Self Organizing Maps and Locally Linear Embedding for Dimensionality Reduction

Vishakh Gopu and Lutz Hamel,Department of Computer Science and Statistics, University of Rhode Island

Introduction

- Can self organizing maps and locally linear embedding find similar lower dimensional structures in high dimensional spaces?
- Interesting because the SOM is a conceptually simpler algorithm
- Convergence has always been an issue, we use a population based convergence method in this investigation and avoid comparisons involving non-converged maps

Background

Self Organizing Maps or SOM :

```
1)Start with a grid of neurons with  vectors of
the same dimensions as the input data.

For every iteration with a decreasing neighborhood
size:
For every observation:
2)Find the neuron most similar to the data point
3) Make the neighborhood around that weighted
vector more like the matched vector
4) Smooth neighborhood of neurons
```

Locally Linear Embedding or LLE :

```
1)Assign neighbors to each data point

2)Compute weights that best linearly reconstruct
each point from its neighbors

3) Compute the low dimensional embedding vectors
best reconstructed by the weights
```

Population based convergence criterion for SOMs

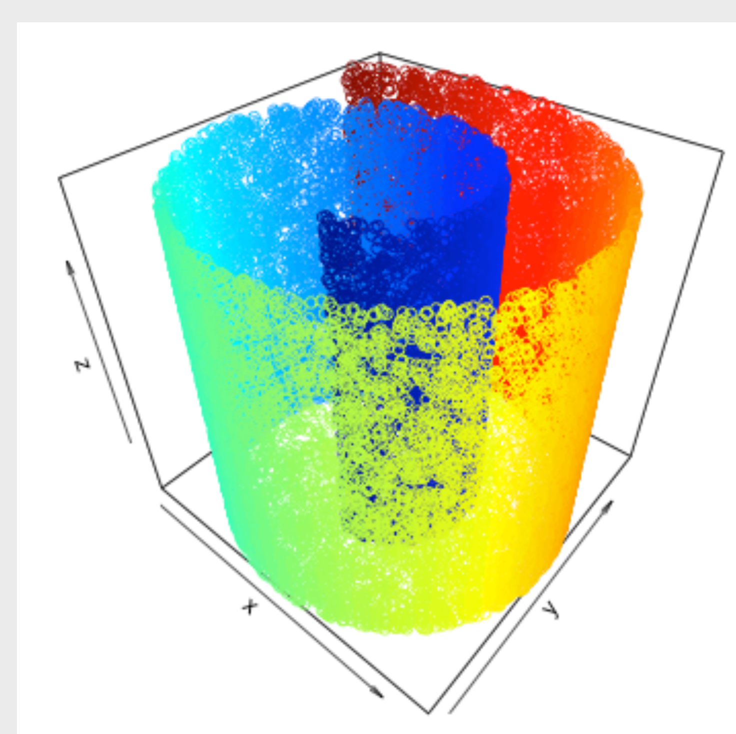
- A way of evaluating convergence
- Data set as one population, SOM as another
- Two sample test to see if they come from the same distribution
- 0-1 rating of convergence

Fundamental Clustering Problem Suite

- Datasets to test clustering and dimensionality reduction
- Each contain a different characteristics to test specific qualities in an algorithm

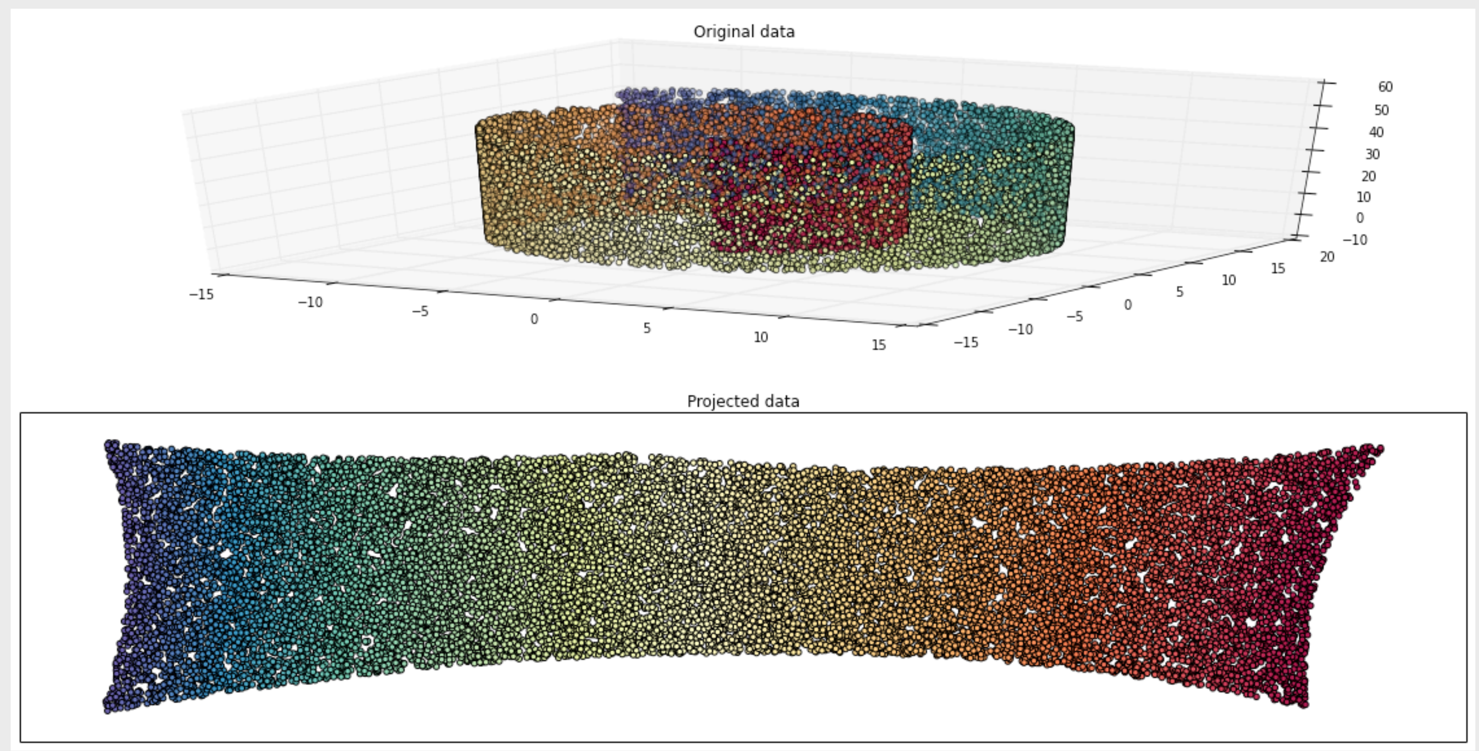
The Swiss Roll Data Set

- A 2-D manifold hidden in a 3-d structure.
- A good test case for dimensionality reduction
- Comes with approximate 2-d positions on manifold



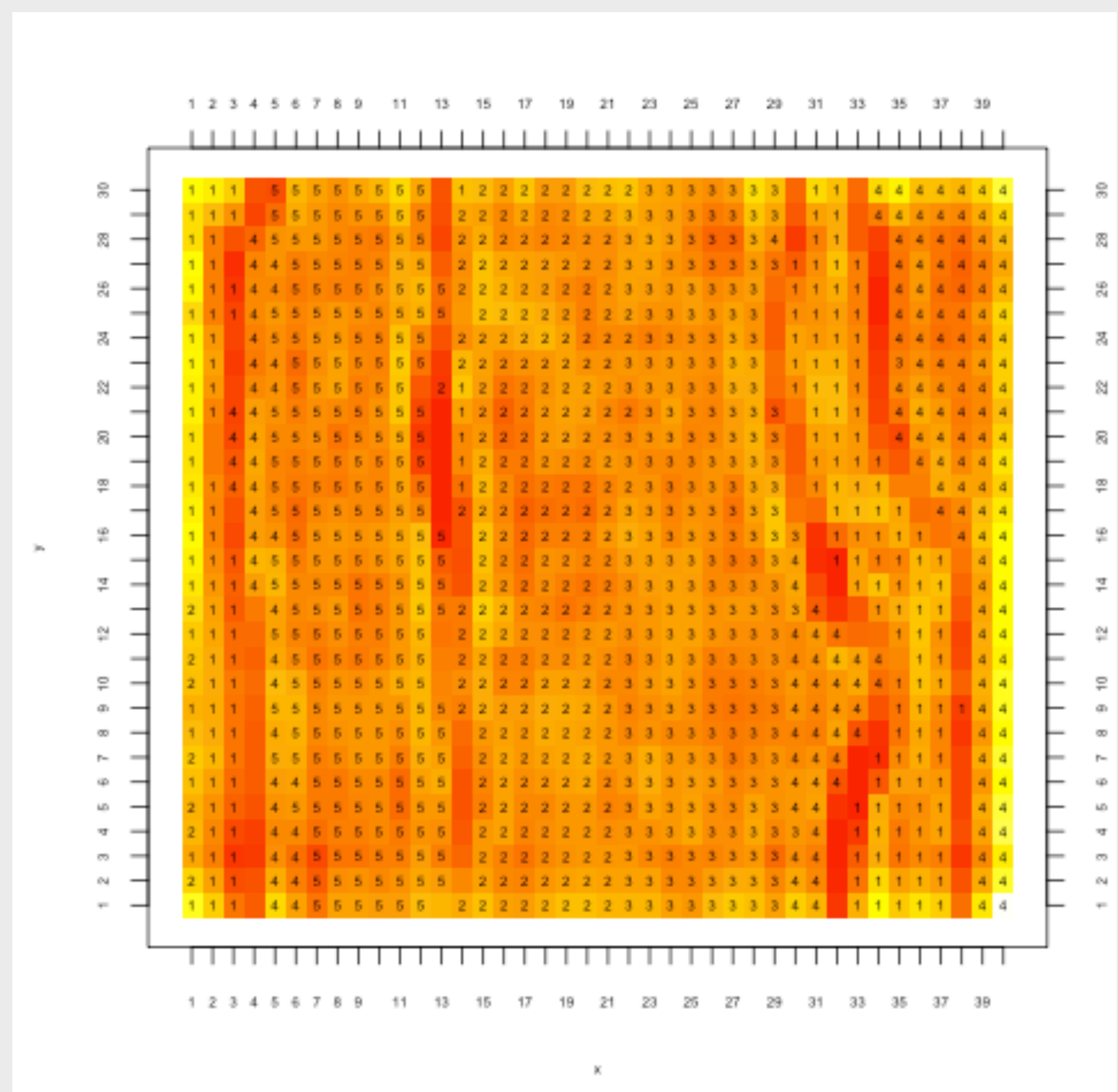
Results

LLE on Swiss Roll



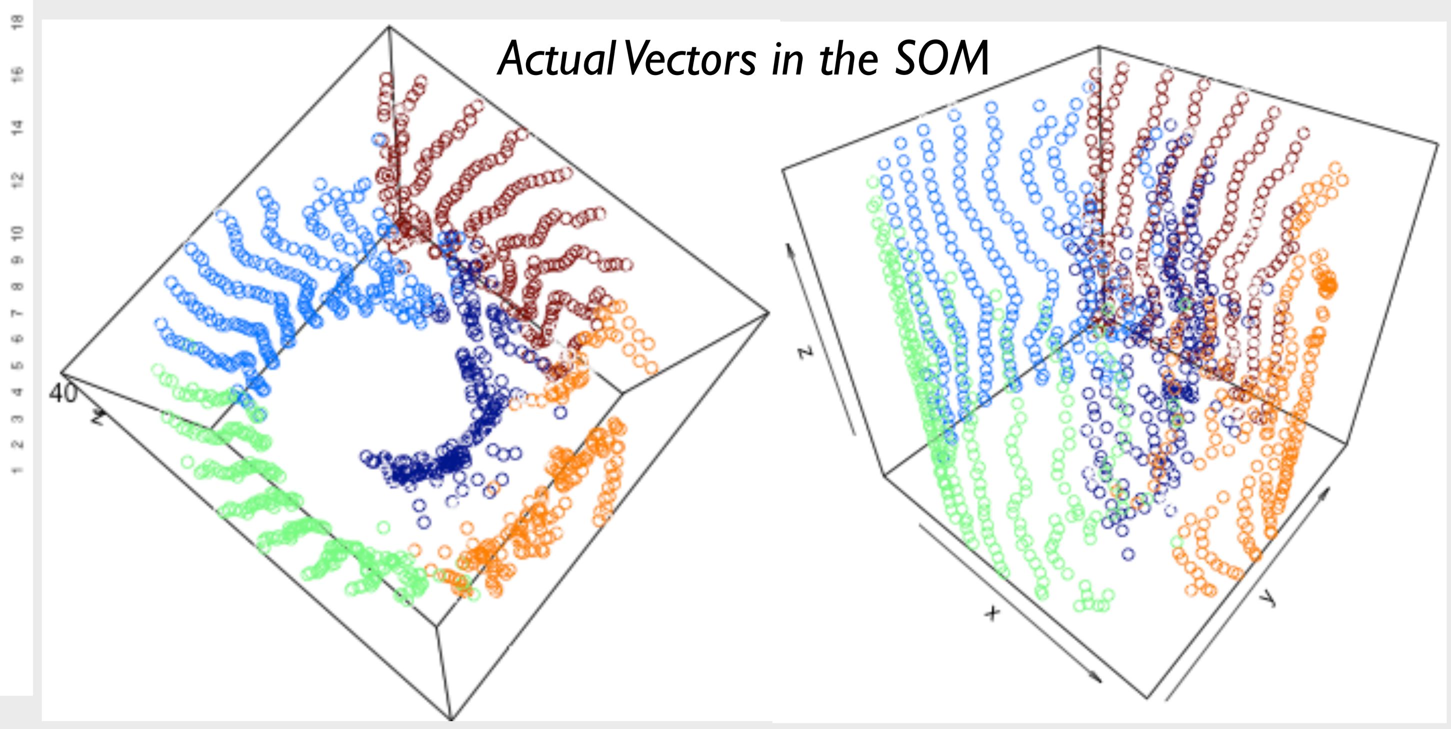
- The Color Map shows local ordering preservation in the LLE, because the projection is a set of 2-D points.

SOM on Swiss Roll



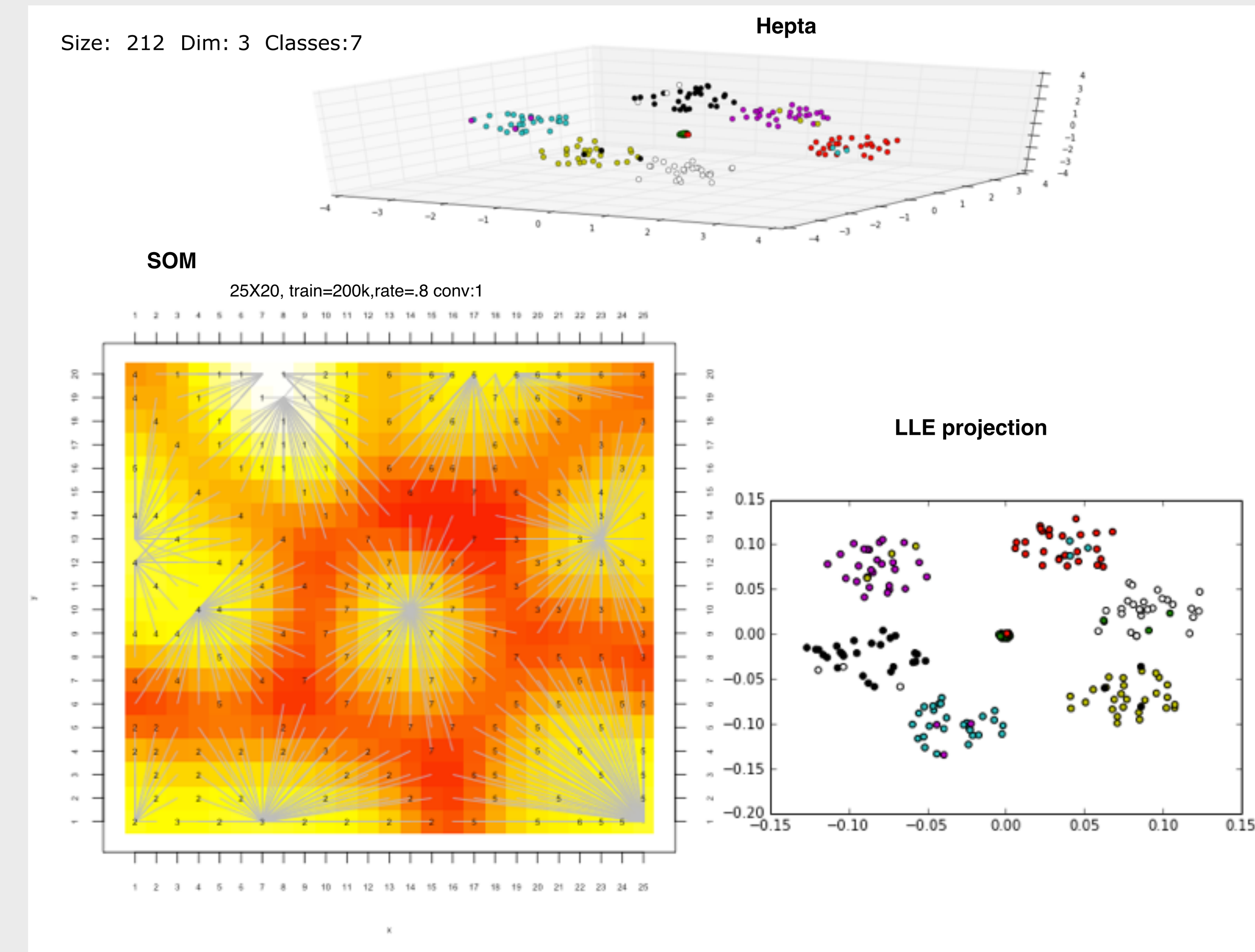
UMAT with labels to show approximate ordering in manifold is preserved

- The striping of the labels is indicative of columns close together in the 3-D data being put close together in the 2-D projection.
- The same striping is shown in the LLE 2 dimensional embedding.
- This was only clear once convergence criteria became 1.



Actual Vectors in the SOM

FCPS Example: Hepta



- They do similarly well on the 3-D FCPS
- Both miss and find the same aspects of the 3-D examples
- It is interesting to see how the visualizations provide different versions of similar information

References

A Population Based Convergence Criterion for Self-Organizing Maps, Lutz Hamel and Benjamin Ott. Proceeding of the 2012 International Conference on Data Mining (DMIN'12), pp98-104, July 16-19, 2012

Kohonen, Teuvo. Self-organizing Maps. Berlin: Springer, 2001. Print.

Roweis, S.T. "Nonlinear Dimensionality Reduction by Locally Linear Embedding." Science 290.5500 (2000): 2323-326. Web.

Tenenbaum, J. B., J. C. Langford, and V. De Silva. Swiss Roll Dataset- A Global Geometric Framework for Nonlinear Dimensionality Reduction. N.d. Raw data. Science, n.p.

Ultsch, A.: Clustering with SOM: U*C, In Proc. Workshop on Self-Organizing Maps, Paris, France, (2005) , pp. 75-82

Methods

- Running the LLE on the Swiss Roll creates 2-D points , so it is easy to check whether the 2-d manifold was found in the 3-d space
- The color map then shows whether local ordering is preserved
- To simulate the same idea on the SOM, we create 5 labels based on each 3-D point's approximate position on the 2-D manifold
- This comes from the accompanying approximate position dataset

Discussion

- LLE and SOM both find the 2-D manifold in the swiss roll
- LLE and SOM both do similarly when faced with the 3-D clustering problems from the FCPS
- SOMs are conceptually simpler but appear to miss some aspects of Swiss Roll
- The results were only clear with a convergence score of 1.