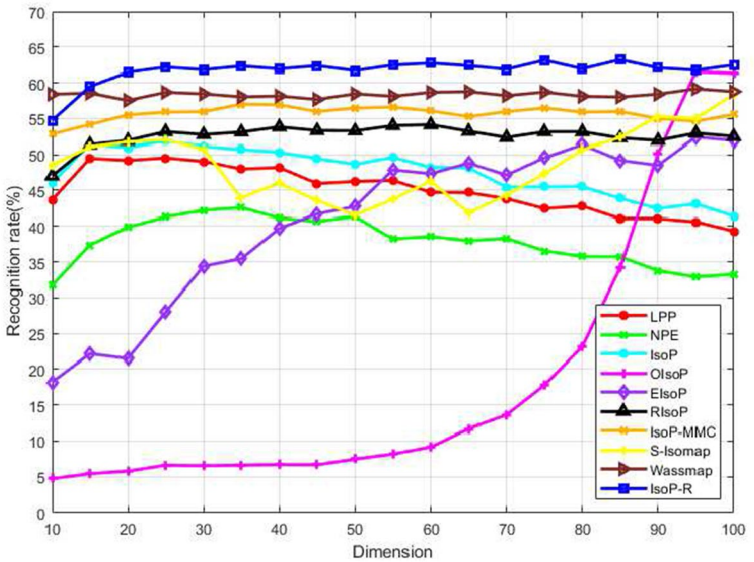

Fig. 3 The recognition rates versus the subspace dimension on Handwritten Alphadigits data set (9 training samples)

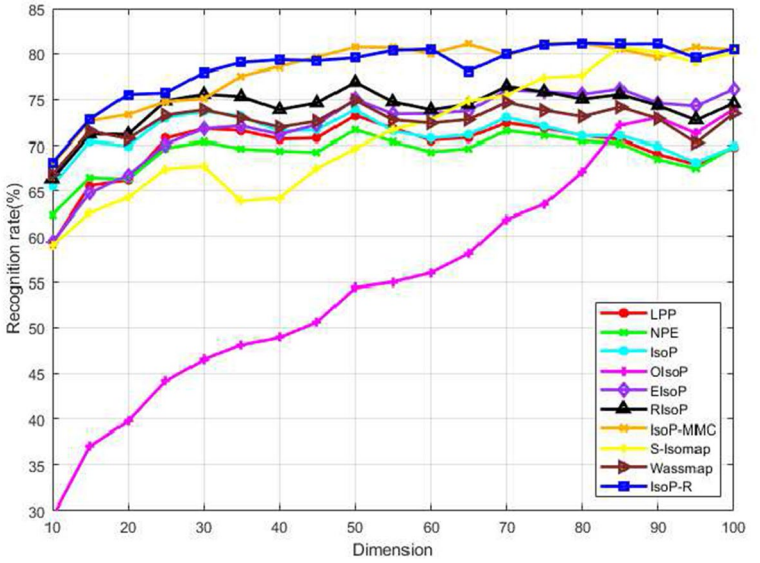

Fig. 4 The recognition rates versus the subspace dimension on COIL-100 data set (8 training samples)

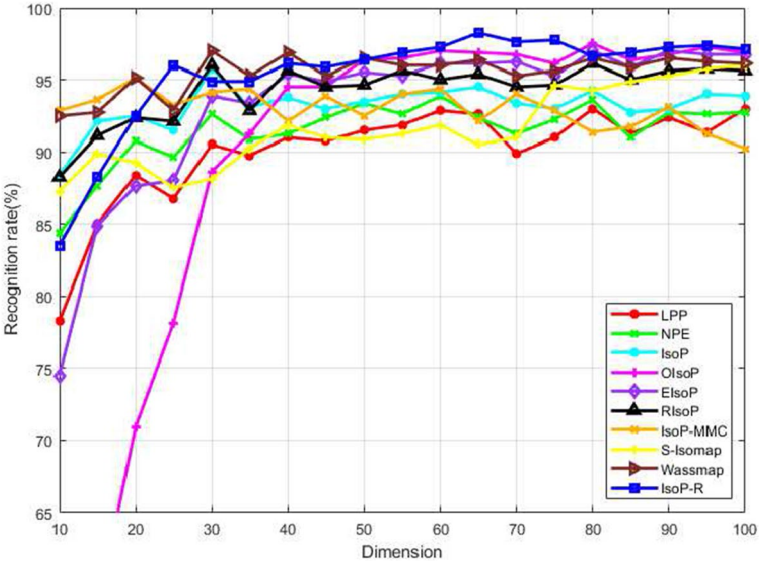

Fig. 5 The recognition rates versus the subspace dimension on ORL face database (8 training samples)

experimental steps in the four datasets. Bold in the table indicates the maximum recognition rate among the ten methods under the same training sample.

Generally, the performances of these methods are also related to the reduced dimension. Based on the above experimental results, the recognition results of

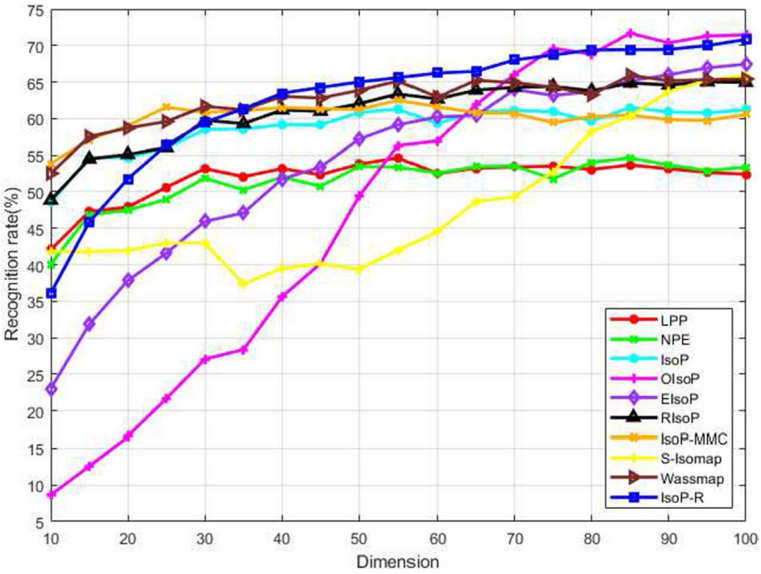

Fig. 6 The recognition rates versus the subspace dimension on Georgia Tech face data set (9 training samples)

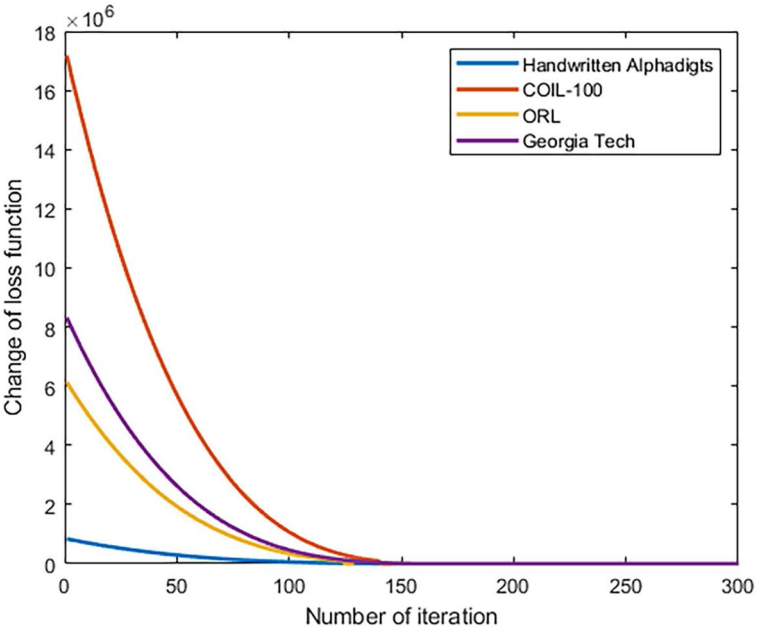

Fig. 7 The convergence curve of IsoP-R

each method in different reduced dimensions are analyzed, as shown in Figs. 3, 4, 5 and 6.

4.4 Convergence analysis

The convergence curves of IsoP-R on four datasets are presented in Fig. 7. Where, the training samples are 5, 6, 8 and 9 in Handwritten Alphadigits, COIL-100, ORL and Georgia Tech datasets, respectively, the neighborhood size parameter $k = 10$, and the subspace dimension is 40. As can be seen from Fig. 7, IsoP-R can converge in only 100–150 steps on four data sets. The good convergence ensures the reliability of the model.

4.5 Analysis

From the above experiments, it can be concluded that: (1) In Tables 2, 3, 4 and 5, the best recognition rates from the best subspace dimensions are reported. And, from Figs. 3, 4, 5 and 6, it can be seen that when the subspace dimension varies, the performance of IsoP-R is significantly better than that of IsoP and its variants. (2) As a new method of IsoP, IsoP-R is compared with LPP, NPE, IsoP and its improved methods. Although the performance gains of IsoP-R are different for different data sets, Iso-R can achieve the best performance compared with other methods in most cases. Specifically, compared with IsoP, IsoP-R improves the recognition rate accuracy by about 12% on Handwritten Alphadigits and improves by 8%, 4% and 8% on COIL-100, ORL and Georgia Tech, respectively. (3) From Table 3, it can be seen that when dataset is COIL-100 and the training sample is 10, IsoP-R is less effective than EIsoP and S-Isomap. However, S-Isomap can only be mapped to the training set, while IsoP-R can be mapped to the test set. (4) As can be seen from Table 5, IsoP-R is slightly lower than OIsoP on the Georgia Tech dataset but better than other methods. In addition, from Fig. 6, it can be seen that OIsoP performs poorly in lower dimension, surpassing IsoP-R only in a few dimensions, while IsoP-R has a relatively stable trend. In fact, OIsoP performs poorly on the Georgia Tech, ORL, COIL-100, and Handwritten Alphadigits datasets when the dimensionality is low.

5 Conclusion

In this paper, based on the "encoding-decoding" mechanism, a new IsoP method, called IsoP-R (Isometric Projection with Reconstruction), has been proposed. The main idea of IsoP-R is to take the conventional projection of IsoP as the encoding stage and use the decoder to reconstruct the original data from the projected low-dimensional data. Compared with IsoP, IsoP-R makes the low-dimensional features more accurately and effectively "represent" the original high-dimensional data with the additional reconstruction stage. The experimental results on Handwritten Alphadigits, COIL-100, ORL and Georgia Tech face datasets verify the effectiveness and superiority of the new methods.

Based on the similarities between IsoP and the other manifold-based dimensionality reduction methods (for example LPP, NPE, linear LTSA (LLTSA) [47], etc.), the further research is to use the framework of IsoP-R to the above methods so that these methods can be extended to new versions and achieve better performance.

Author contributions RR proposed this new idea, and QZ completed the experiment. QZ wrote the main manuscript text. RR and QZ reviewed the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by National Natural Science Foundation of China under Grant 61876026, Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K202100505, Chongqing Technology Innovation and Application Development Project under Grant cstc2020jscx-msxmX0190 and cstc2019jscxmbdxX0061.

Data availability The datasets generated during and/or analyzed during the current study are available in the https://cs.nyu.edu/~roweis/data.html, http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php, https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html, http://www.anefian.com/face_reco.htm.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Ethical approval This declaration is "Not applicable".

References

1. Berisha V, Krantsevich C, Hahn PR et al (2021) Digital medicine and the curse of dimensionality. NPJ Digit Med 4(1):153
2. Bach F (2017) Breaking the curse of dimensionality with convex neural networks. J Mach Learn Res 18(1):629–681
3. El Naqa I, Murphy MJ (2015) What is machine learning? Machine learning in radiation oncology. Springer, Cham, pp 3–11
4. Jia W, Sun M, Lian J et al (2022) Feature dimensionality reduction: a review. Complex Intell Syst 8(3):2663–2693
5. Sarveniazi A (2014) An actual survey of dimensionality reduction. Am J Comput Math 2014(4):55–72
6. Ayesha S, Hanif MK, Talib R (2020) Overview and comparative study of dimensionality reduction techniques for high dimensional data. Inf Fusion 59:44–58
7. Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 374(2065):20150202
8. Tharwat A, Gaber T, Ibrahim A et al (2017) Linear discriminant analysis: a detailed tutorial. AI Commun 30(2):169–190
9. Sharma A, Paliwal KK (2015) Linear discriminant analysis for the small sample size problem: an overview. Int J Mach Learn Cybern 6:443–454
10. Valencia XPB, Becerra M, Ospina AC et al (2017) Kernel-based framework for spectral dimensionality reduction and clustering formulation: a theoretical study. ADCAIJ Adv Distrib Comput Artif Intell J 6(1):31–40
11. Anowar F, Sadaoui S, Selim B (2021) Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). Comput Sci Rev 40:100378

12. Ghojogh B, Ghodsi A, Karray F, et al (2020) Locally linear embedding and its variants: Tutorial and survey. arXiv preprint arXiv:2011.10925
13. Li W, Zhang L, Zhang L et al (2017) Gpu parallel implementation of isometric mapping for hyperspectral classification. IEEE Geosci Remote Sens Lett 14(9):1532–1536
14. Wang R, Nie F, Hong R et al (2017) Fast and orthogonal locality preserving projections for dimensionality reduction. IEEE Trans Image Process 26(10):5019–5030
15. Li B, Li YR, Zhang XL (2019) A survey on Laplacian eigenmaps based manifold learning methods. Neurocomputing 335:336–351
16. Torres L, Chan KS, Eliassi-Rad T (2020) Glee: geometric Laplacian eigenmap embedding. J Complex Netw 8(2):cnaa007
17. Al-juboori AM, Bu W, Wu X et al (2014) Palm vein verification using multiple features and isometric projection. Int J Signal Process Image Process Pattern Recognit 7(1):33–44
18. Hout MC, Papesh MH, Goldinger SD (2013) Multidimensional scaling. Wiley Interdiscip Rev Cogn Sci 4(1):93–103
19. Blouvshtein L, Cohen-Or D (2018) Outlier detection for robust multi-dimensional scaling. IEEE Trans Pattern Anal Mach Intell 41(9):2273–2279
20. Ran R, Fang B, Wu X (2018) Exponential neighborhood preserving embedding for face recognition. IEICE Trans Inf Syst 101(5):1410–1420
21. Modarresi K (2015) Unsupervised feature extraction using singular value decomposition. Proc Comput Sci 51:2417–2425
22. Xue J, Li J, Gong Y (2013) Restructuring of deep neural network acoustic models with singular value decomposition. In: Frederic B, Christophe C, Cecile F et al (eds) 14th Annual Conference of the International Speech Communication Association. Interspeech. France, Lyon, pp 2365–2369
23. Gui J, Sun Z, Jia W et al (2012) Discriminant sparse neighborhood preserving embedding for face recognition. Pattern Recogn 45(8):2884–2893
24. Sun Y, Lin ZH, Zhang RB (2012) A tensor based isometric projection algorithm. Advanced Engineering Forum, vol 4. Trans Tech Publ, Tsukuba, pp 183–188
25. Jiangyi L (2020) Research on isometric projection algorithm and its application in image recognition. Master's thesis, Chongqing Normal University
26. Zheng Y, Tang YY, Fang B, et al (2012) Orthogonal isometric projection. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). IEEE, Tsukuba, pp 405–408
27. Ge B, Shao Y, Shu Y (2012) Uncorrelated discriminant isometric projection for face recognition. International Conference on Information Computing and Applications. Springer, Heidelberg, pp 138–145
28. Liu B, Xia SX, Meng FR et al (2015) Extreme spectral regression for efficient regularized subspace learning. Neurocomputing 149:171–179
29. Wang SJ, Yan S, Yang J et al (2014) A general exponential framework for dimensionality reduction. IEEE Trans Image Process 23(2):920–930
30. Cai D, He X, Han J et al (2007) Isometric projection. In: Anthony C (eds) Proceedings of the 22nd National Conference on Artificial Intelligence - Volume 1. AAAI Press. Vancouver, British Columbia, Canada, pp 528–533
31. Zabalza J, Ren J, Zheng J et al (2016) Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. Neurocomputing 185:1–10
32. Bengio Y et al (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127
33. Meng L, Ding S, Xue Y (2017) Research on denoising sparse autoencoder. Int J Mach Learn Cybern 8:1719–1729
34. Seyfioğlu MS, Özbayoğlu AM, Gürbüz SZ (2018) Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. IEEE Trans Aerosp Electron Syst 54(4):1709–1723
35. Kusner MJ, Paige B, Hernández-Lobato JM (2017) Grammar variational autoencoder. In: International Conference on Machine Learning, PMLR, pp 1945–1954
36. Vincent P, Larochelle H, Bengio Y, et al (2008) Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine learning, pp 1096–1103
37. Diallo B, Hu J, Li T et al (2021) Deep embedding clustering based on contractive autoencoder. Neurocomputing 433:96–107
38. Ranzato M, Boureau YL, Cun Y et al (2007) Sparse feature learning for deep belief networks. Adv Neural Inf Process Syst 20:1185–1192

39. Kodirov E, Xiang T, Gong S (2017) Semantic autoencoder for zero-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3174–3183
40. Sun W, Shao S, Zhao R et al (2016) A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement 89:171–178
41. Zeng N, Zhang H, Song B et al (2018) Facial expression recognition via learning deep sparse autoencoders. Neurocomputing 273:643–649
42. He X, Cai D, Yan S et al (2005) Neighborhood preserving embedding. Tenth IEEE International Conference on Computer Vision (ICCV'05), vol 2. IEEE, Heidelberg, pp 1208–1213
43. Li K, Zhang S, Yan D et al (2020) Prediction of hot spots in protein-dna binding interfaces based on supervised isometric feature mapping and extreme gradient boosting. BMC Bioinform 21:1–10
44. Wang Z, Yao L, Cai Y (2020) Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine. Measurement 156:107574
45. Hamm K, Henscheid N, Kang S (2022) Wassmap: Wasserstein isometric mapping for image manifold learning. arXiv preprint arXiv:2204.06645
46. Wan M, Lai Z, Yang G et al (2017) Local graph embedding based on maximum margin criterion via fuzzy set. Fuzzy Sets Syst 318:120–131
47. Su Z, Tang B, Liu Z et al (2015) Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. Neurocomputing 157:208–222