The University of Texas at Austin
Department of Computer Science
College of Natural Sciences

Diffusion Maps

SDS 384 Scientific Machine Learning, Spring 2022

Chris Lawson, Jeffrey Gordon
Department of Computer Science, The University of Texas at Austin

Talk Points

- Dimensionality Reduction
- Common Reduction Techniques
- Diffusion Maps
  - How do they differ
  - How they work
- Comparing Reduction Algorithms
- Conclusions

Motivation - Need for Dimension Reduction

- Simple problems can easily explode in the number of features -> Curse of dimensionality (Images Case)

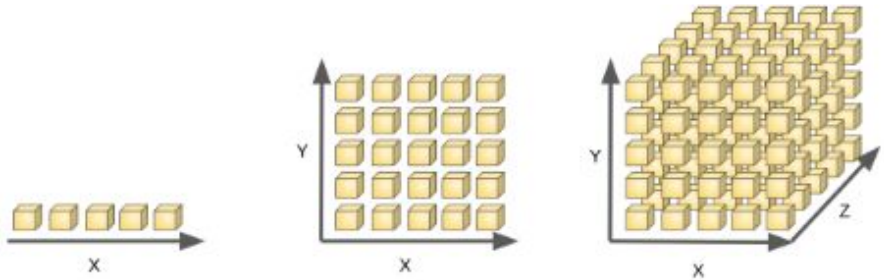
- Extra features obscure or obfuscate the truly informative features

- Local Feature Space -> Global Feature Space

- Goal: Find underlying low dimensional structure, for global similarity

Figure 1: Two images of the same digit at different rotation angles.

https://inside.mines.edu/~whereman/talks/delaPorte-Herbst-Hereman-vanderWalt-DiffusionMaps-PRASA2008.pdf



https://www.i2tutorials.com/what-do-you-mean-by-curse-of-dimensionality-what-are-the-different-ways-to-deal-with-it/

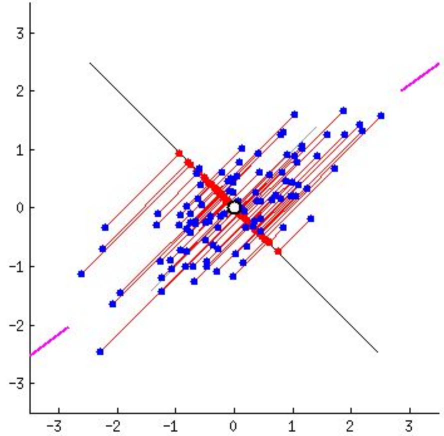
Common Dimensionality Techniques

Principal Component Analysis

- Linear reduction technique
- Principal Components -> Eigenvectors
- Focuses on capturing variability in the data

Draw backs:

- Cannot transform non-linear, real world data very well



https://sagarsaha455.medium.com/pca-for-visualization-and-dimension-reduction-14492e2acf2b

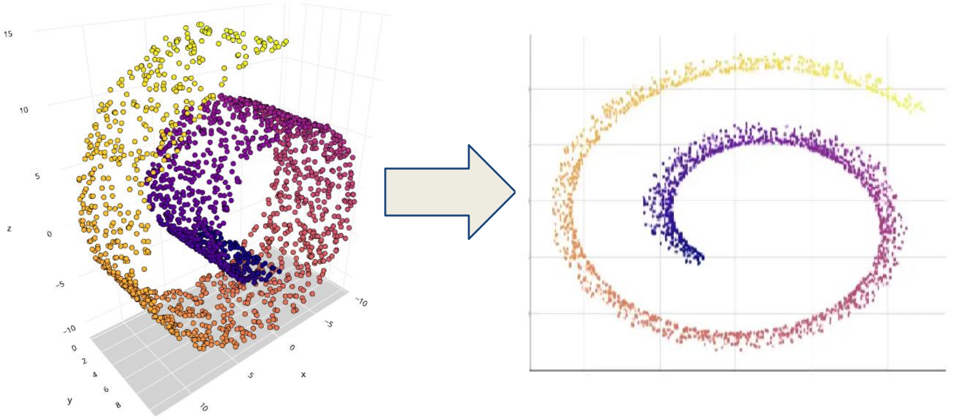
Common Dimensionality Techniques

Multidimensional Scaling (MDS)

- Linear Reduction Technique
- Focuses on preserving pairwise euclidean distances while minimizing strain cost function

Draw backs:

- Dependent on what distance metric you use
- Breaks down with sparse, separated data

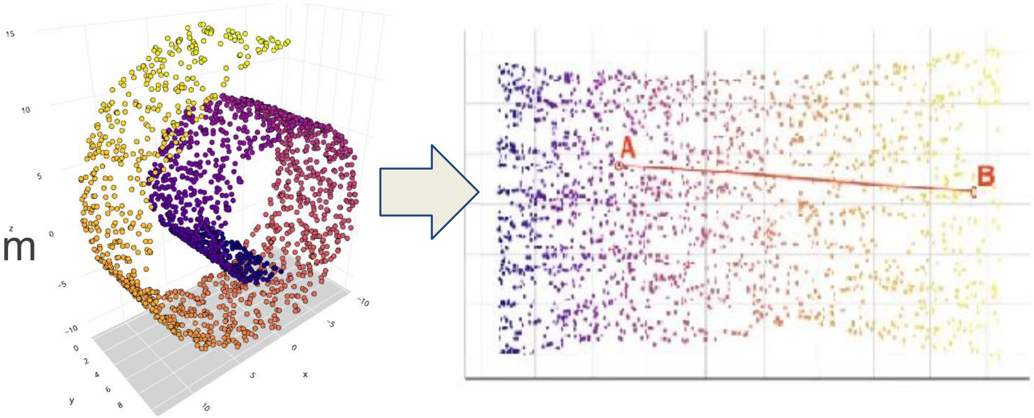


https://towardsdatascience.com/mds-multidimensional-scaling-smart-way-to-reduce-dimensionality-in-python-7c126984e60b

Common Dimensionality Techniques

Isometric Feature Mapping (Isomaps)

- Non-linear reduction technique
- Builds on MDS but focuses on geodesic distance
- Geodesic distance approximated from euclidean using neighborhoods



Draw backs:

- Susceptible to noise in the data
- Need prior knowledge of how to represent the geodesic distance.

https://towardsdatascience.com/isomap-embedding-an-awesome-approach-to-non-linear-dimensionality-reduction-fc7efbca47a0

The University of Texas at Austin
Department of Computer Science
College of Natural Sciences

Diffusion Maps

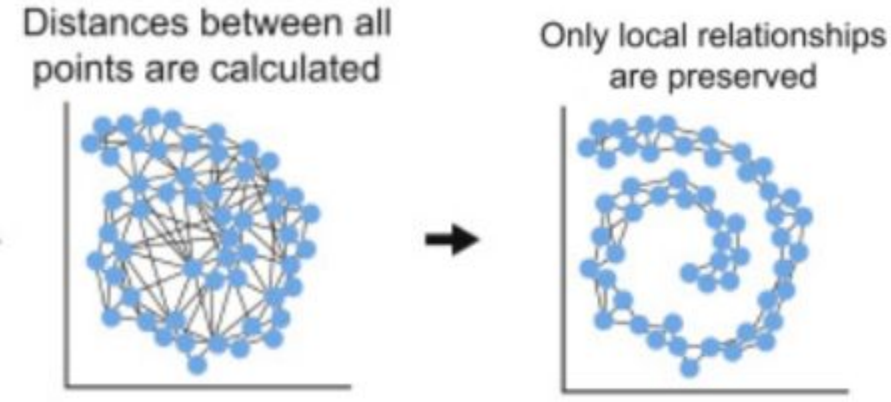
Diffusion Maps

How do they work?

- Consist of 3 parts:
  1) Connectivity

  2) Diffusion

  3) Mapping



Diffusion Maps

Connectivity

- Define a connective measure (i.e. Euclidean distance) between points **if known**

OR

- Define a kernel to help define local measures of similarity

$$k(x, y) = \exp\left(-\frac{|x-y|^2}{\alpha}\right)$$

Create a connectivity matrix P

$$P_{n\times n} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

Where P[i, j] refer to connectivity of X_i and X_j

Diffusion Maps

Diffusion

Define a matrix D^t where t is the time step

$$D^t_{n\times n} = \begin{bmatrix} d^t_{11} & d^t_{12} & \cdots & d^t_{1n} \\ d^t_{21} & d^t_{22} & \cdots & d^t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d^t_{n1} & d^t_{n2} & \cdots & d^t_{nn} \end{bmatrix}$$

Where $d^t_{i,j}$ is: $\sum_{u\in X} |p_t(X_i,u) - p_t(u,X_j)|^2$

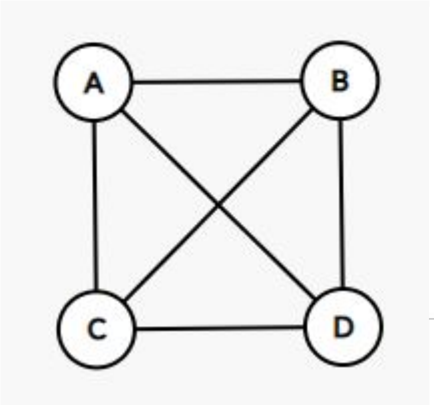
Advantages:

- Helps avoid noise in the data
- With higher values of t, you conform to the geometry of the data.

Diffusion Maps

Example Problem - Probabilistic Jumping

Connectivity



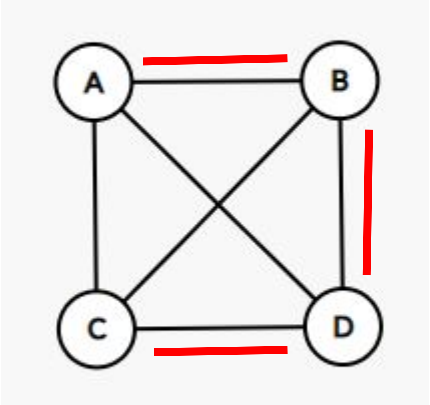
$$P = \begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0.00 & 0.10 & \underline{0.70} & 0.10 \\ \underline{0.70} & 0.00 & 0.10 & 0.20 \\ 0.20 & 0.20 & 0.00 & \underline{0.70} \\ 0.10 & \underline{0.70} & 0.20 & 0.00 \end{bmatrix} \end{array}$$

$$P^3 = \begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0.17 & 0.17 & 0.22 & 0.17 \\ 0.54 & 0.20 & 0.14 & 0.54 \\ 0.17 & 0.17 & 0.16 & 0.17 \\ 0.12 & 0.46 & 0.48 & 0.12 \end{bmatrix} \end{array}$$

Diffusion Maps

Example Problem - Probabilistic Jumping

Diffusion



$$P^3 = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{cccc} A & B & C & D \\ \end{array}$$

$$P^3 = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0.17 & 0.17 & 0.22 & 0.17 \\ B & 0.54 & 0.20 & 0.14 & 0.54 \\ C & 0.17 & 0.17 & 0.16 & 0.17 \\ D & 0.12 & 0.46 & 0.48 & 0.12 \end{array}$$

$$D_3 = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0.14 & 0.08 & 0.10 & 0.14 \\ B & \underline{0.43} & 0.14 & 0.11 & \underline{0.43} \\ C & 0.14 & 0.09 & 0.10 & 0.14 \\ D & 0.11 & \underline{0.29} & \underline{0.35} & 0.11 \end{array}$$

$$\sum_{u \in X} |p_t(X_i, u) - p_t(u, X_j)|^2$$

Diffusion Maps

Mapping

- Where the actual reduction occurs

- Creates the following relation:

  Diffusion Distance in Original Space

  Euclidean Distance in the Mapped Space

And Establish them into a mapping matrix:

$$Y_i := \begin{bmatrix} p_t(X_i, X_1) \\ p_t(X_i, X_2) \\ \vdots \\ p_t(X_i, X_N) \end{bmatrix} = P_{i*}^T.$$

Extract the *m* dominant eigenvalues and vectors from the equation:

$$Y_i' = \begin{bmatrix} \lambda_1^t \psi_1(i) \\ \lambda_2^t \psi_2(i) \\ \vdots \\ \lambda_n^t \psi_n(i) \end{bmatrix}$$

The University of Texas at Austin
Department of Computer Science
College of Natural Sciences

Side by Side Experiments

Presenting the Data

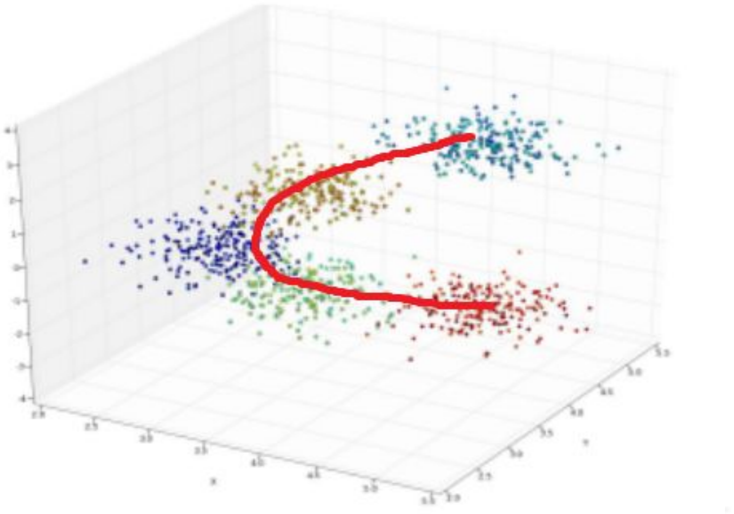
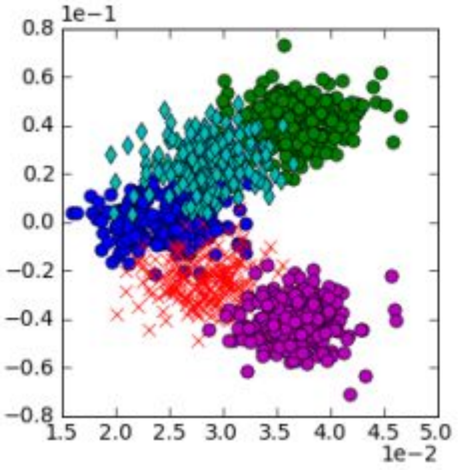


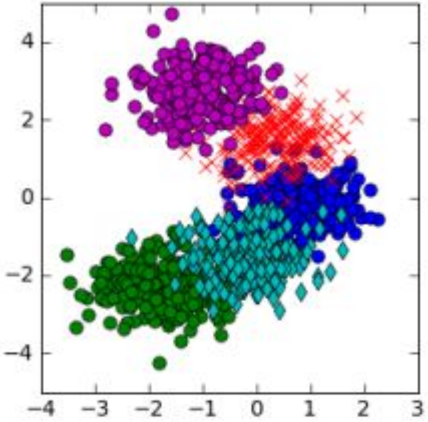
Figure 5: Original Data

PCA, MDS, Isomap Results

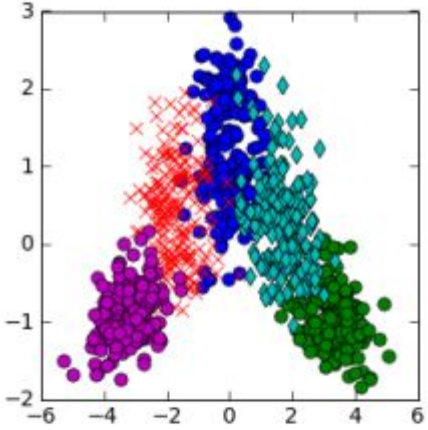
2 Dimensions PCA



2 Dimensions MDS

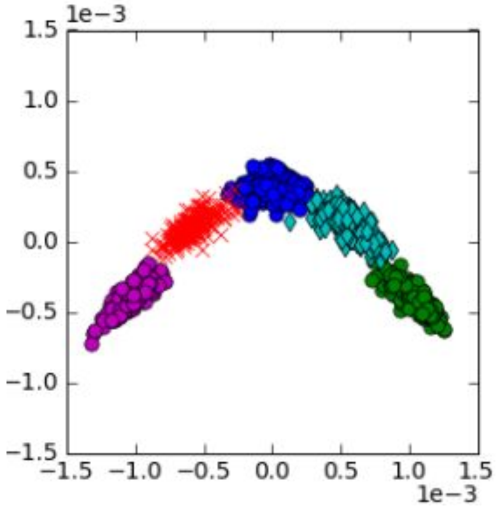


2 Dimensions Isomap

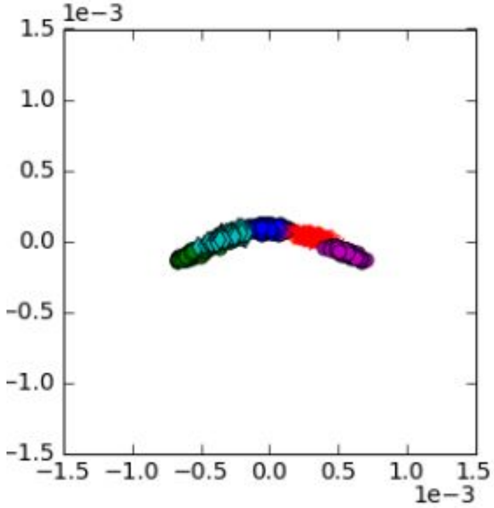


Diffusion Map Results

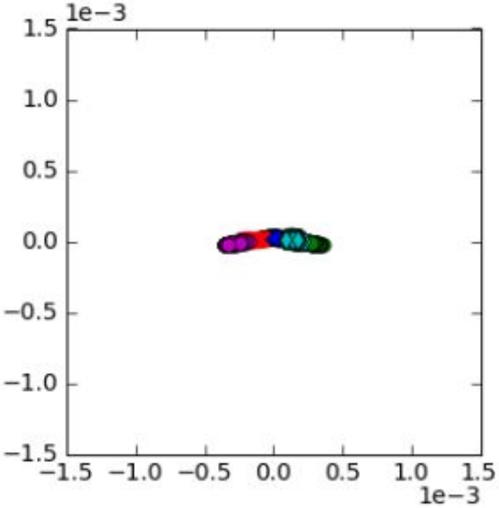
Diffusion t = 1



Diffusion t = 2

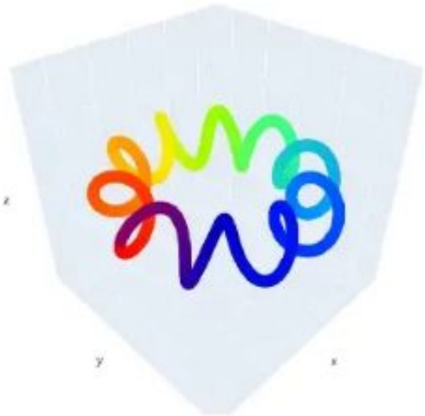


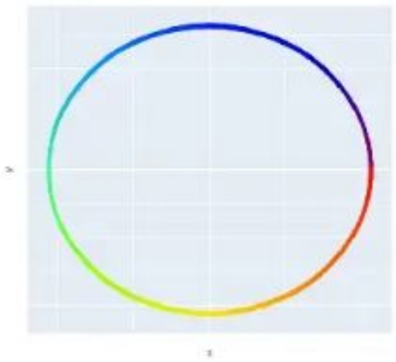
Diffusion t = 3



Takeaways

- Helps to expose the underlying Geometric Structure of the data through time steps
- Diffusion Translation allows for Euclidean measures to not be distorted
- Extremely expensive for the best results





https://towardsdatascience.com/unwrapping-the-swiss-roll-9249301bd6b7

The University of Texas at Austin
Department of Computer Science
College of Natural Sciences

Contributions

- Chris Lawson - 65%
- Slides Presented:
  - 9-17

- Jeffrey Gordon - 35%
- Slides Presented:
  - 1-8, 18