AMSC808N-HW5

Chenyang Fang
https://github.com/Chenyang-Fang/AMSC808N

November 19, 2020

1. First I will consider the following optimization problem:

$$\min \sum_{i,j} k_{ij}\|y_i - y_j\|_2^2 = tr(Y^T LY) = \sum_{j=1}^{m} y_i^T L y_j,$$

subject to

$$Y^T QY = I$$

where $Y = [y_1, y_2, \ldots, y_m]$. The Lagrangian is

$$\begin{aligned} L(Y, \boldsymbol{\lambda}) &= \sum_{j=1}^{m} y_i^T L y_j - \boldsymbol{\lambda}(Y^T QY - I) \\ &= \sum_{j=1}^{m} y_i^T L y_j - \sum_{i,j=1}^{n} \lambda_{ij}(y_i^T Q y_j - \theta_i j) \end{aligned}$$

The second equality is because $(Y^T QY)_{ij} = y_i^T Q y_j$ and θ_{ij} are the index variable, $\theta_{ij} = 1$ when $i = j$ and 0 otherwise. Then taking the gradient of $L(Y, \boldsymbol{\lambda})$ with respect to y_i, we can get

$$\nabla_{y_i} = 2\sum_{j=1}^{m} L y_i - 2\sum_{j=1}^{n} \lambda_{ij} Q y_j.$$

Set $\nabla_{y_i} = 0$ and $\lambda_{ij} = 0$ when $i \neq j$, then we can get

$$L y_i = \lambda_{ii} Q y_i.$$

Based on the definition of graph Laplacian L, we can get $1_{n\times 1}$ is the eigenvector corresponding to the eigenvalue 0. Laplacian is a symmetrix, positive semidefinite matrix, so all other eigenvectors are corresponding to the positive eigenvalues. To eliminate this trivial solution, and the solution is given by the eigenvector with smallest nonzero eigenvalue, the second constrained $Y^T Q 1_{n\times 1} = 0$ is satisfied by removing the trivial value. For the other eigenvalue $\lambda_{ii} > 0$, by normalizing the eigenvector y_i, then we can get

$$y_i^T Q y_i = 1.$$

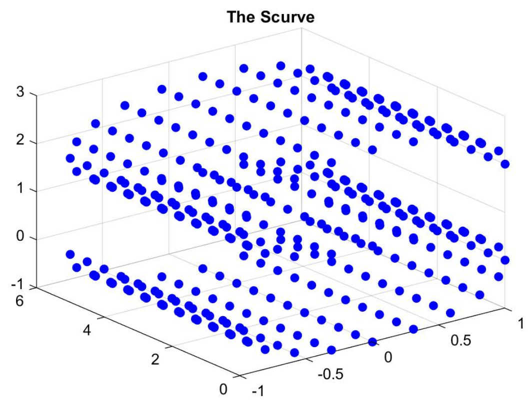
For the condition $i \neq j$, $y_i^T Q y_i = 0$ are satified due to the orthogonal property. According to the lecture notes, the following generalized eigenvalue problem is solved by R:

$$LR = QRM, \quad M = diag\{\mu_0, \mu_1, \ldots, \mu_{n-1}\},$$

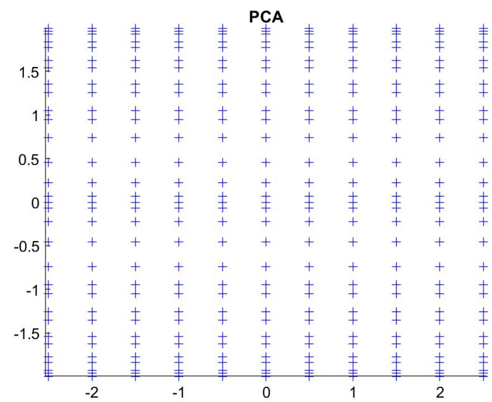
where $0 = \mu_0 \leq \mu_1 \leq \cdots \leq \mu_{n-1}$. This generalized eigenvalue problem is indetical to the $L y_i = \lambda_{ii} Q y_i$. So it is shown that the Laplacian eigenmap to $\mathbb{R}^m$ is the solution to the optimization problem.

2. (a) **Dataset 1** In this problem, five methods will be used to do the dimension reduction for the s-curve data. For the PCA and t-SNE method, the build-in matlab function will be used. For the isomap method, the code in the lecture note will be used. S.Roweis's code will be used for the LLE method. In the isomap and LLE mehtod, $k = 12$ is used.

- **PCA**



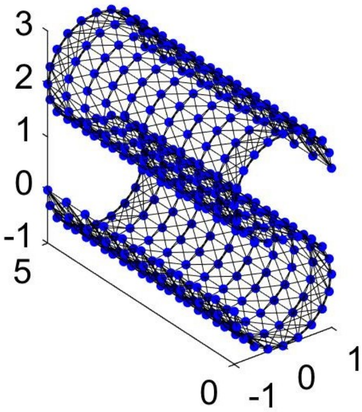
(a) The scurve.



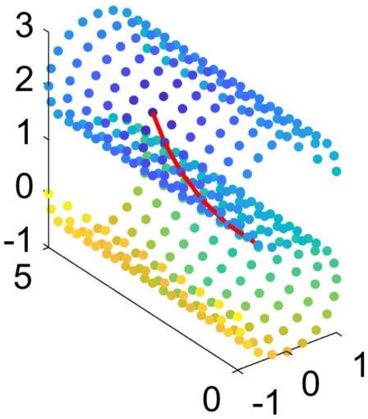
(b) PCA.

Figure 1: The PCA method applied to S-curve data.

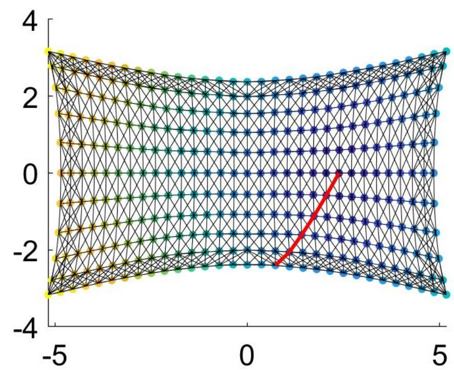
- **Isomap**



(a) The scurve.



(b) The distances from a randomly chosen point and the shortest path between it and a randomly chosen endpoint.



(c) The embedding.

Figure 2: The k-isomap applied to S-curve data.

- **LLE**

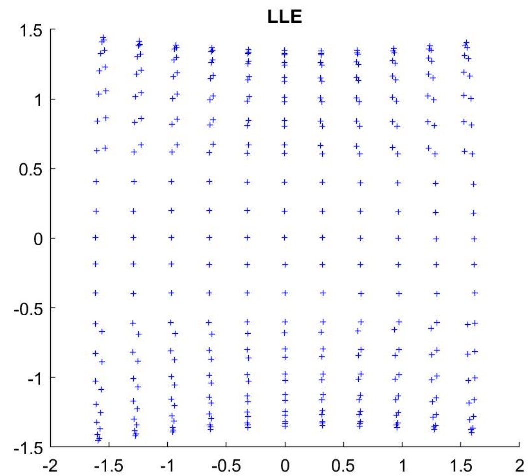


Figure 3: The LLE method applied to S-curve data.

- **t-SNE**

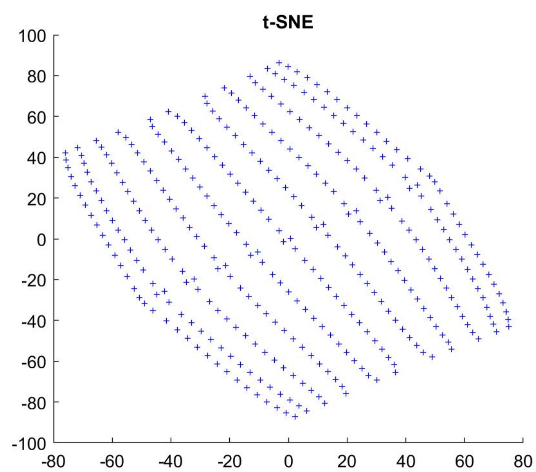


Figure 4: The t-SNE method applied to S-curve data.

- **DIffusion map**

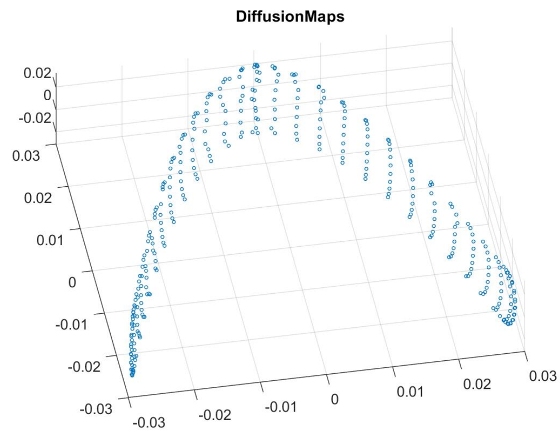
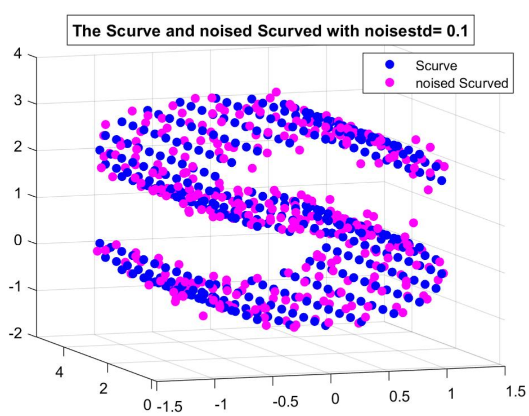


Figure 5: The t-SNE method applied to S-curve data.

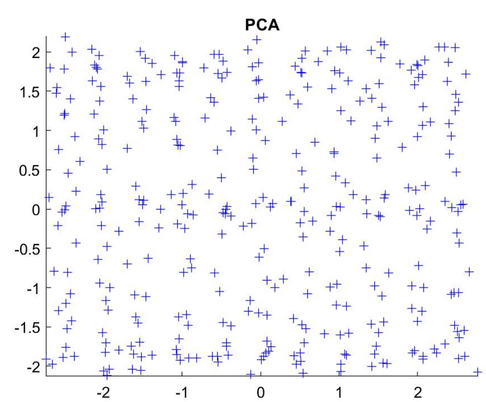
(b) **Dataset 2**

In the dataset2, Gaussian noise is added to the s-curve data to test the performance of each method. For the PCA method,the parameter Gaussian standard deviation is chosen 0.1, 0.3, 0.5 respectively, and the scurve and noised scurve are plotted in the following figures. And the embedding results are also plotted correspondingly.
For the isomap method, the parameter Gaussian standard deviation is chosen 0.1, 0.5. If the parameter Gaussian standard deviation is continuous increasing, the isomap method will fail.
For the LLE method and t-SNE method, the parameter Gaussian standard deviation is chosen 0.1, 0.3, 0.5 respectively, and the corresponding results are plotted. The LLe method works for the large noise, but the result is not accurate. The matlab shows that first input matrix is close to singular or badly scaled.
Based on the result, it is illustrated that the isomap method is sensitive to the noise. When the noise is large, the method will fail. While for the PCA, LLE and t-SNE method, even the mehtod can work, the results are not accurate. Only the diffusion map method can generate a reasonable result.

- **PCA**

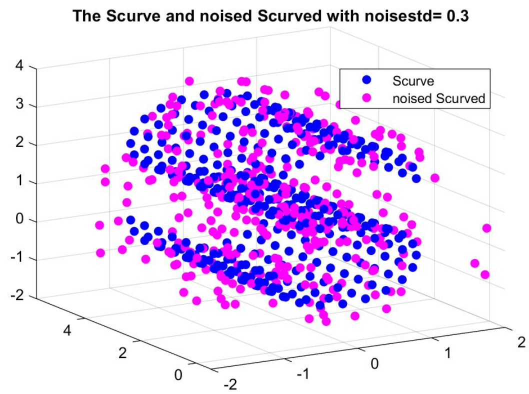


(a) The scurve and noised scurve data.

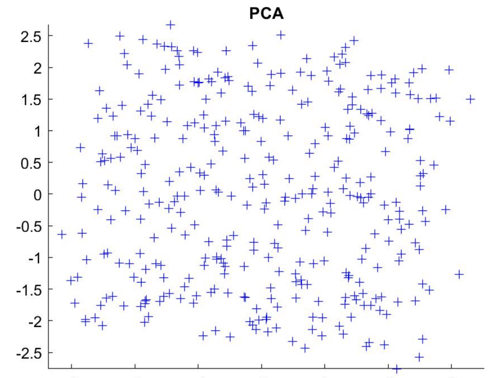


(b) The embedding.

Figure 6: The PCA method applied to noised S-curve data with noisestd=0.1.

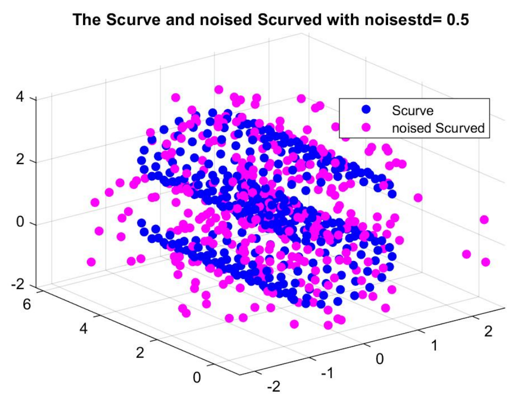


(a) The scurve The scurve and noised scurve data.

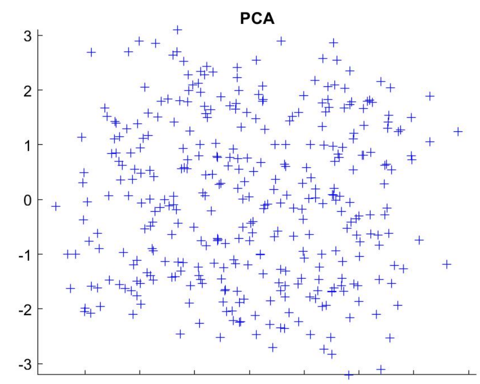


(b) The embedding.

Figure 7: The PCA method applied to noised S-curve data with noisestd=0.3.



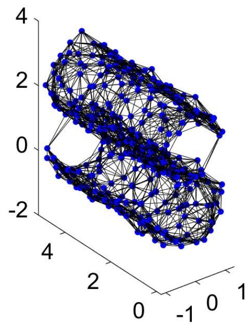
(a) The scurve The scurve and noised scurve data.



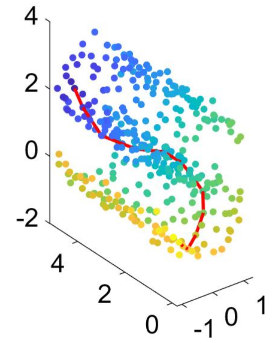
(b) The embedding.

Figure 8: The PCA method applied to noised S-curve data with noisestd=0.5.

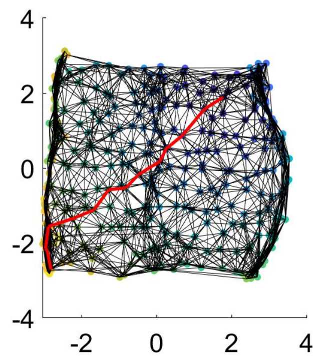
- **Isomap**



(a) The noised scurve.

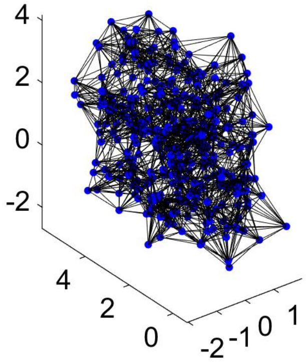


(b) The distances from a randomly chosen point and the shortest path between it and a randomly chosen endpoint.

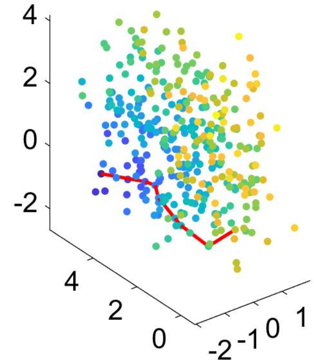


(c) The distances from a randomly chosen point and the shortest path between it and a randomly chosen endpoint.

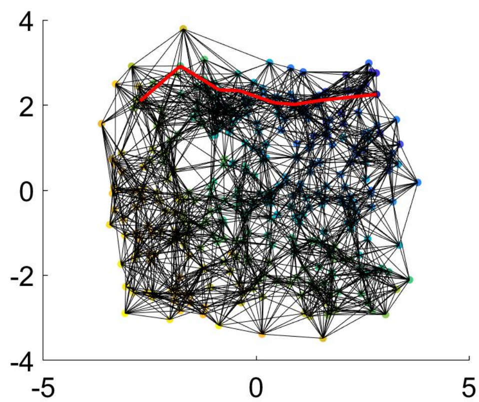
Figure 9: The k-isomap applied to noised S-curve data with noisestd=0.1.



(a) The noised scurve.



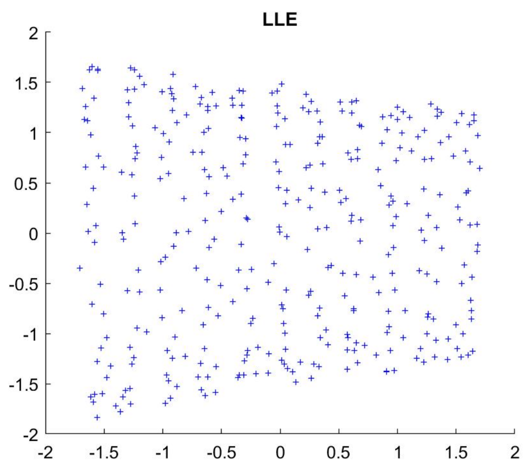
(b) The distances from a randomly chosen point and the shortest path between it and a randomly chosen endpoint.



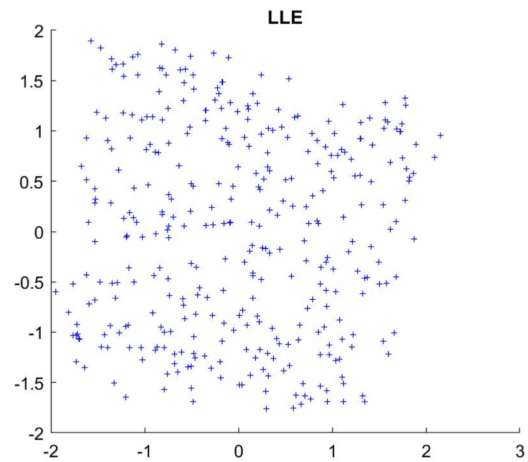
(c) The distances from a randomly chosen point and the shortest path between it and a randomly chosen endpoint.

Figure 10: The k-isomap applied to noised S-curve data with noisestd=0.5.

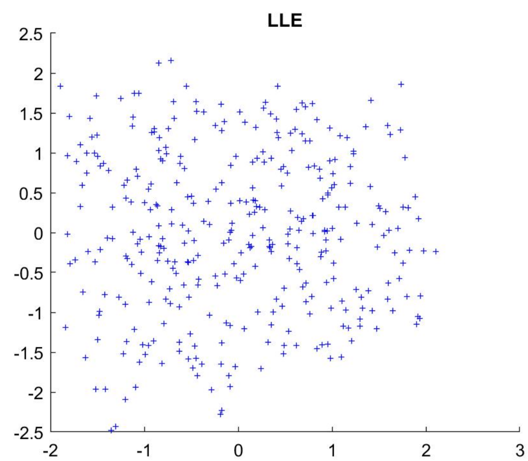
- **LLE**



(a) The embedding with noisestd=0.1.



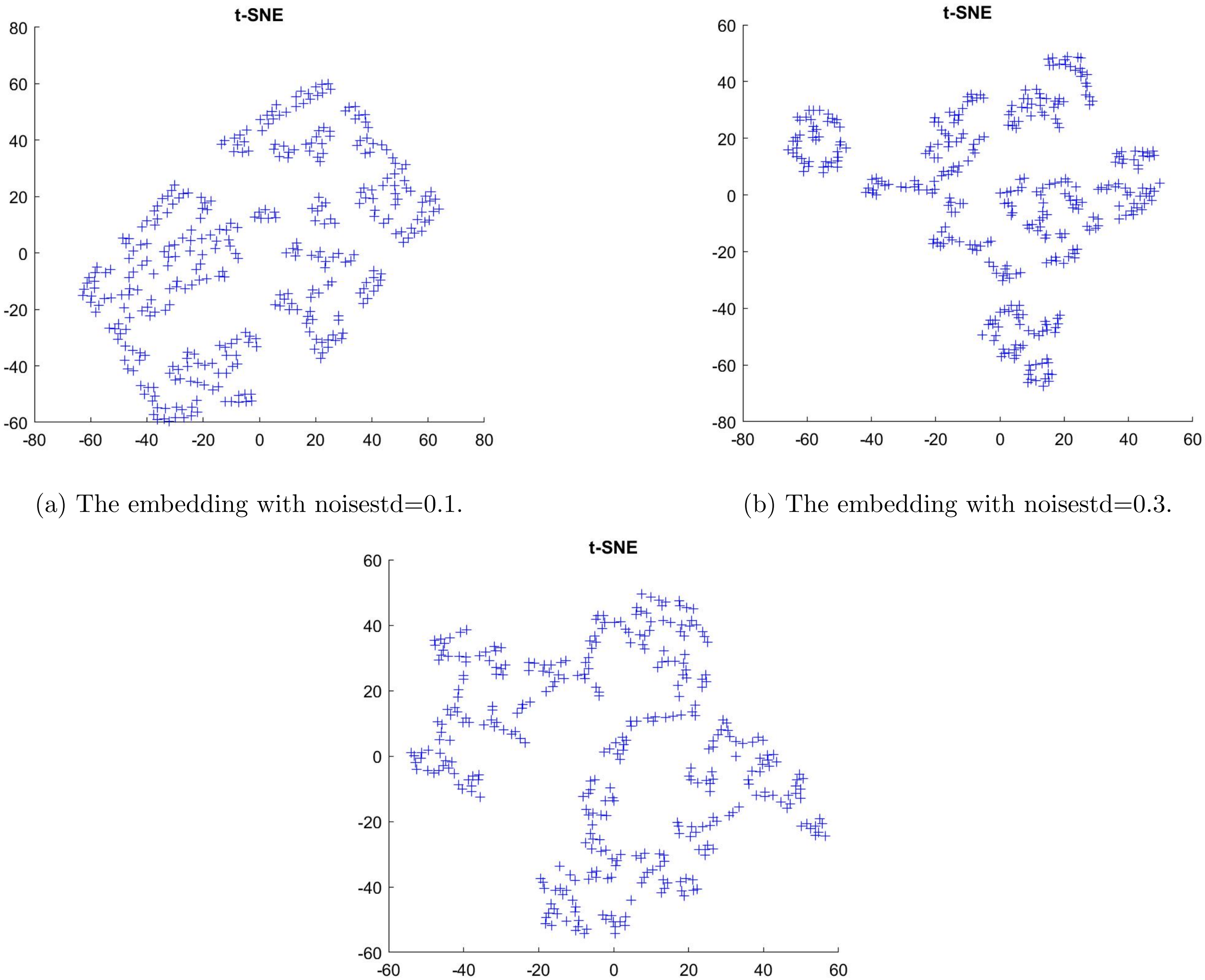
(b) The embedding with noisestd=0.3.



(c) The embedding with noisestd=0.5.

Figure 11: The LLE method applied to noised S-curve data.

- **s-SNE**



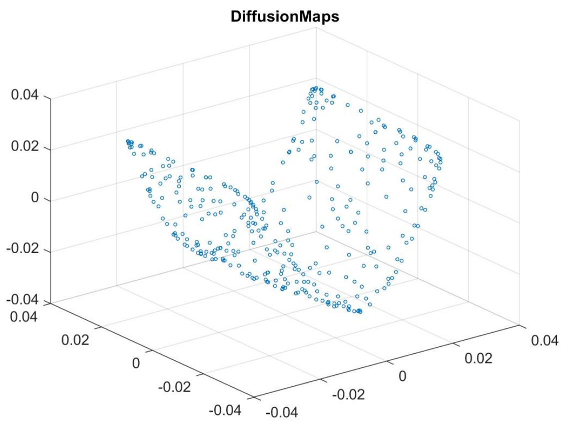
(a) The embedding with noisestd=0.1.

(b) The embedding with noisestd=0.3.

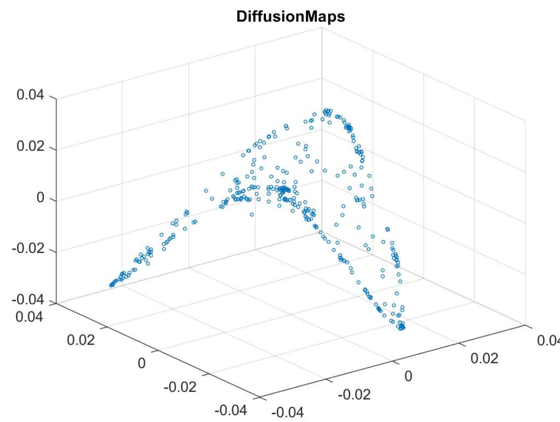
(c) The embedding with noisestd=0.5.

Figure 12: The t-SNE method applied to noised S-curve data.

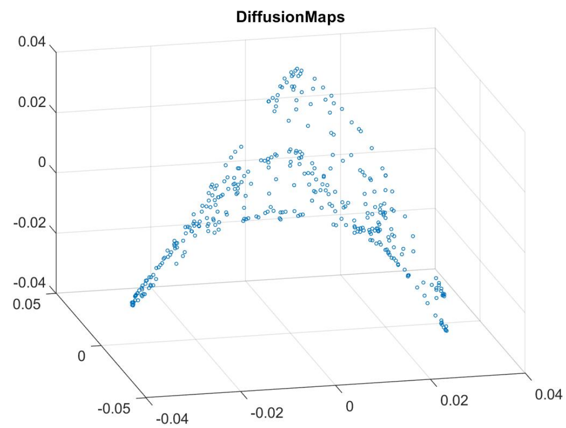
- **DIffusion maps**



(a) The embedding with noisestd=0.1.



(b) The embedding with noisestd=0.3.



(c) The embedding with noisestd=0.5.

Figure 13: The diffusion map method applied to noised S-curve data.

(c) **Dataset 3** In this problem, the facedata is used to test the performance of each method. The results are shown from figure 14 to 16. For the isomap and LLE method, they do not work on this dataset. For the diffusion map method, $\epsilon = 150$ and $\delta = 0.2$

- **PCA**

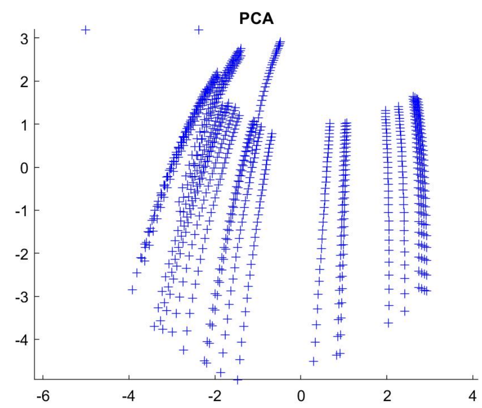


Figure 14: The PCA method applied to Facedata

- **Isomap**
- **LLE**
- **t-SNE**

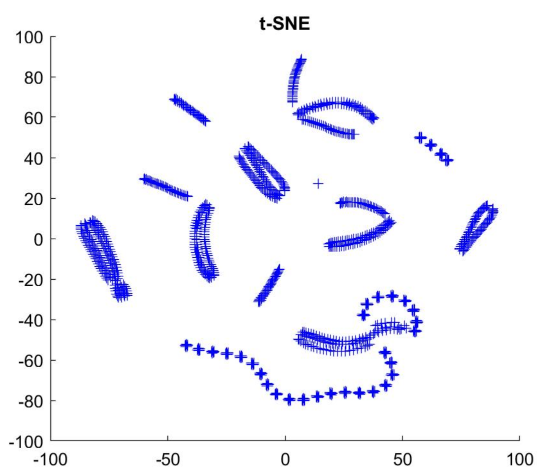


Figure 15: The t-SNE method applied to Facedata

- **DIffusion maps**

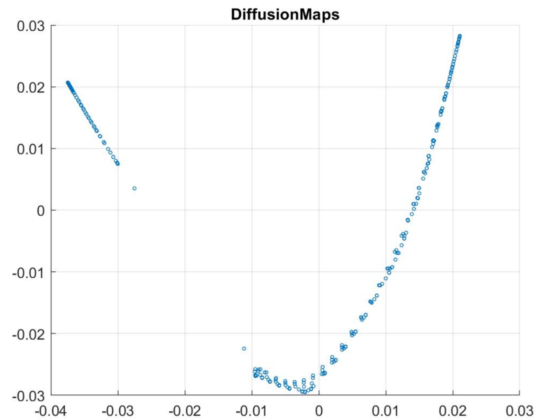


Figure 16: The t-SNE method applied to Facedata