Dimensionality Reduction Tutorial
Part 2: Nonlinear methods (MDS and tSNE)

Gabrielle M. Schroeder

CNNP Journal Club

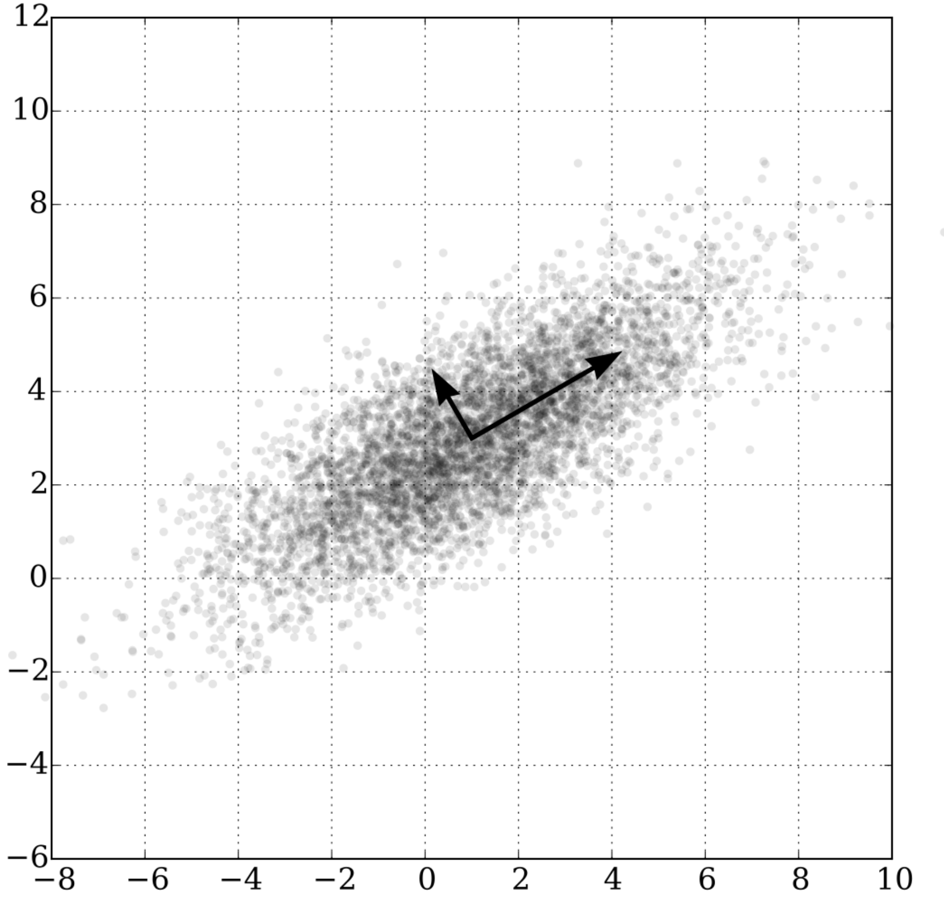
24 April 2019

Dimensionality reduction

Transforming **high dimensional data** with many features into a **lower dimensional space**

Examples of high dimensional data:

- Expression levels of 1000 genes in cancerous vs. non-cancerous cells (observations = cells, features = genes)
- Network connectivity of brain regions in patients with epilepsy vs. healthy subjects (observations = subjects' brains, features = connections)
- Spatiotemporal changes in EEG gamma bandpower over time in a single subject (observations = time windows, features = gamma bandpower of channels)



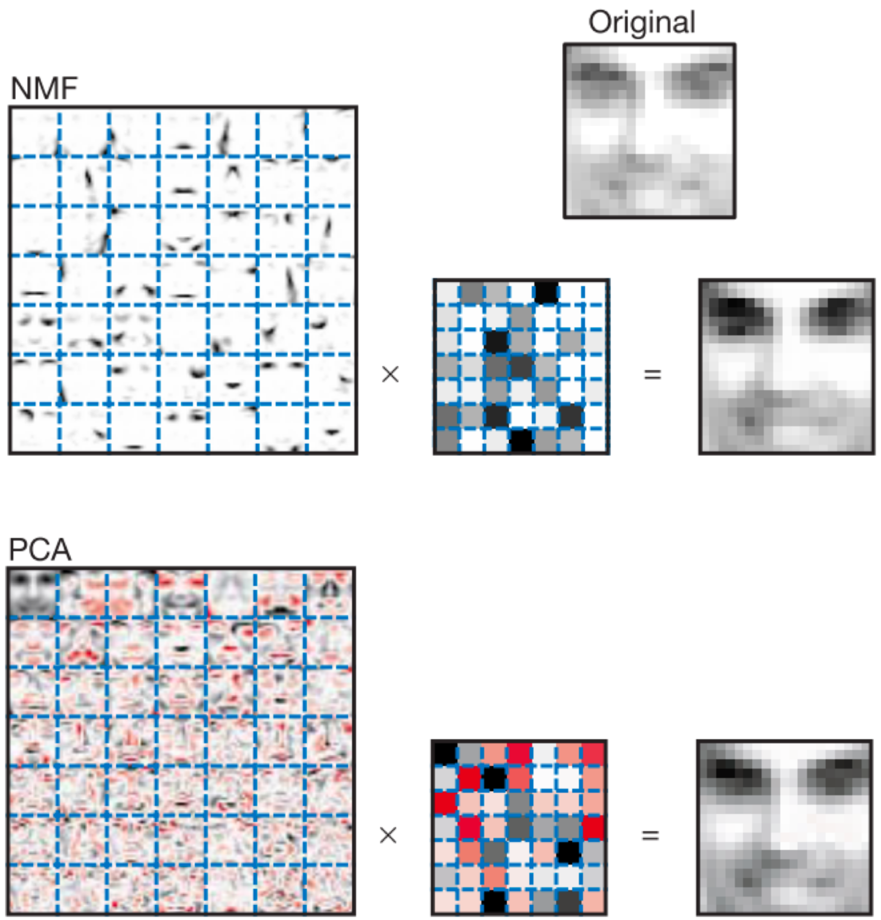
Outline

- Matrix factorization vs. embedding methods
- Applications for nonlinear dimensionality reduction methods
- Multidimensional scaling
- t-Distributed Stochastic Neighbour Embedding
- Cautions

Matrix factorization vs. embedding

Last time: matrix factorization methods for dimensionality reduction

- Start with a **matrix** that describes m features of n observations
- Find **components** (vectors, length m)
- Each observation can be described as a weighted sum of the components



Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791 Available at: http://www.ncbi.nlm.nih.gov/pubmed/10548103.

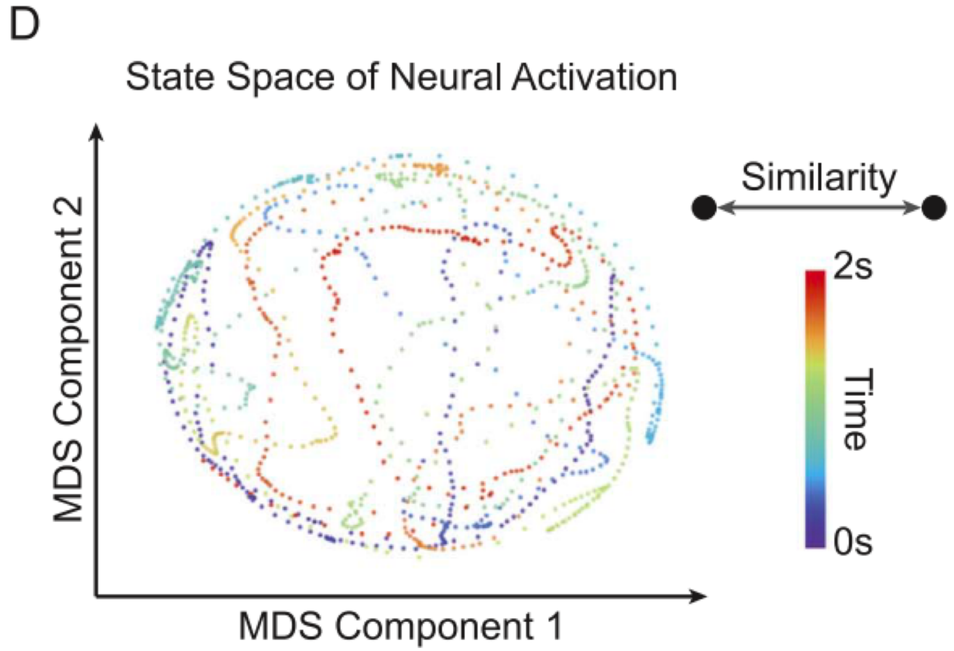
This tutorial: embedding approaches to dimensionality reduction

- General goal: maintain distances between observations in a lower-dimensional space
- Different techniques emphasize different distances (e.g., local structure vs. large distances) and use different algorithms
- Unlike matrix factorization techniques, does not yield interpretable components

Applications for nonlinear dimensionality reduction

Visualizing high dimensional data in low-dimensional space

- Usually visualize in 2-3 dimensions
- Able to visualize data even if intrinsic dimensionality is >3 dimensions



Khambhati AN, Sizemore AE, Betzel RF, Bassett DS (2017) Modeling and interpreting mesoscale network dynamics. Neuroimage:1–13 Available at: http://dx.doi.org/10.1016/j.neuroimage.2017.06.029.

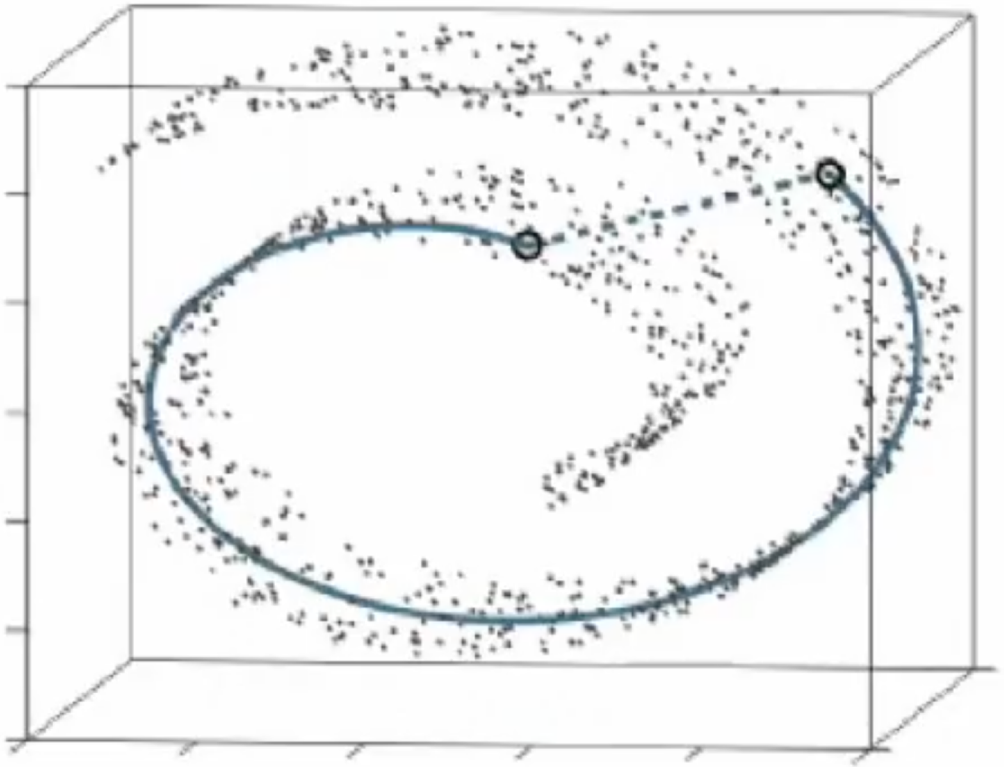
Visualizing high dimensional data in low-dimensional space

- Evaluate features to use for machine learning

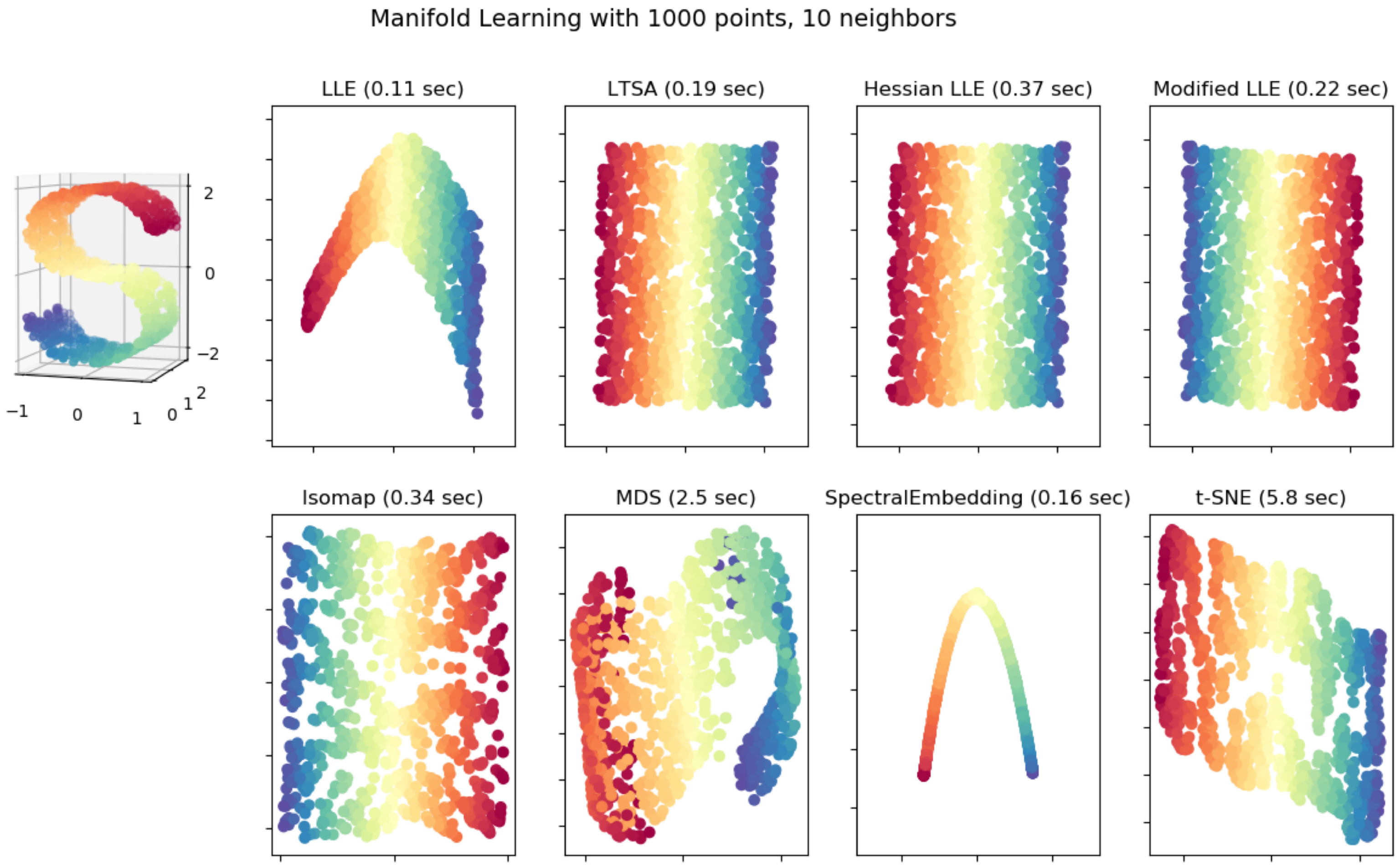
One cluster:

https://www.youtube.com/watch?v=RJVL80Gg3lA&list=UUtXKDgv1AVoG88PLl8nGXmw

Identifying non-linear structure



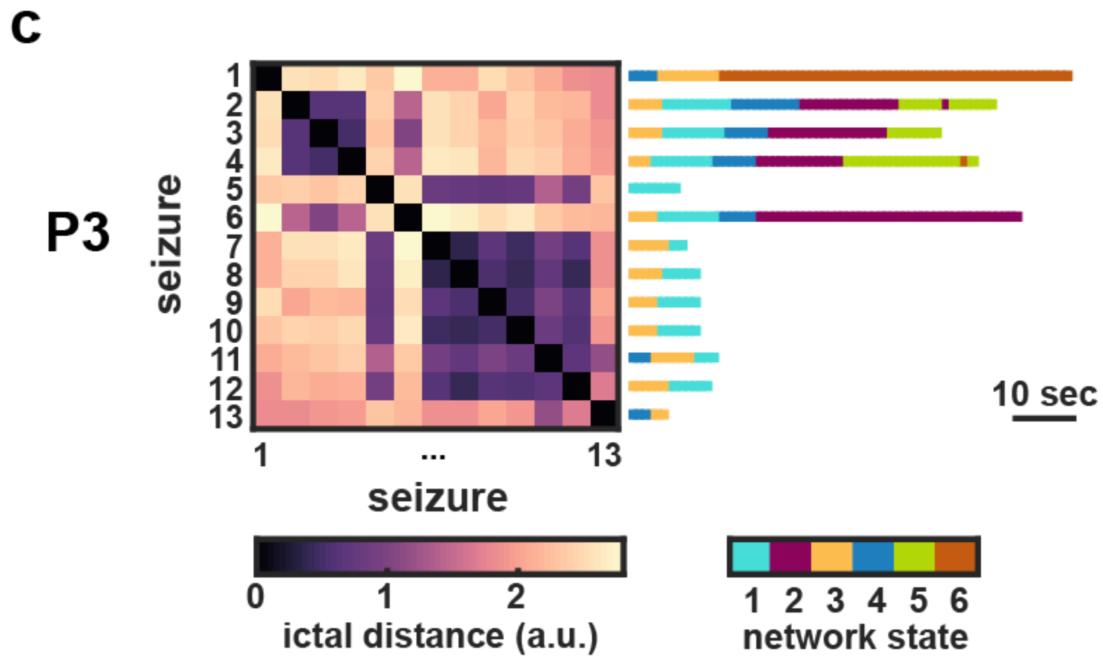
Identifying non-linear structure



https://scikit-learn.org/stable/auto_examples/manifold/plot_compare_methods.html

Convert a distance matrix to coordinates

- To use matrix factorization techniques, must start with a matrix (*features* of each observations)
- Sometimes, only know the *distances* between observations



Multidimensional scaling (MDS)

http://www-stat.wharton.upenn.edu/~buja/PAPERS/paper-mds-jcgs.pdf

http://www.lse.ac.uk/Statistics/Assets/Documents/chapter3.pdf

https://en.wikipedia.org/wiki/Multidimensional_scaling

Multidimensional scaling

Given a symmetric **dissimilarity matrix** D, find coordinates for the points such that the dissimilarities between points are maintained

i.e., map objects $i = 1,...,N$ to points $x_1,...,x_N \in IR^k$ in such a way that the given **dissimilarities** $D_{i,j}$ are well-approximated by the **distances** $\|x_i - x_j\|$

Flavours of MDS

The terminology for MDS is very confusing because

1. there are multiple types of distinctions
2. the same terms are often used to mean different things!

Most common distinction (I have also seen these terms used in different ways, so pay attention to the algorithm used…)

- **Metric MDS:** tries to maintain the observed dissimilarities in the embedding
- **Non-metric (ordinal) MDS:** tries to preserve the order (i.e., ranks) of the observed dissimilarities in the embedding

The term "**classic/classical**" MDS is also used, but is ambiguous – sometimes means metric MDS, sometimes means Torgerson's MDS (see next slide)

Torgerson MDS

- Type of metric MDS; often called classical MDS
- Also called Principal Coordinate Analysis (PCoA)
- Minimizes "strain"
- Solved using eigenvalue decomposition (vs. other forms of MDS, which are solved iteratively)
- Equivalent to PCA if use Euclidean distances (minus translations/reflections)
- Matlab function: cmdscale

https://stats.stackexchange.com/questions/14002/whats-the-difference-between-principal-component-analysis-and-multidimensional

Iterative forms of MDS

- Choose number of dimensions (usually 2-3)
- Use a cost function to minimize difference between dissimilarities and the distances between points in the embedding
- Common cost function for metric MDS: normalized stress

Distances in embedding

Dissimilarities (original distances)

$$\sqrt{\frac{\sum_{i<j}(d_{ij}-\delta_{ij})^2}{\sum_{i<j} d_{ij}^2}}.$$

- Lower stress = better fit
- Can plot stress vs. number of dimensions to determine dimensionality
- Matlab function: mdscale

Nonmetric MDS

- Preserve the *order* of the distances between observations (useful if the order of the distances, rather than the absolute values, is most meaningful)
- Kruskal's normalized stress, type 1:

Distances in embedding

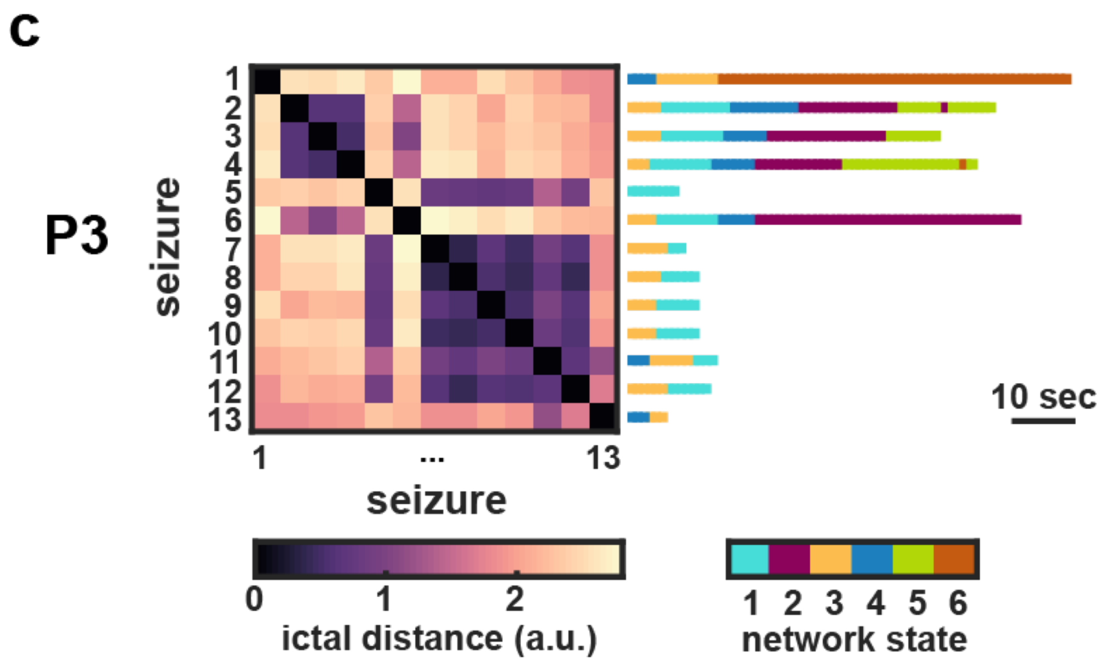
Disparities (original distances mapped to new distances using a monotonic function)

$$\sqrt{\frac{\sum_{i<j}(d_{ij}-\hat{d}_{ij})^2}{\sum_{i<j} d_{ij}^2}}.$$

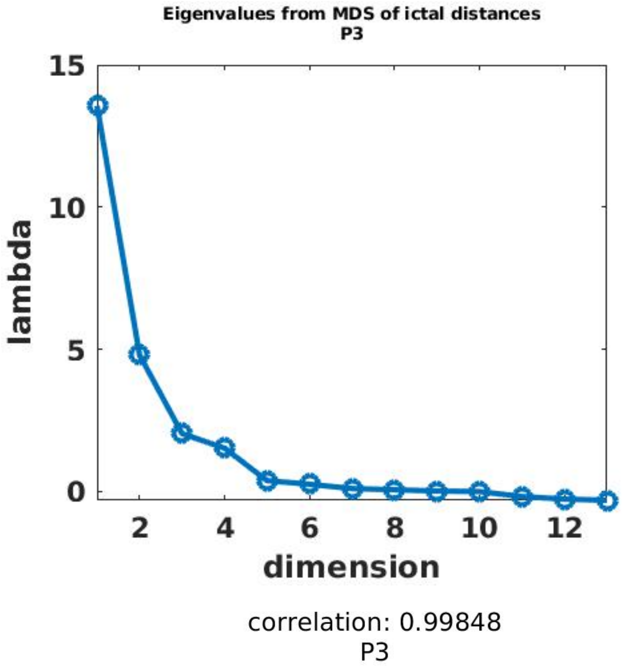
Tutorial, Part 1: Great Britain Cities

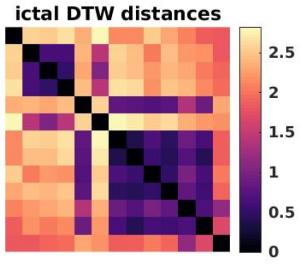
https://github.com/gmschroe/dim-reduction-tutorial/tree/master/embedding_tutorial

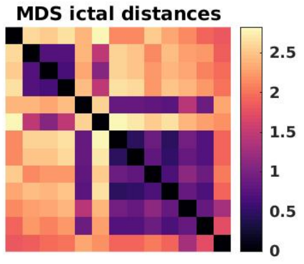
Example – embedding seizures in Cartesian space

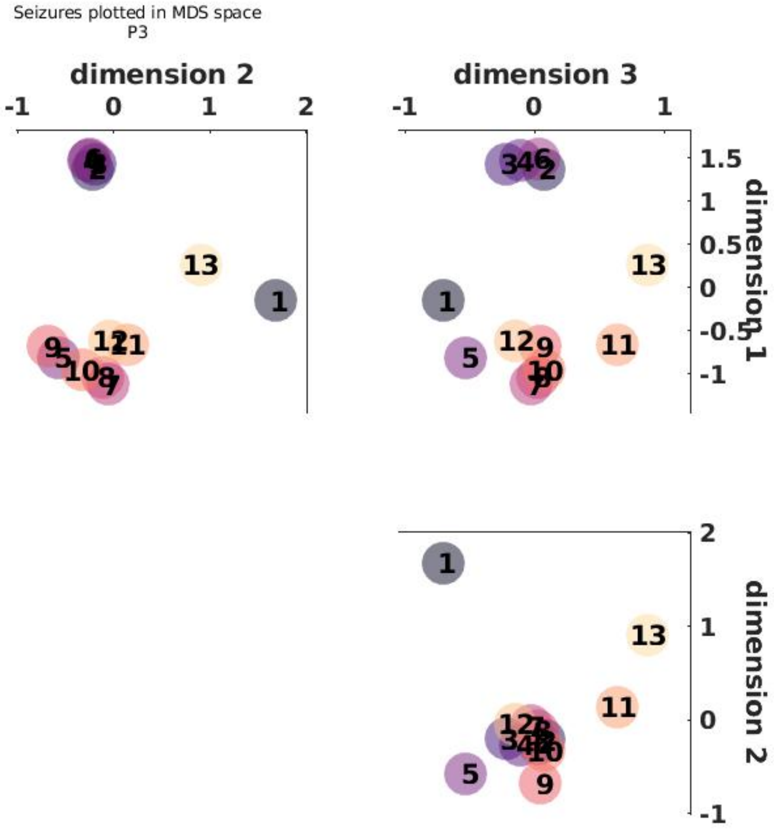


Example – embedding seizures in Cartesian space









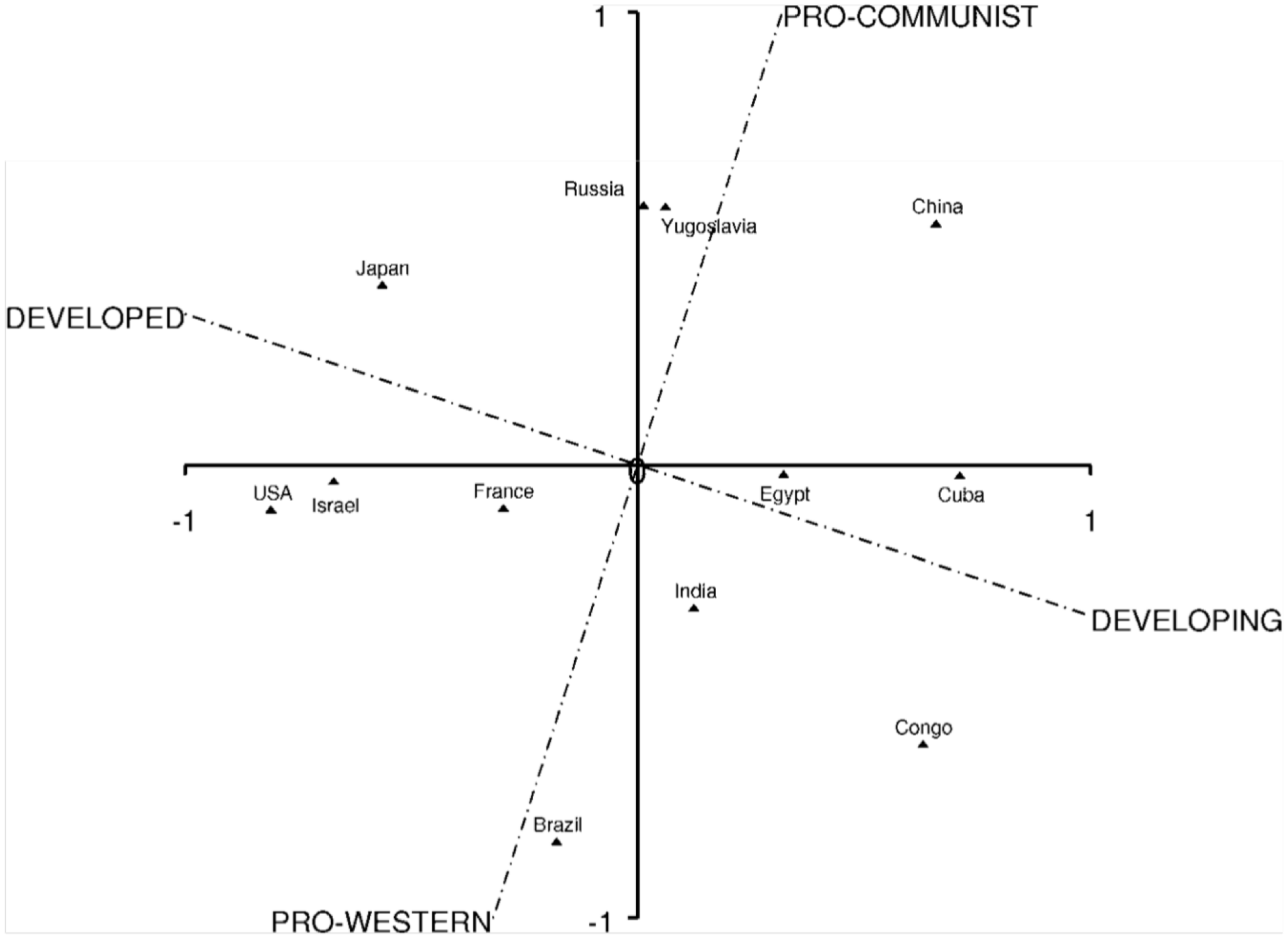
Used Torgerson's MDS to embed the seizures in Cartesian space, based on the ictal distances.

Used maximum number of dimensions in order to preserve the distances as closely as possible. The coordinates were then used for cluster evaluation.

Do not use lower-dimensional embeddings of iterative MDS for downstream analysis!

Can the dimensions be interpreted?

- Can be somewhat interpretable – but may need to rotate the axes
- E.g.: MDS of subjective similarity between pairs of countries (http://www.lse.ac.uk/Statistics/Assets/Documents/chapter3.pdf)
- In general, a technique that provides interpretable dimensions (e.g., PCA, NMF) is preferable when possible
- Clusters/shapes may be more insightful than the dimensions



t-Distributed Stochastic Neighbour Embedding

https://lvdmaaten.github.io/tsne/ (recommend the Google tech talk)

https://distill.pub/2016/misread-tsne/ (interactive walkthrough)

t-SNE maintains local structure

Cost function focuses on preserving small distances; if two points are far apart, any errors contribute very very little to the cost function.

See the Google tech talk on https://lvdmaaten.github.io/tsne/ for a good walk-through of the maths.

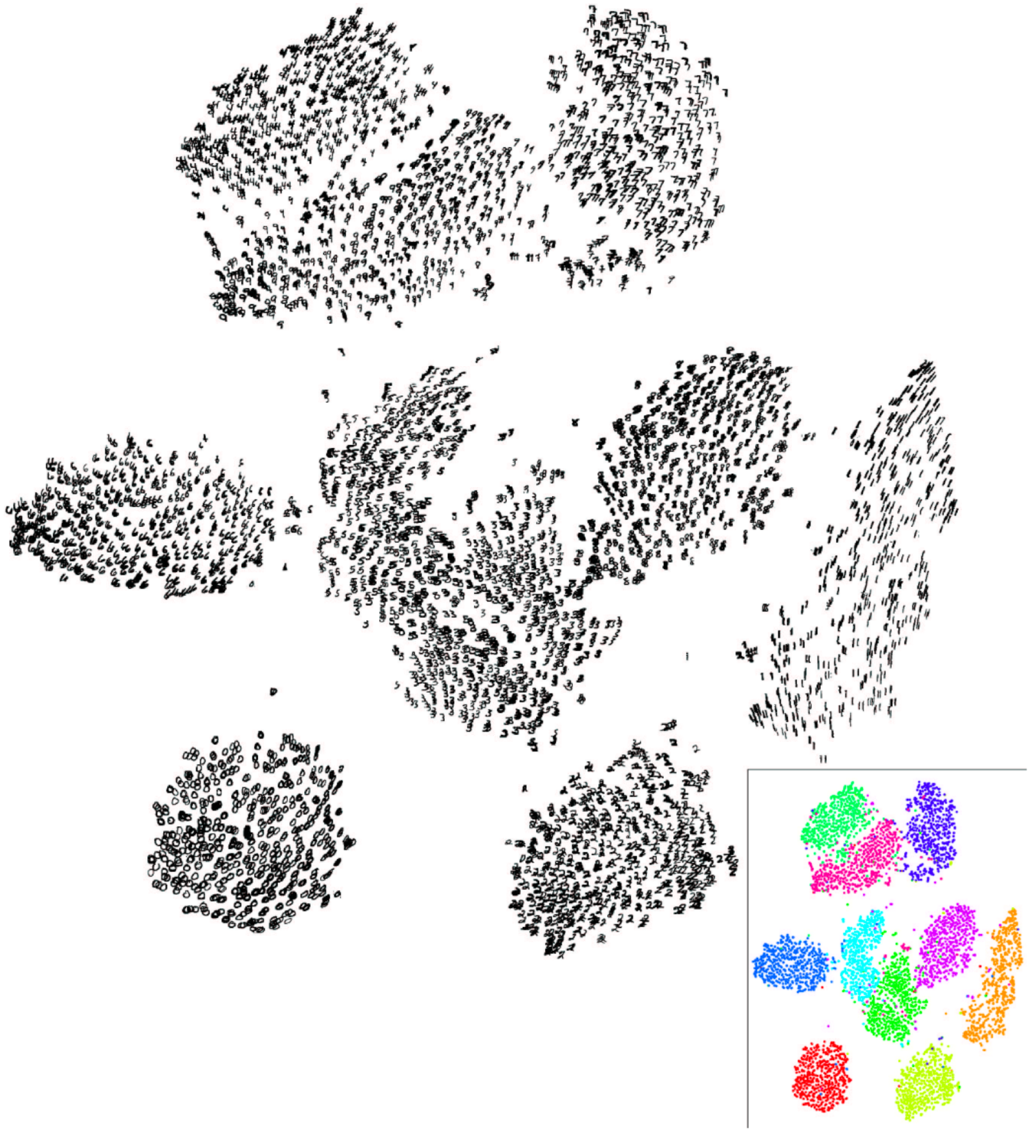
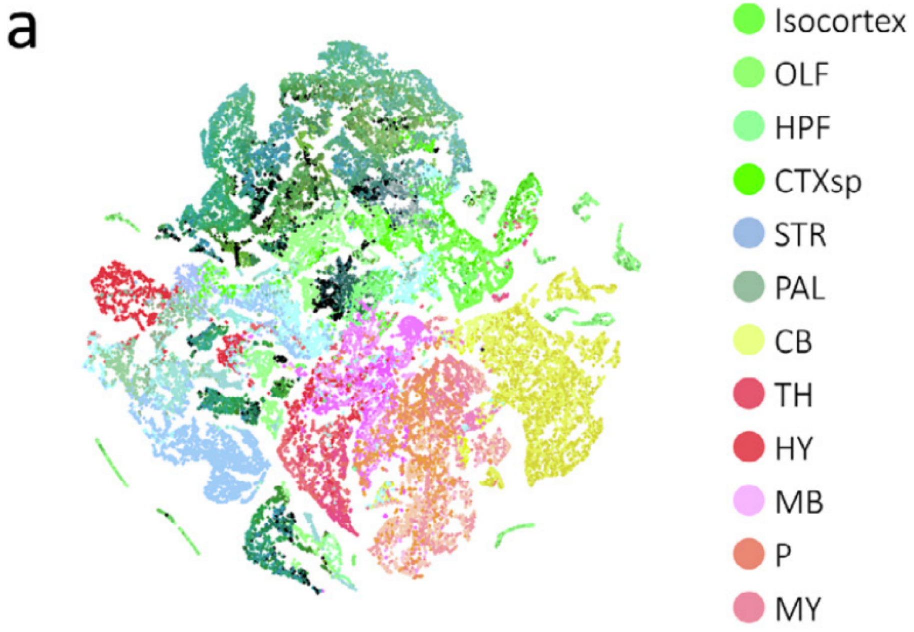


Figure 7: Visualization of 6,000 digits from the MNIST data set produced by the random walk version of t-SNE (employing all 60,000 digit images).

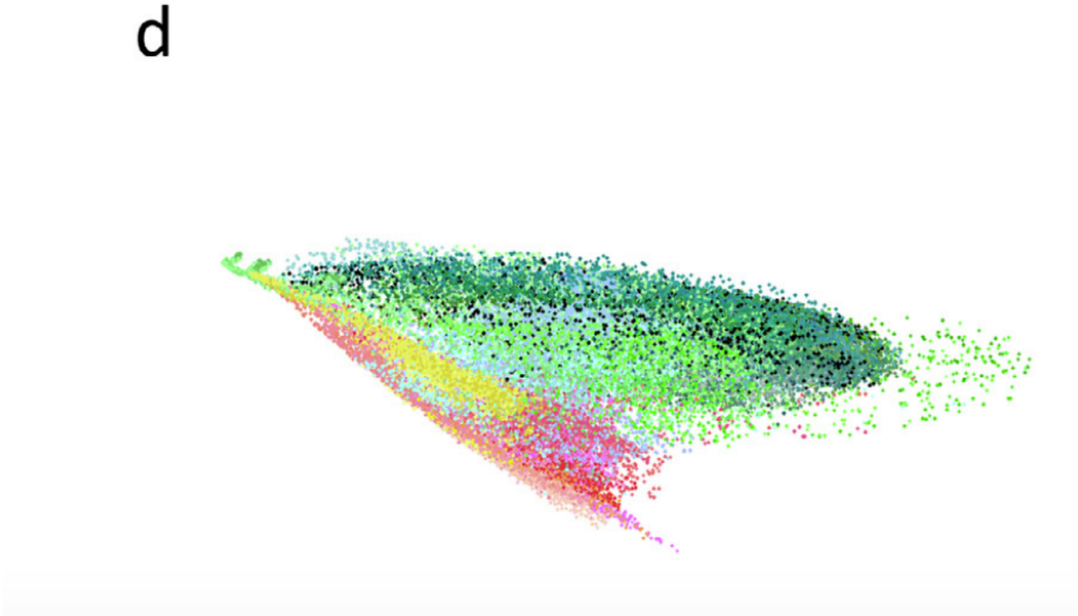
Maaten L Van Der, Hinton G (2008) Visualizing Data using t-SNE. J Mach Learn Res 1 620:267–284 Available at: http://link.springer.com/10.1007/s10479-011-0841-3%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/20652508.

Mouse brain transcriptome similarities

t-SNE

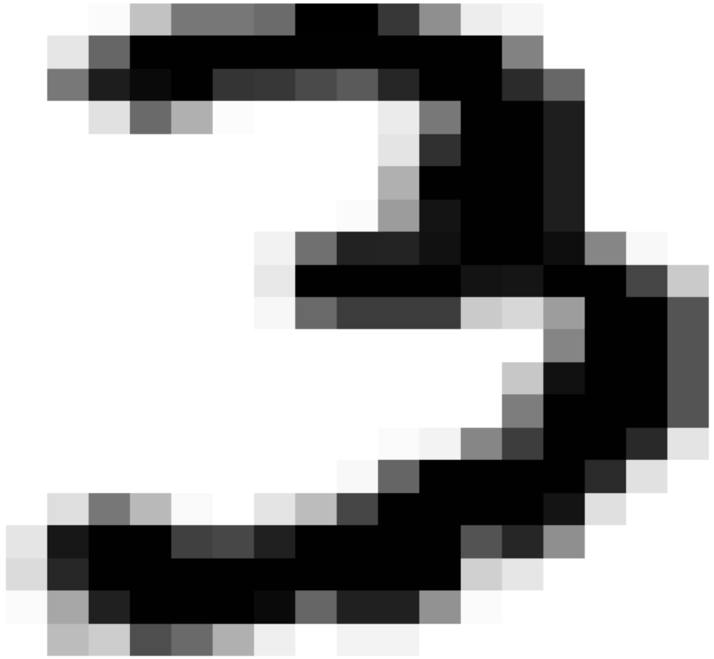


PCA (first two components)



Mahfouz A, van de Giessen M, van der Maaten L, Huisman S, Reinders M, Hawrylycz MJ, Lelieveldt BPF (2015) Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. Methods 73:79–89 Available at: http://dx.doi.org/10.1016/j.ymeth.2014.10.004.

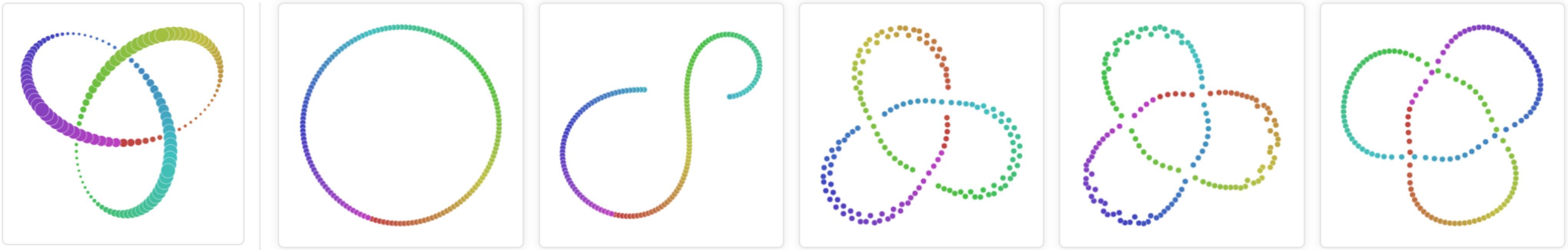
Tutorial: Digits



https://github.com/gmschroe/dim-reduction-tutorial/tree/master/embedding_tutorial

Cautions

- These techniques are best used for **visualizing** data
- Be wary of using the embedding (especially a low-dimensional embedding) for purposes other than visualization
  - Errors in the embedding may not be distributed among all points – instead, there may be particular points that are incorrectly represented
  - Only certain distances may be accurately represented (especially using tSNE)
  - See https://stats.stackexchange.com/questions/263539/clustering-on-the-output-of-t-sne for a discussion of clustering using tSNE output
- The dimensions may be difficult to interpret
  - May make more sense if you rotate that data
  - May not correspond to any particular features (esp. t-SNE, since only local structure is meaningful)
- Cannot add new data to the map - no function from high dimensional space to low dimensional space (t-SNE creator may be working on this; see also https://stats.stackexchange.com/questions/368331/project-new-point-into-mds-space)
- In t-SNE embeddings, cluster sizes and distances between clusters do not convey any information
- t-SNE parameter can greatly affect the results (https://distill.pub/2016/misread-tsne/)



Original | Perplexity: 2 Step: 5,000 | Perplexity: 5 Step: 5,000 | Perplexity: 30 Step: 5,000 | Perplexity: 50 Step: 5,000 | Perplexity: 100 Step: 5,000

There are many more nonlinear dimensionality reduction techniques

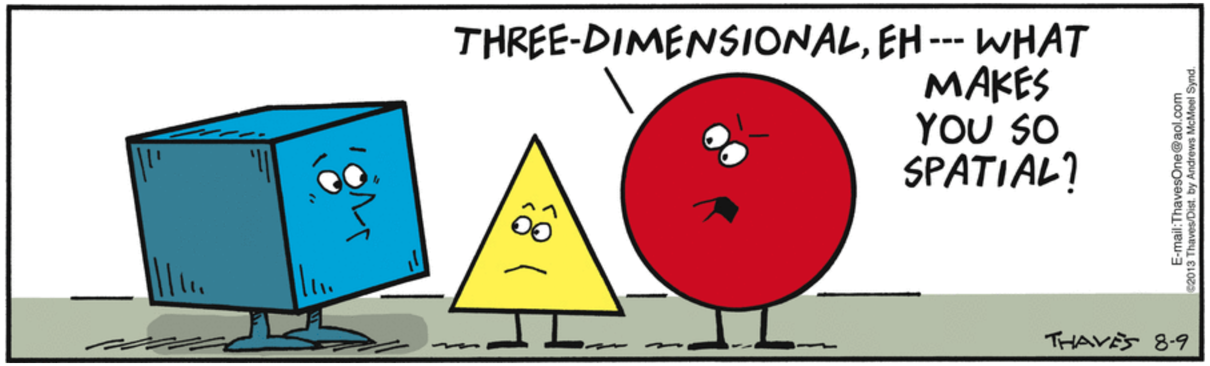
Good (albeit slightly biased) review:

Maaten L Van Der, Postma E, Herik J (2009) Dimensionality Reduction : A Comparative Review Dimensionality Reduction : A Comparative Review.

More on t-SNE:

Maaten L Van Der, Hinton G (2008) Visualizing Data using t-SNE. J Mach Learn Res 1 620:267–284

Questions?


THREE-DIMENSIONAL, EH --- WHAT MAKES YOU SO SPATIAL?
E-mail: ThavesOne@aol.com
©2013 Thaves/Dist. by Andrews McMeel Synd.
THAVES 8-9