CSIC 5011 Final Project
Manifold Learning on MNIST Dataset

Xiaowen FU, Jiayi LI

1. Introduction

In this project, we explores different manifold learning methods on the MINST dataset, including PCA, Isomap, LLE, MLE, LTSA, MDS, t-SNE, Random forest, Spectral, and then compares the results of dimension reduction to distinguish hand written digits from 0 to 9.

2. MNIST Dataset

The MNIST dataset has 60000 examples of handwritten digits from 0-9 with different hand-writing style. A sample is shown as below:

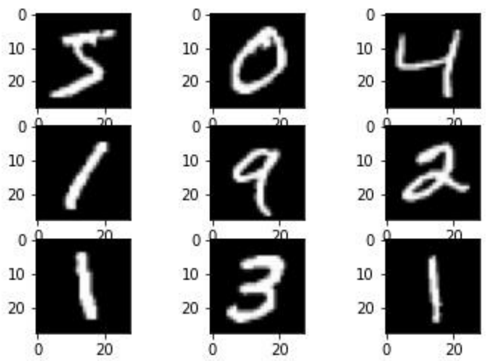


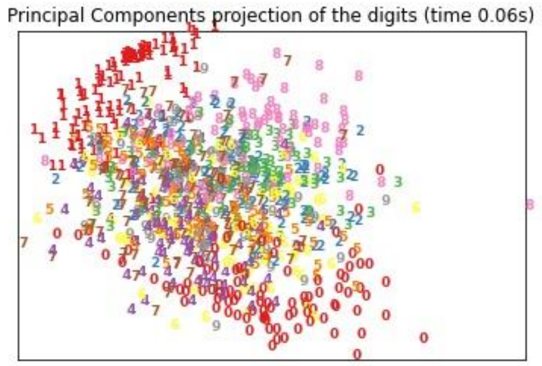
Figure 1: a sample of MNIST dataset

3. Methodology

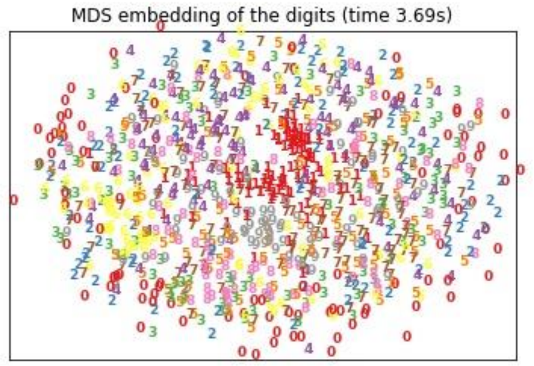
To analysis the high dimensional MNIST data, the following methods are applied for dimension reduction and classification.

- **Principal Component Analysis (PCA):** It is defined as an orthogonal linear transformation by projection such that the greatest variance lies on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on[1]
- **Multidimensional scaling (MDS):** It is a visual representation of distances or dissimilarities between sets of objects[2]
- **Isomap:** It extends MDS by incorporating the geodesic distances imposed by a weighted graph.
- **Locally Linear Embedding (LLE):** It take advantage of sparse matrix algorithms for dimension reduction[3]
- . **Modified Locally Linear Embedding (MLE):** It is a LLE variant which uses multiple weights in each neighborhood to address the local weight matrix conditioning problem which leads to distortions in LLE maps[4]
- **Local Tangent Space Alignment (LTSA):** It is based on the intuition that when a manifold is correctly unfolded, all of the tangent hyperplanes to the manifold will become aligned [5]
- **t-Distributed Stochastic Neighbor Embedding (t-SNE):** It computes the probability that pairs of are related, and then chooses low-dimensional embeddings which produce a similar distribution[6]
- **Random forest:** It constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean/average prediction of the individual trees[7]
- **Spectra clustering:** It make use of the spectrum of the similarity matrix of the data to perform dimensionality reduction before clustering in fewer dimensions.

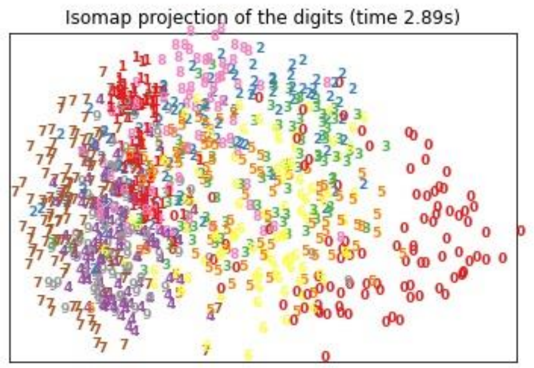
4. Dimension Reduction Result



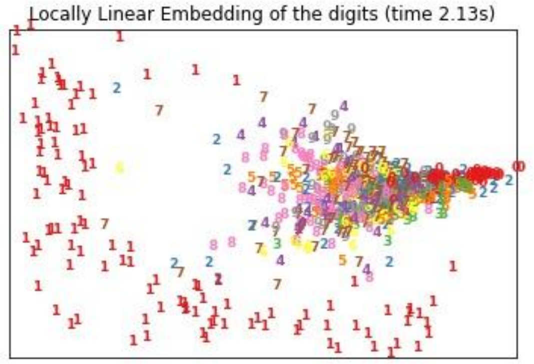
1. PCA



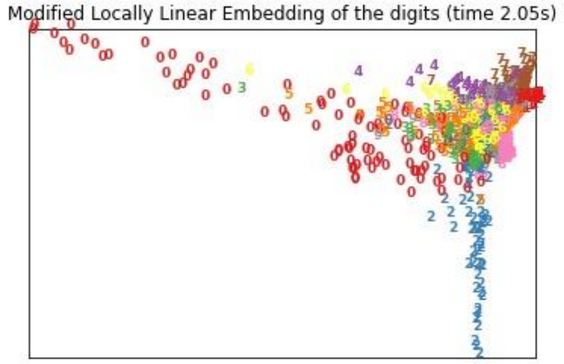
2. MDS



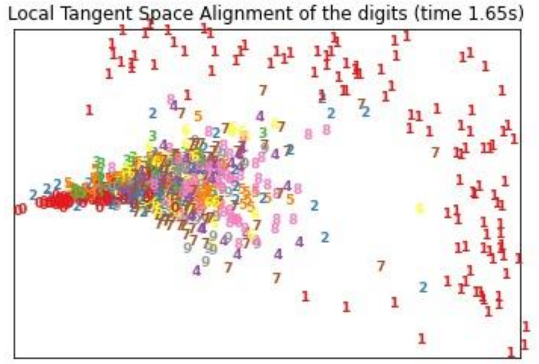
3. Isomap



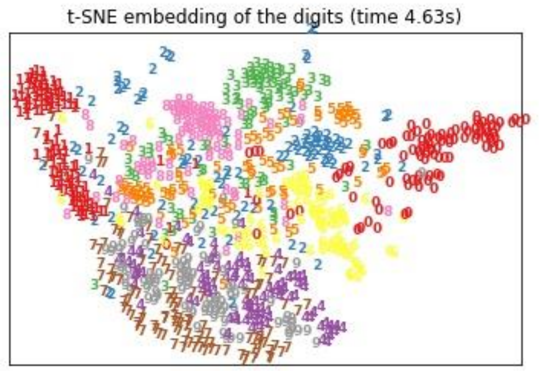
4. LLE



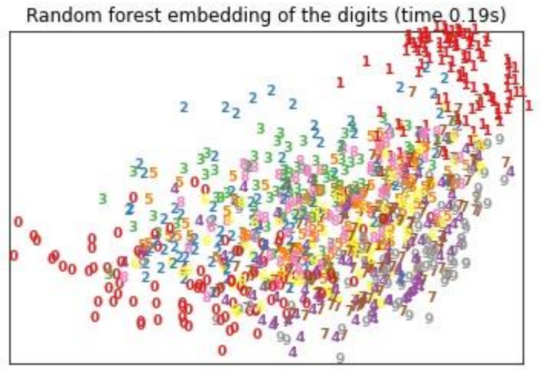
5. MLE



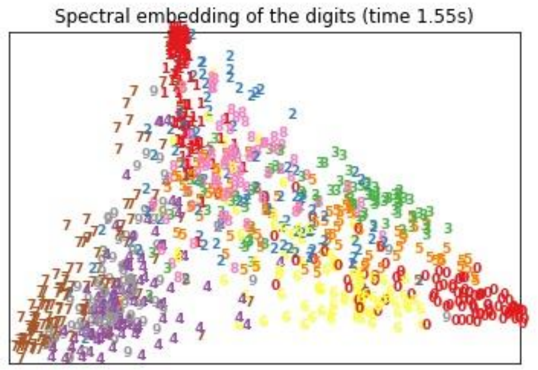
6. LTSA



6. t-SNE



7. Random forest



9. Spectral

Figure 2: dimension reduction results of manifold learning

6. Result Analysis and Conclusion

As shown in Figure 2, we can know that : PCA performs well on clustering but performs badly on distinguishing after projection. MDS turns to show bad results, with only locally classification results. Isomap improves MDS by incorporating the geodesic distances imposed by a weighted graph, and it shows improved results based on MDS. LLE shows an unsatisfying results, where the projection of digit 1 scatters around, and the projection of other digits mixes and squeezes together. MLE improves the result of LLE, where there are distinguishable districts for different digits, although the districts for digit 1 and digits 3-9. LSTA shows a bad result similar to that of LLE. t-SNE and Spectral clustering shows good results, where the districts for different digits are distinguishable enough, while random forest shows a bad result similar to that of PCA.

In conclusion, among all the manifold learning methods for MNIST dataset, it turns out that Isomap, MLE, t-SNE and spectral clustering shows relatively satisfying dimension reduction results, among which t-SNE and spectral clustering performs best due to the structure of MNIST data.

7. Reference

[1] Jolliffe, I. T. (2002). Principal Component Analysis. Springer Series in Statistics. New York: Springer-Verlag. doi:10.1007/b98835

[2] Buja A , Swayne D F , Littman M L , et al. Data Visualization With Multidimensional Scaling[J]. Journal of Computational & Graphical Statistics, 2008, 17(2):444-472.

[3] Roweis S , Saul L . Nonlinear Dimensionality Reduction by Locally Linear Embedding[J]. Science, 2000, 290(5500):2323-2326.

[4] Zhang Z , Jing W . MLLE: Modified Locally Linear Embedding Using Multiple Weights[C]// International Conference on Neural Information Processing Systems. MIT Press, 2006.

[5] Zhang Z , Zha H . Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment[J]. Journal of Shanghai University, 2004, 8(4):406-424.

[6] Maaten L J P V , Hinton G E . Visualizing High-Dimensional Data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9:2579-2605.

[7] Ho T K . The random subspace method for constructing decision forests[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(8):832-844.

8. Contribution

- **Xiaowen Fu:** parts for MDS, t-SNE, Random forest, spectral clusting; poster
- **Jiayi Li**: parts for PCA, Isomap, LLE, MLE, LSTA; video presentation