Apply MLT to Model 3D Chromatin Structures

Weiyi Chen, Tianyi Hao

Overview

- T.H.
  - Dimension Reduction (MLT)
  - Example tech: Locally Linear Embedding (LLE)
- W.C.
  - Modified MLT for Structure Prediction
  - Implementation and results
  - Former step and next step (L.W. and Z.W.)

Part 1:Dimension Reduction

Motivations and Techniques

Dimension Reduction

- Manifold learning techniques
  - Embed data from high to low dimensional
  - Goal 1: Reduce Complexity
  - Goal 2: Extract Key Low Dimensional Features
  - Goal 3: Data Visualization

Dimension Reduction

- Possible MLTs
  - Principle Component Analysis (PCA)
    - SVD, high speed and good result
  - Multidimensional Scaling (MDS)
    - Hw2, a bit slower but better
  - Locally Linear Embedding (LLE)
  - Improved LLE

Part 2: Locally Linear Embedding (LLE)

Intuitions and Algorithm

Assumptions

- Data points are sampled from a low dimensional manifold
- Data points can be locally approximately embedded into a plane
- The nearer two data points, the more similar they are

Intuitions

- Every data point can be approximately expressed as an combination of its neighbors
- LLE constructs a neighborhood preserving mapping by
  1. Calculate combinations for each data point
  2. Fit into a low dimensional Euclidean space

LLE Algorithm:

Parameter: K (The number of neighbors per point)

Input: $X_1, \cdots, X_N \in \mathbb{R}^D$

Output: $Y_1, \cdots, Y_N \in \mathbb{R}^d$

0. Compute the nearest K neighbors of each data points X_i, as $\Gamma(i)$.

1. Compute the weighs W_{ij} that best reconstruct each data point X_i from its neighbors, minimizing the cost in $E(W) = \sum_i \left| X_i - \sum_{j \in \Gamma(i)} W_{ij} X_j \right|^2$ subjected to $\sum_{j \in \Gamma(i)} W_{ij} = 1$ for every i.

2. Compute the vectors Y_i best reconstructed by the weights W_{ij}, minimizing the quadratic form $\Phi(W) = \sum_i \left| Y_i - \sum_{j \in \Gamma(i)} W_{ij} Y_j \right|^2$ by its bottom nonzero eigenvectors.

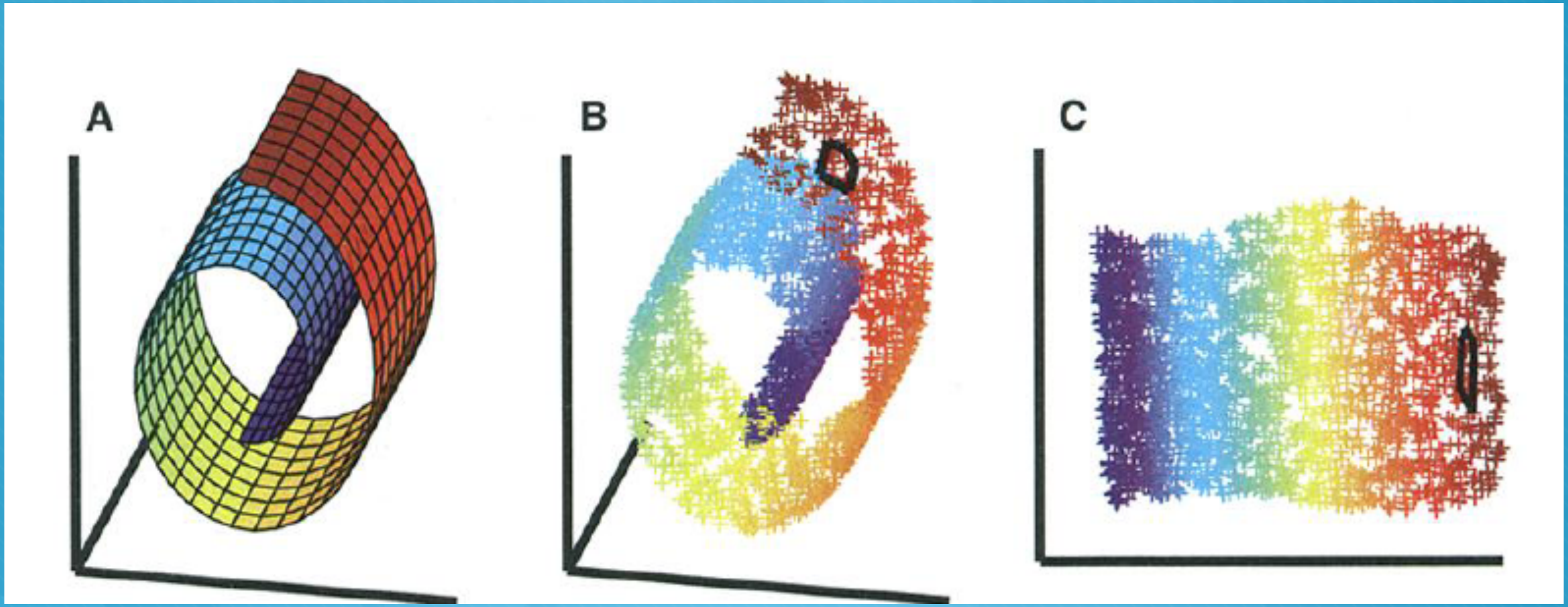


Illustration of LLE [4]

Part 3: Modified LLE for Structure Prediction

Problem Specification and Ideas

Implementations

- Step 1 & 2 are both quadratic programming such that they can be solved efficiently
- Step 2 has a quadratic form,
- Constraint 1:
- Constraint 2:
- Obtained by SVD

Algorithm (Z.W.)

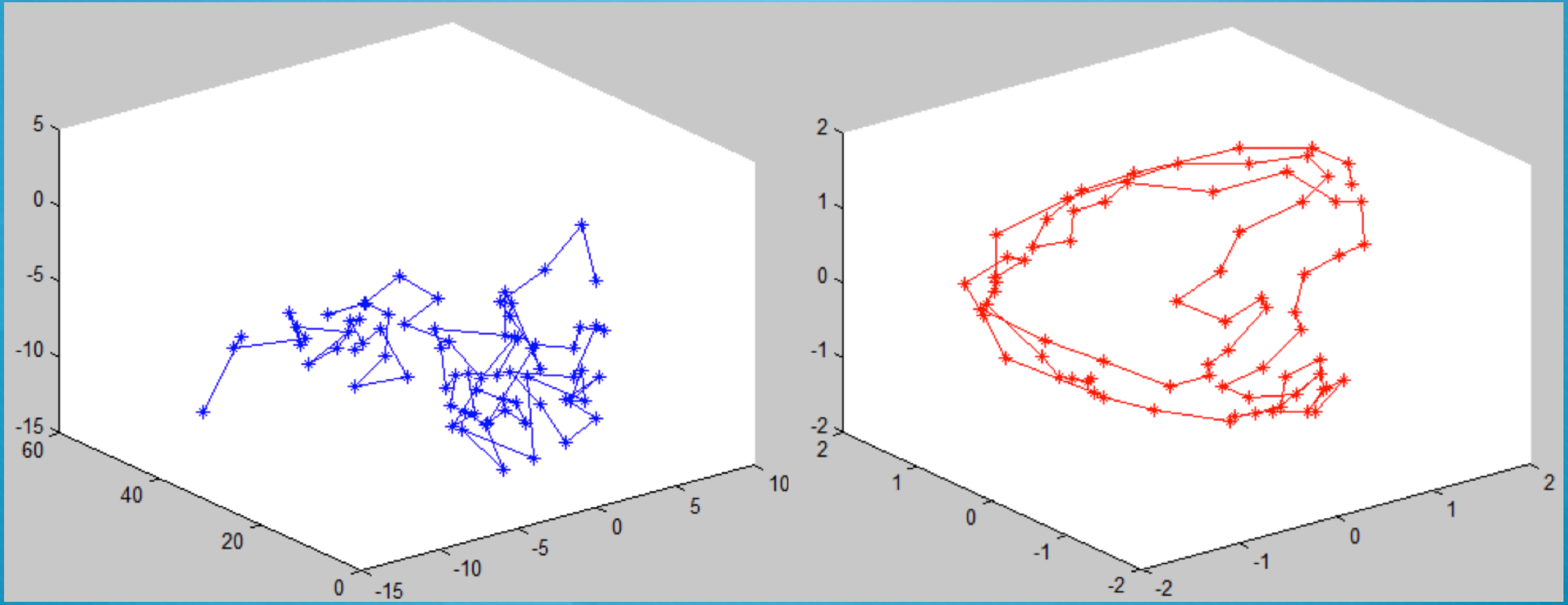
- Search for K nearest nodes
  - By searching the highest K IFs
- Estimate W
  - Naïve way
  - Covariance using distance
- Compute embedding from eigenvectors

Parameters and return

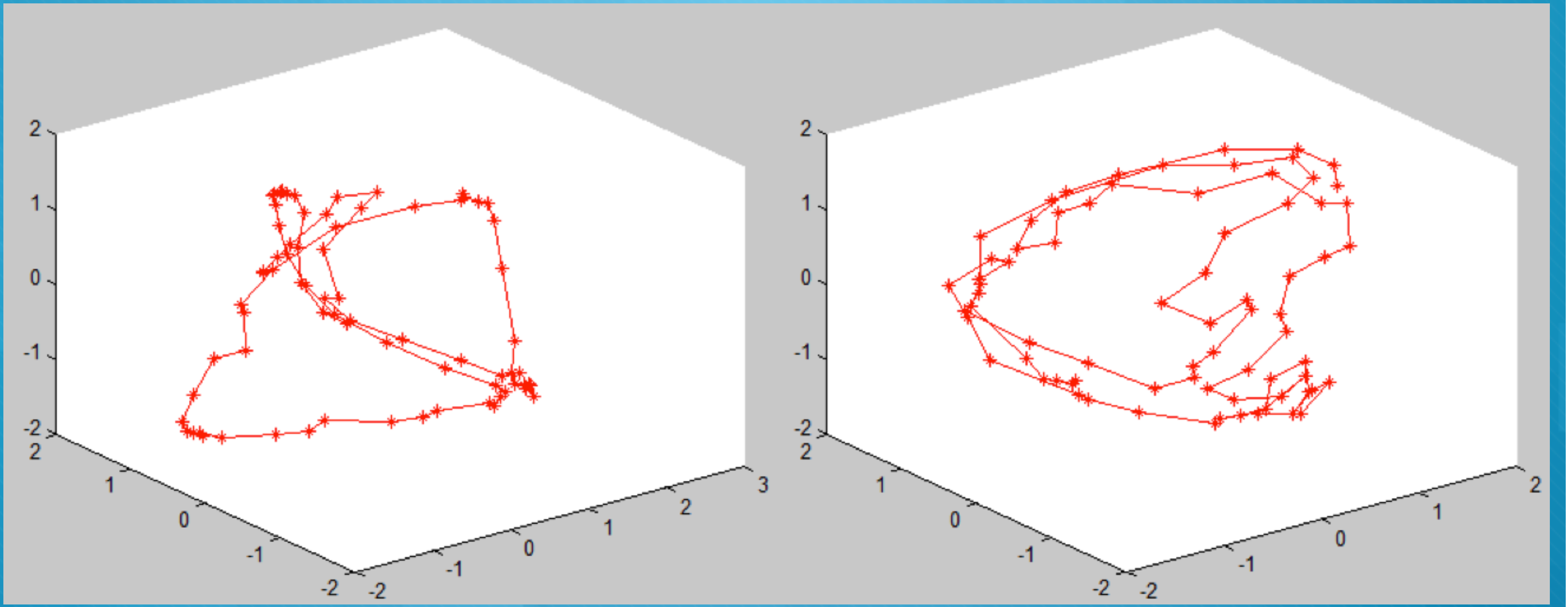
- Input
  - Interaction Frequency information
  - Matrix information
- Output
  - 3 dimensional coordinates for each point

Part 4: Implementation & results

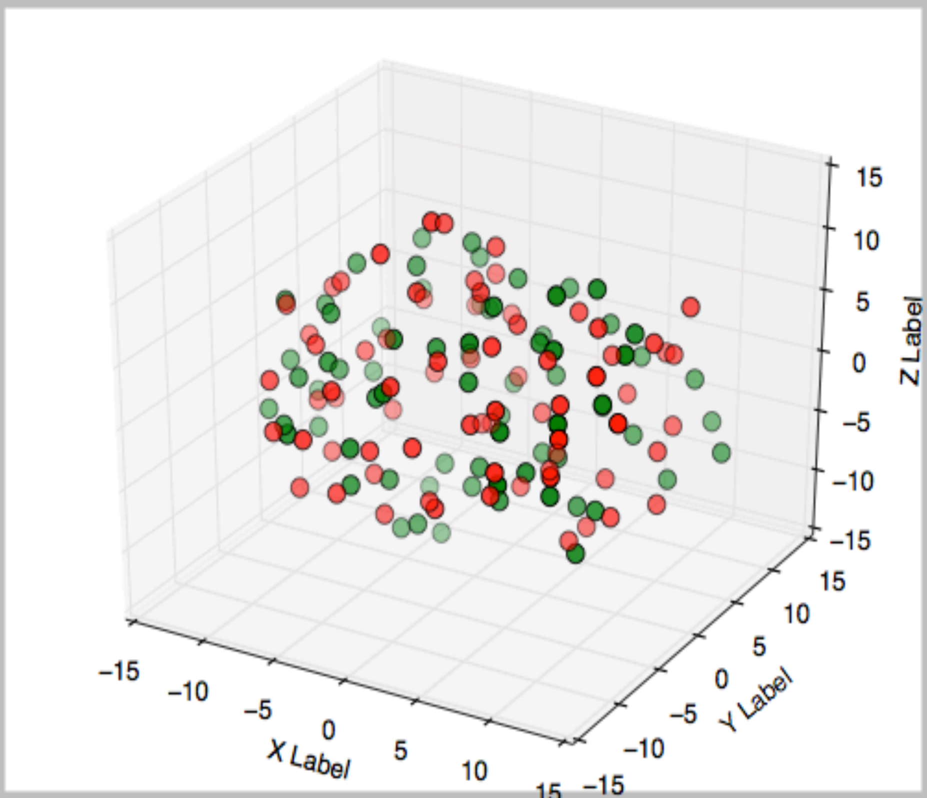
LLE performance

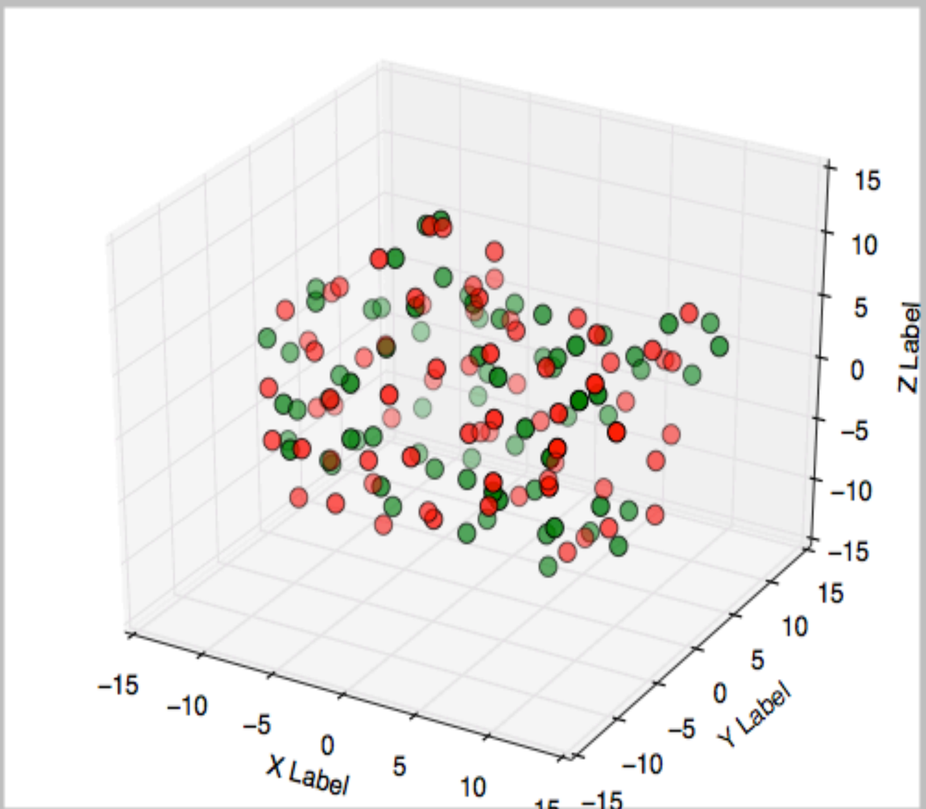


Sensitivity of k

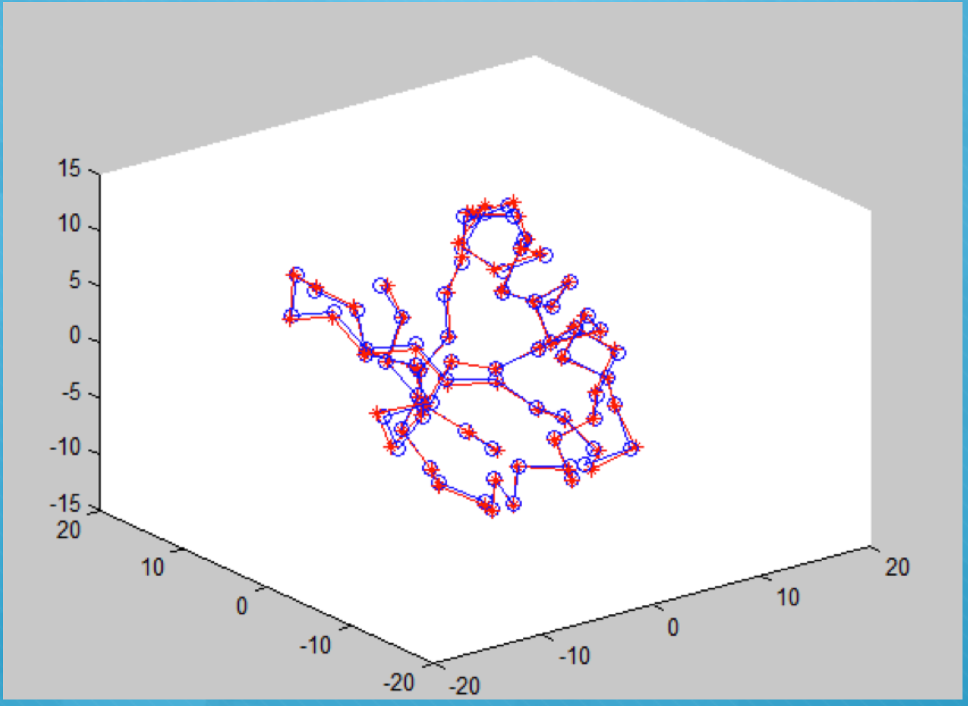


MDS performance

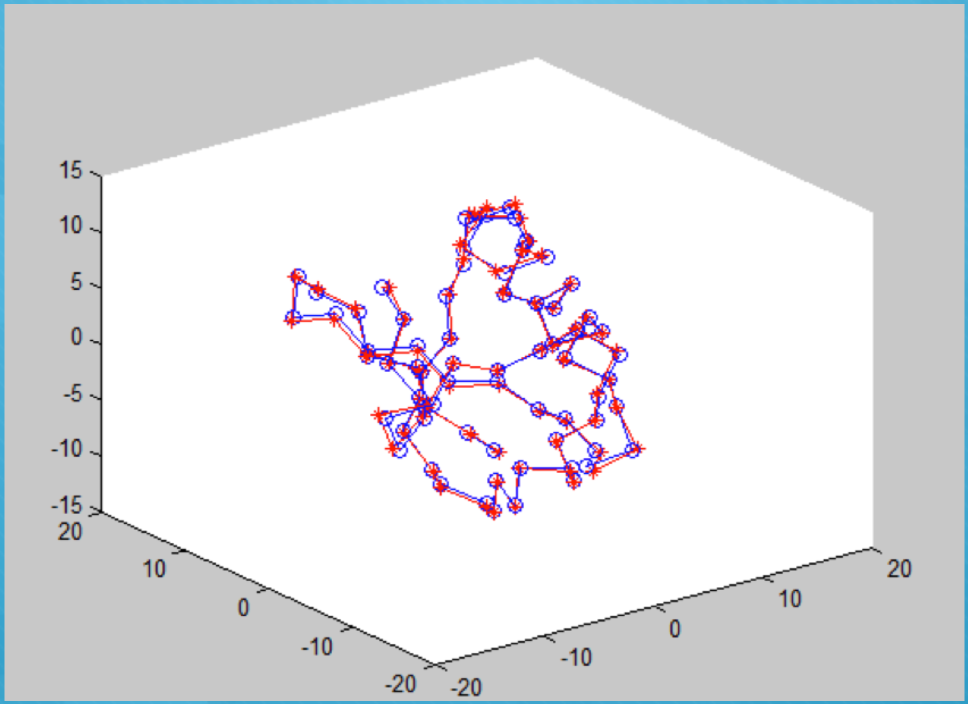




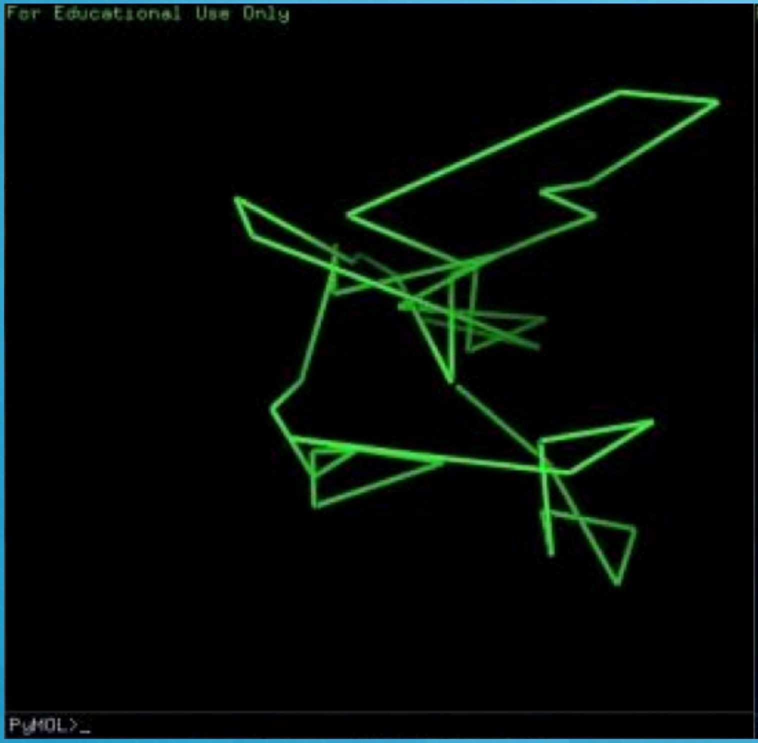
MDS RMSD = 0.01544373

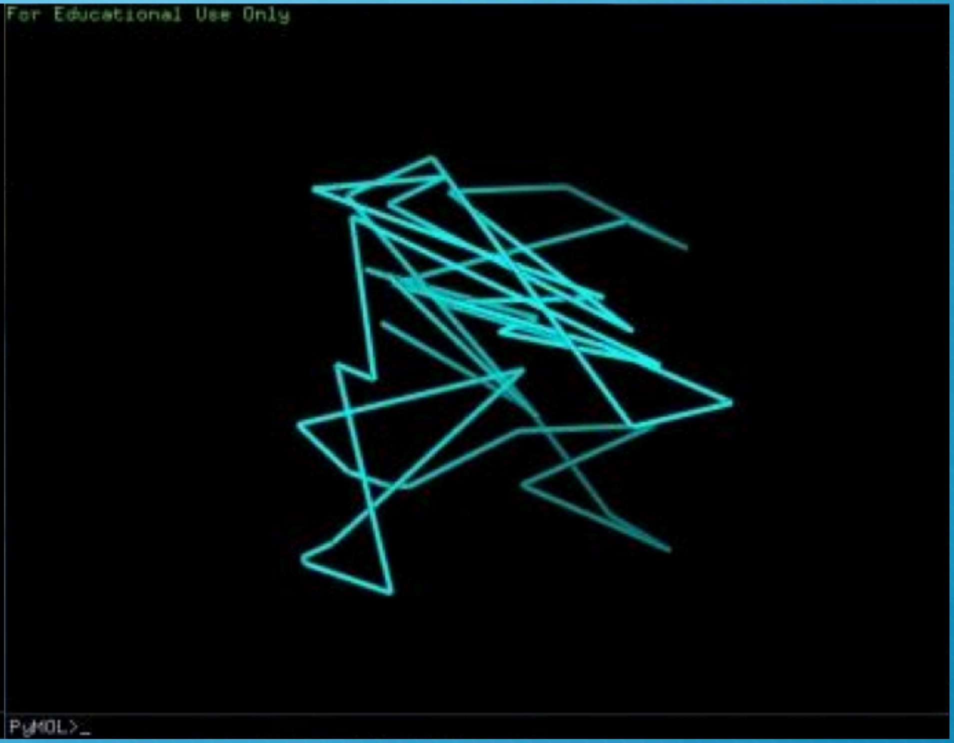


MDS RMSD = 0.01544373



PCA performance



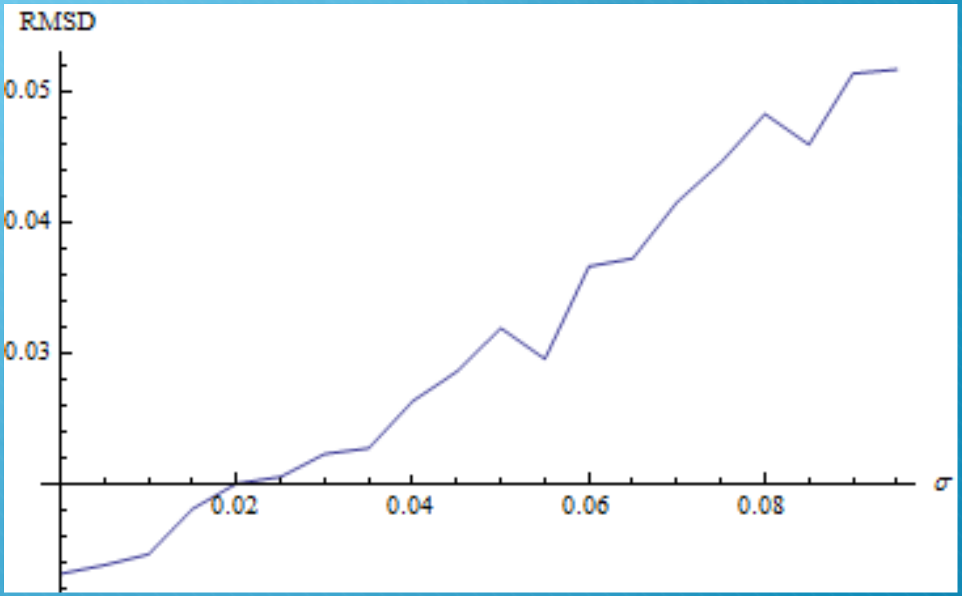


Why? Reason - 1

- The IF information from one to its nearest points is not enough
- need more IF from its neighbor to its neighbor

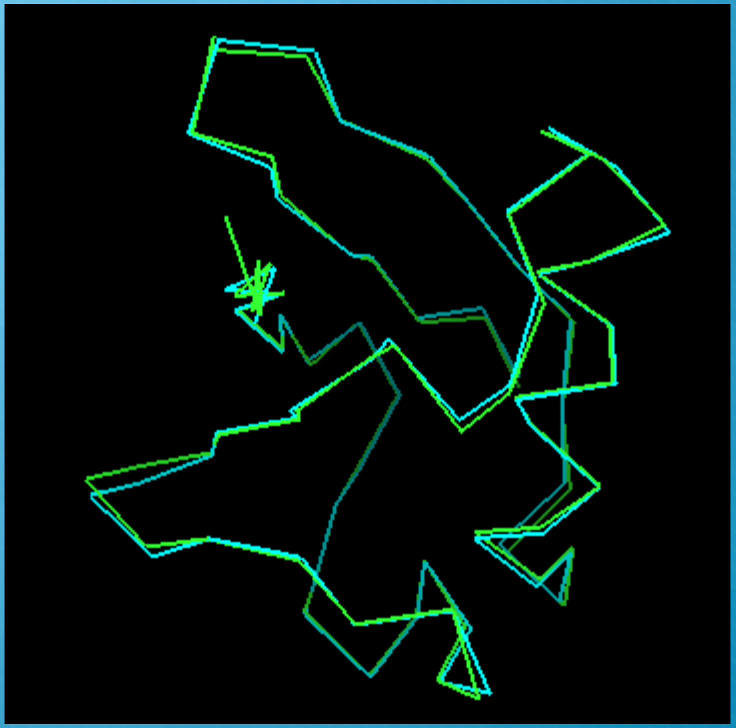
Why? Reason - 2

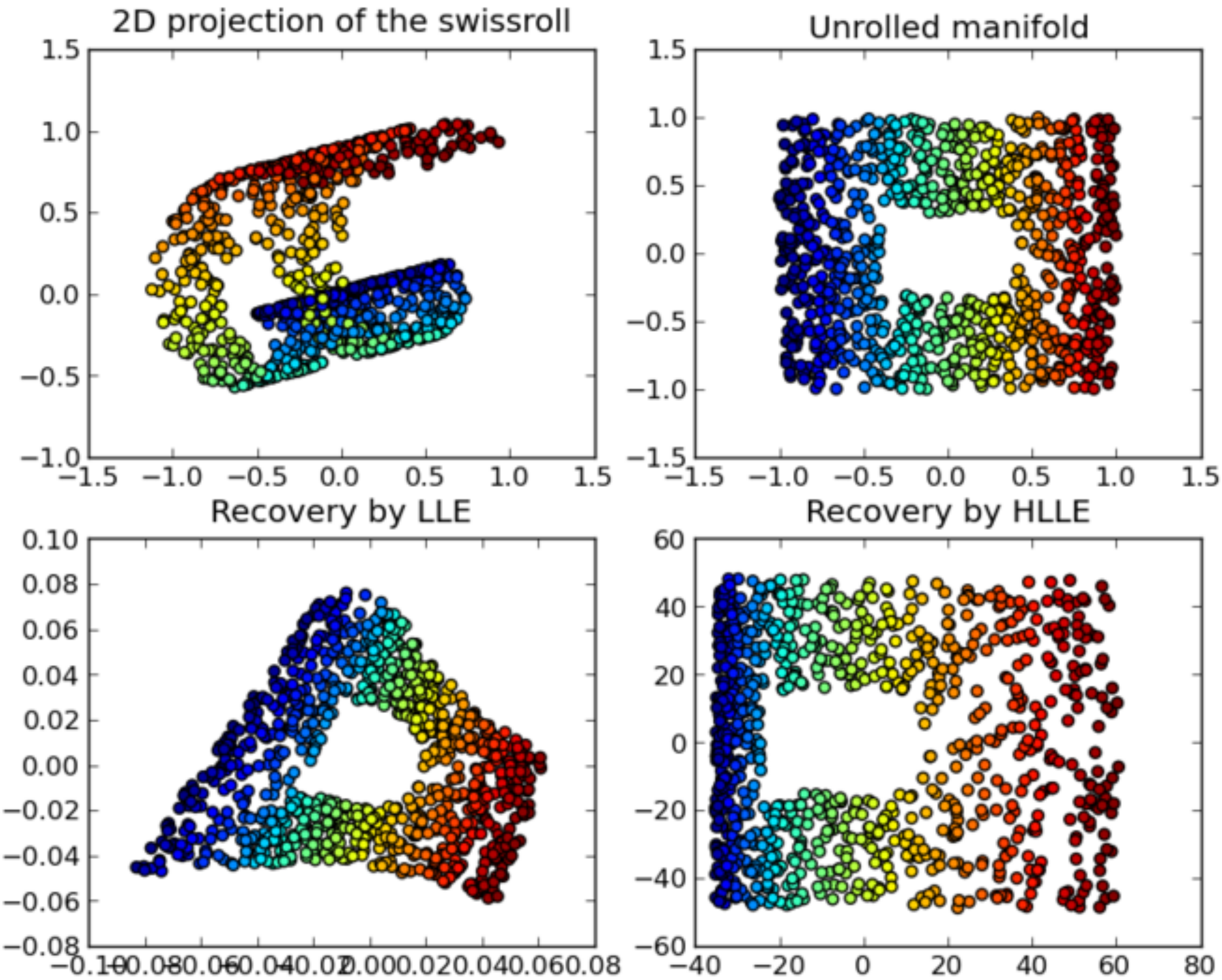
- Random structure is not reasonable
  - W.L's MLT: decrease energy function by random walk
  - Inapplicable cause of complexity



Why? Reason - 3

- Pure LLE doesn't work well, requiring modifications
- Hessian LLE, MLLE




2D projection of the swissroll
Unrolled manifold
Recovery by LLE
Recovery by HLLE

Improvement

- 1. Still have to convert IF to distance
  - Estimate the distance between one's neighbors
- 2. Dataset
  - Modify random structure
  - Search for real data to run more structures
- 3. LLE modification (HLLE works better)

Reference

- [1] An Introduction to Locally Linear Embedding, Lawrence K. Saul, Lawrence K. Saul
- [2] A Survey of Dimension Reduction Techniques, Imola K. Fodor
- [3] Dimension Reduction: A Guided Tour, Christopher J. C. Burges
- [4] Nonlinear Dimensionality Reduction by Locally Linear Embedding Sam T. Roweis and Lawrence K. Saul