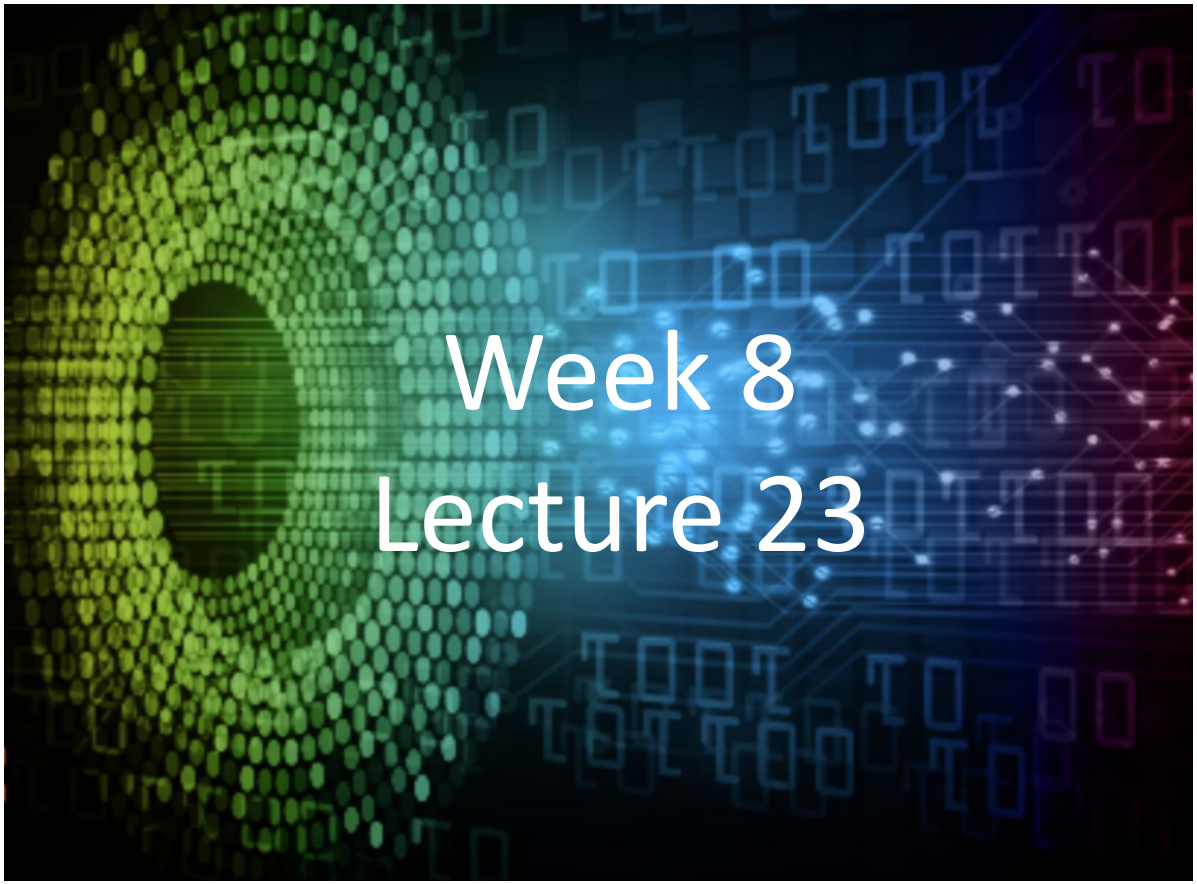
Introduction to Deep Learning
Applications and Theory



AMATH 563

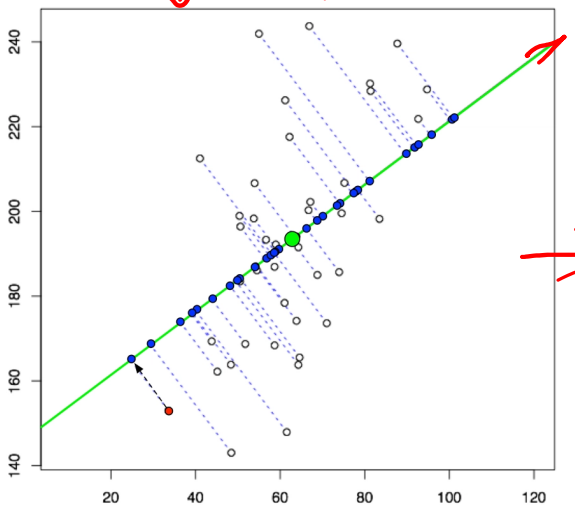
Previous Lecture: Data Decomposition

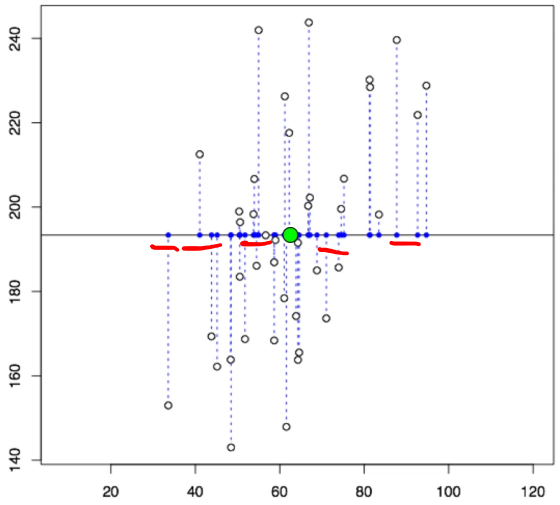
- **Singular Value Decomposition (SVD)**
- Principal Component Analysis (PCA)
- Proper Orthogonal Decomposition (POD)
- Dynamic Mode Decomposition (DMD)
- Time-delayed Embeddings

This Lecture: Manifold Learning and Representation

- Multi-Dimensional Scaling (MDS)
- ISOMAP
- t-SNE
- Force Directed graphs

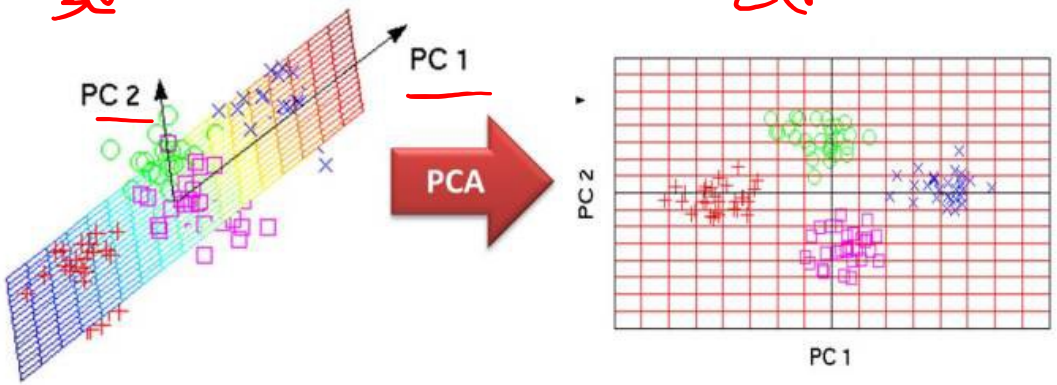
PCA Embedding - Truncation





$$X = U\Sigma V^T$$

$$U_N^T X = \Sigma_N V_N^T$$

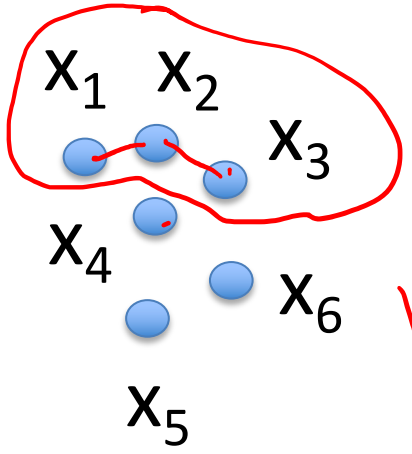


Embedding

50

High

d-dim:



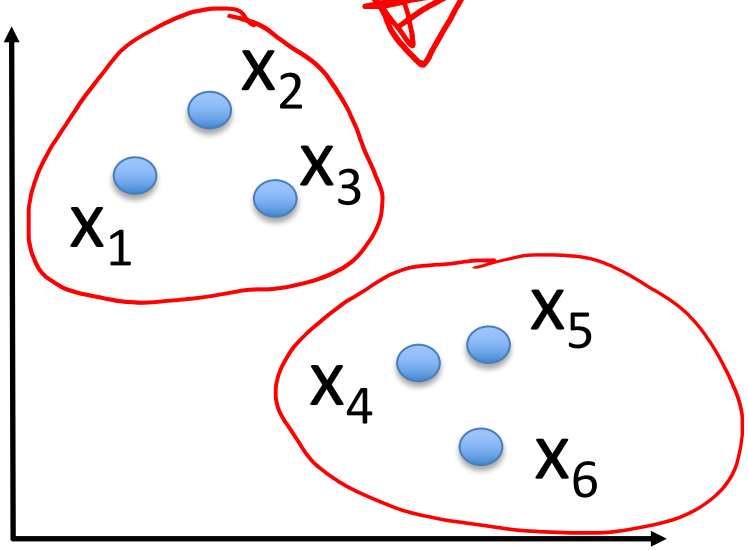
$d(x_1, x_2) < \delta$

$d(x_2, x_3) < \delta$

$d(x_3, x_4) > \delta$

Low

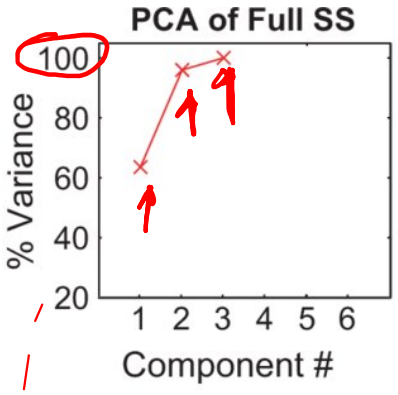
2-dim:

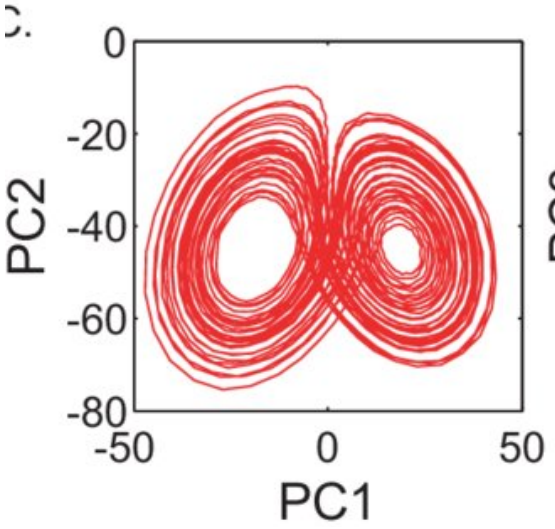


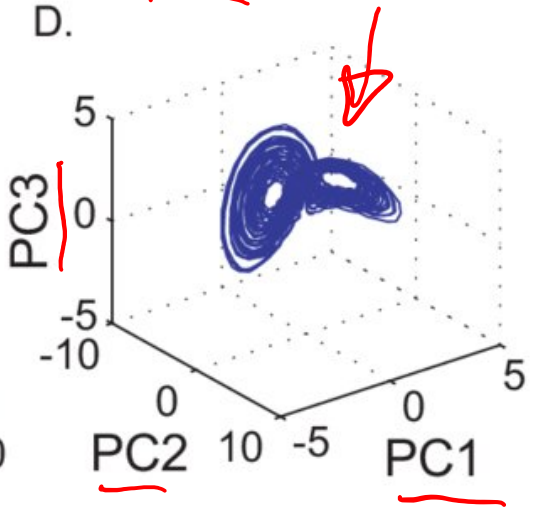
Time Delay Embeddings - Example

$$H = [x_1; x_1^{\delta}; ...; ; x_1^{(k+1)\delta}]$$

$$H = U\Sigma V^T$$







Manifold Learning

Manifold Learning: a subfield of machine learning closely related to dimension reduction based on the assumption that one's observed data lie on a low-dimensional manifold embedded in a higher-dimensional space.

- Manifold Approximation
- Manifold Visualization

Multidimensional Scaling (MDS)

PCoA

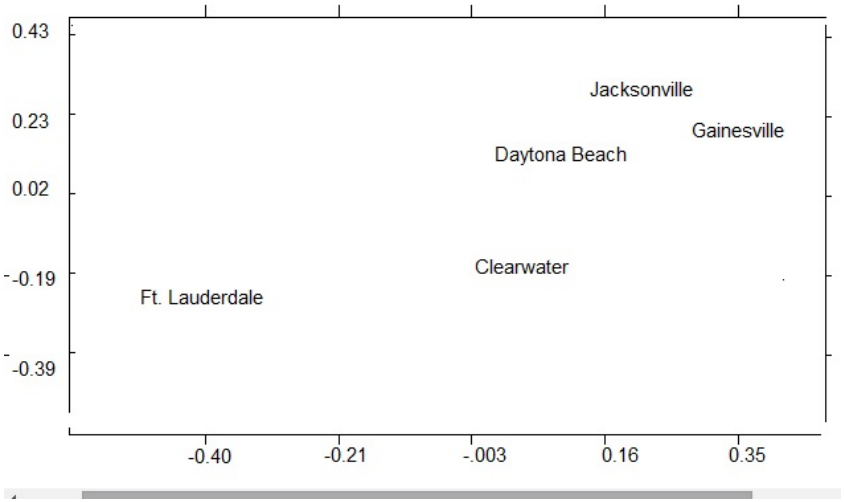
Multidimensional scaling is a **visual representation of distances or dissimilarities between sets of objects**. "Objects" can be any kind of real or conceptual stimuli (e.g. colors, faces, map coordinates, etc).

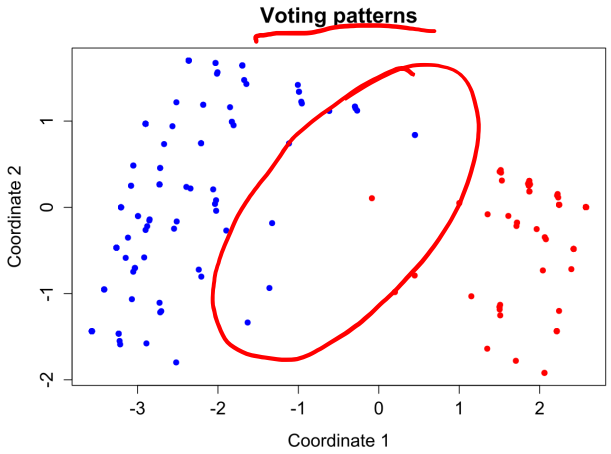
- Objects that are more similar (or have shorter distances) are closer together on the graph than objects that are less similar (or have longer distances).

- MDS can also serve as a dimension reduction technique for high-dimensional data As well as interpreting dissimilarities as distances on a graph.

Multidimensional Scaling (MDS or PCoA)

CITY	Clearwater	Daytona Beach	Ft. Lauderdale	Gainesville	Jacksonville
Clearwater	0	159	247	131	197
Daytona Beach	159	0	230	97	89
Ft. Lauderdale	247	230	0	309	317
Gainesville	131	97	309	0	68
Jacksonville	197	89	317	68	0





MDS Algorithm

Compute pairwise distance between all pairs of points

$$D := \begin{pmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,M} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,M} \\ \vdots & \vdots & & \vdots \\ d_{M,1} & d_{M,2} & \cdots & d_{M,M} \end{pmatrix}$$

$$d_{ij} = ||x_i - x_j||$$

Center the matrix D (double centering in rows and columns)

$$\widetilde{D} \rightarrow^{EVD} U\sqrt{\Sigma}$$

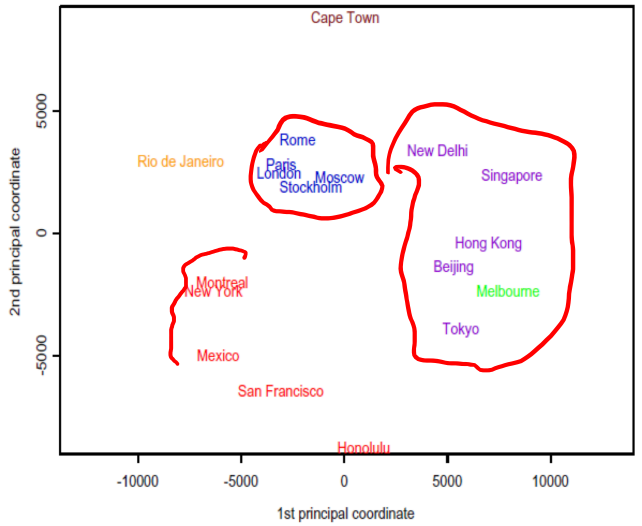
$$X = U\Sigma V^*$$

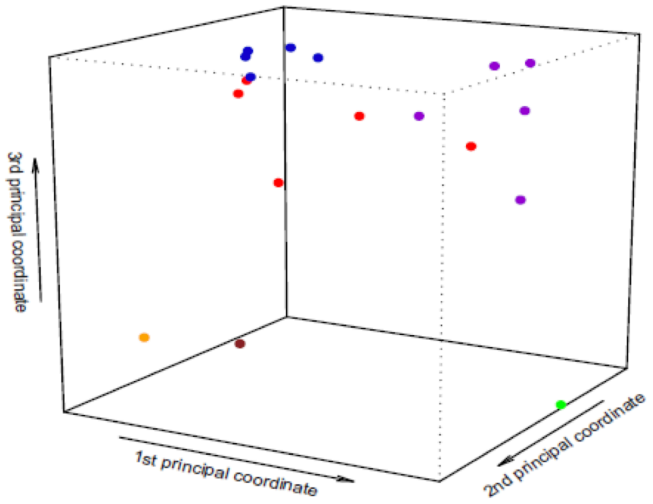
MDS Example

	Beijing	Cape Town	Hong Kong	Honolulu	London	Melbourne
Cape Town	12947					
Hong Kong	1972	11867				
Honolulu	8171	18562	8945			
London	8160	9635	9646	11653		
Melbourne	9093	10338	7392	8862	16902	
Mexico	12478	13703	14155	6098	8947	13557
Montreal	10490	12744	12462	7915	5240	16730
Moscow	5809	10101	7158	11342	2506	14418
New Delhi	3788	9284	3770	11930	6724	10192
New York	11012	12551	12984	7996	5586	16671
Paris	8236	9307	9650	11988	341	16793
Rio de Janeiro	17325	6075	17710	13343	9254	13227
Rome	8144	8417	9300	12936	1434	15987
San Francisco	9524	16487	11121	3857	8640	12644
Singapore	4465	9671	2575	10824	10860	6050
Stockholm	6725	10334	8243	11059	1436	15593
Tokyo	2104	14737	2893	6208	9585	8159

	Mexico	Montreal	Moscow	New Delhi	New York	Paris
Montreal	3728					
Moscow	10740	7077				
New Delhi	14679	11286	4349			
New York	3362	533	7530	11779		
Paris	9213	5522	2492	6601	5851	

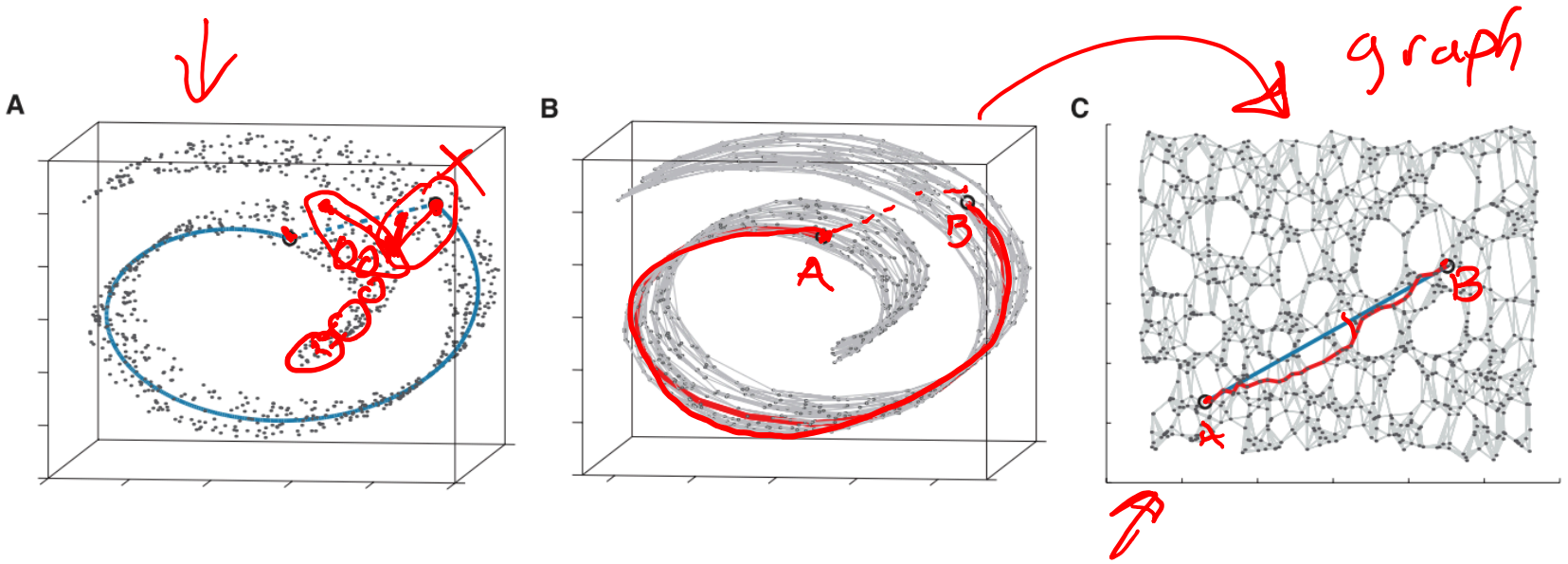
	Eigenvalues	Eigenvectors		
1	471582511	0.245	−0.072	0.183
2	316824787	0.003	0.502	-0.347
3	253943687	0.323	−0.017	0.103
4	−98466163	0.044	−0.487	-0.080
5	−74912121	−0.145	0.144	0.205
6	−47505097	0.366	−0.128	−0.569
7	31736348	−0.281	−0.275	−0.174
8	−7508328	−0.272	−0.115	0.094
9	4338497	−0.010	0.134	0.202
10	1747583	0.209	0.195	0.110
11	−1498641	−0.292	−0.117	0.061
12	145113	−0.141	0.163	0.196
13	−102966	−0.364	0.172	−0.473
14	60477	−0.104	0.220	0.163
15	−6334	−0.140	−0.356	−0.009
16	−1362	0.375	0.139	−0.054
17	100	−0.074	0.112	0.215
18	0	0.260	−0.214	0.173





ISOMAP

Isomap: nonlinear dimensionality reduction estimating the intrinsic geometry of a data manifold based on a rough estimate of each data point's neighbors on the manifold.



J. B. Tenenbaum, V. de Silva, J. C. Langford,
A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290, (2000), 2319–2323.

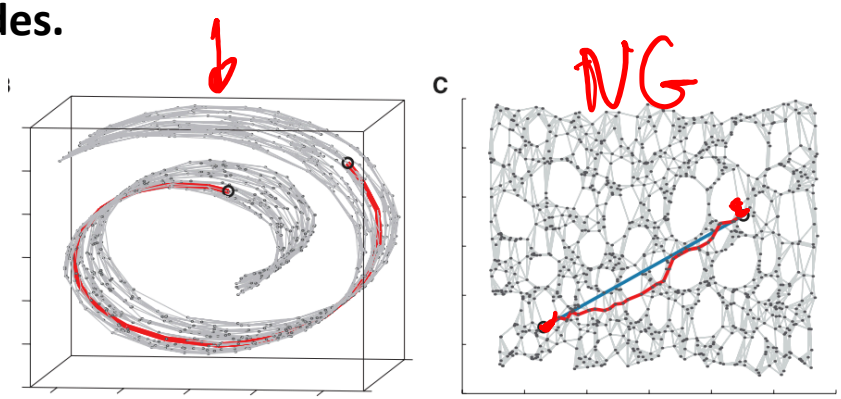
ISOMAP

$$D := \begin{pmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,M} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,M} \\ \vdots & \vdots & & \vdots \\ d_{M,1} & d_{M,2} & \cdots & d_{M,M} \end{pmatrix}$$

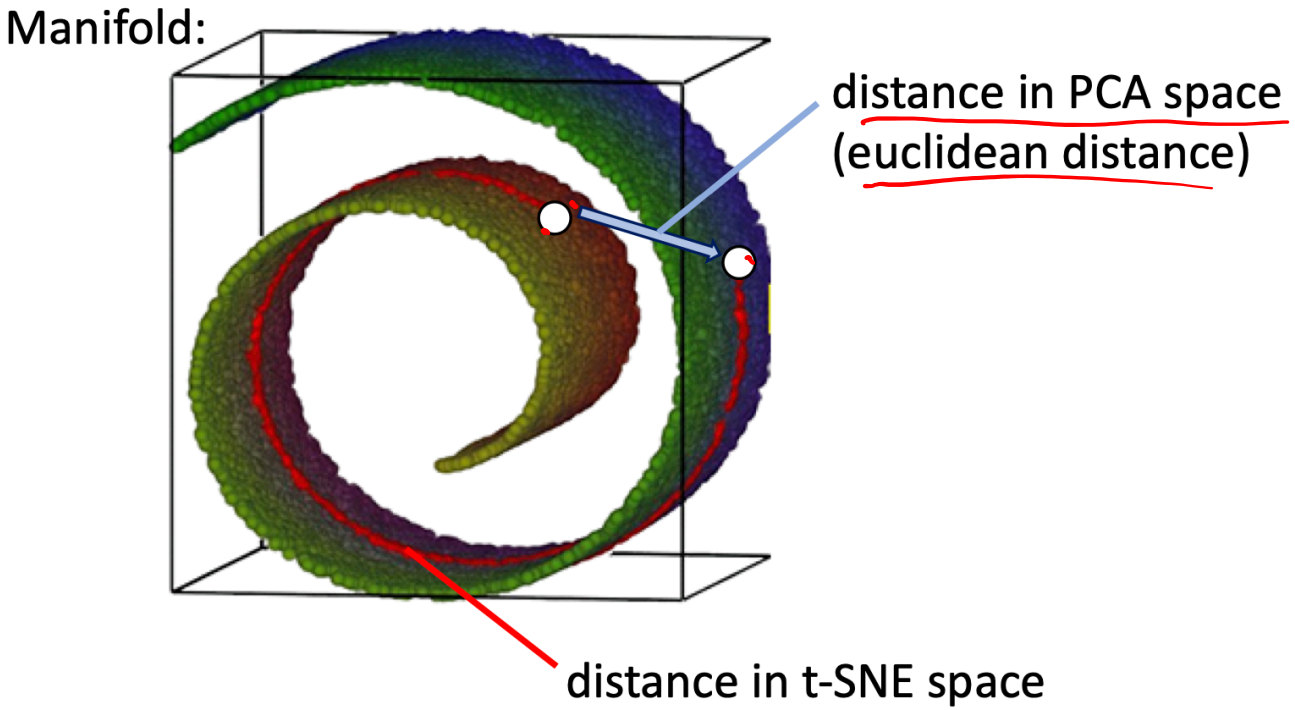
$$d_{ij} = d^g_{ij}$$

$$D \rightarrow^{SVD} U\sqrt{\Sigma}$$

- **Determine the neighbors of each point.**
  - All points in some fixed radius.
  - *K* nearest neighbors.
- **Construct a neighborhood graph.**
  - Each point is connected to other if it is a *K* nearest neighbor.
  - Edge length equal to Euclidean distance.
- **Compute shortest path between two nodes.**
  - Dijkstra's algorithm
  - Floyd–Warshall algorithm
- **Compute lower-dimensional embedding.**
  - Multidimensional scaling



t-SNE: T Stochastic Neighborhood Embedding

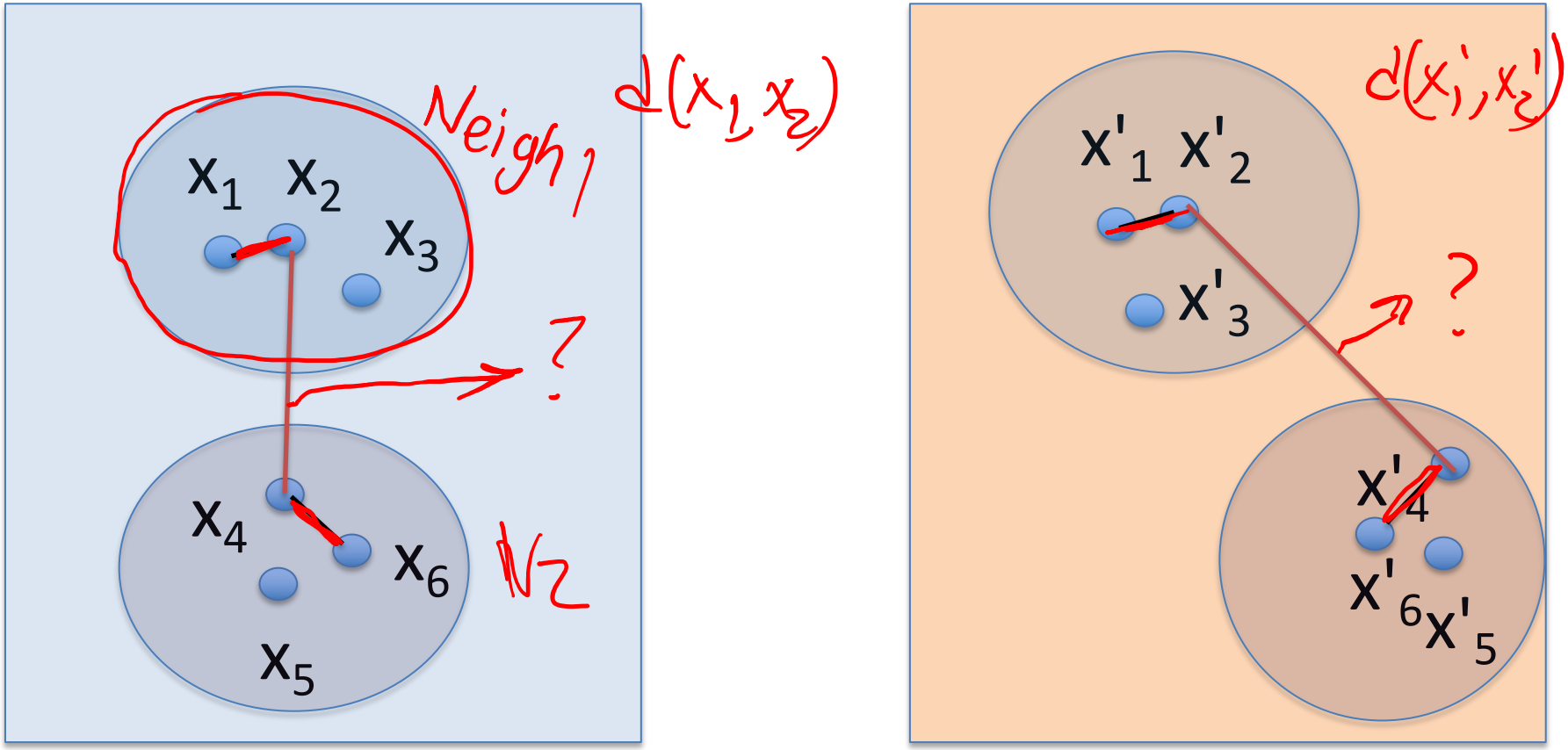


Maaten et al (2008) *Journal of Machine Learning Research*

t-SNE

high-dim:

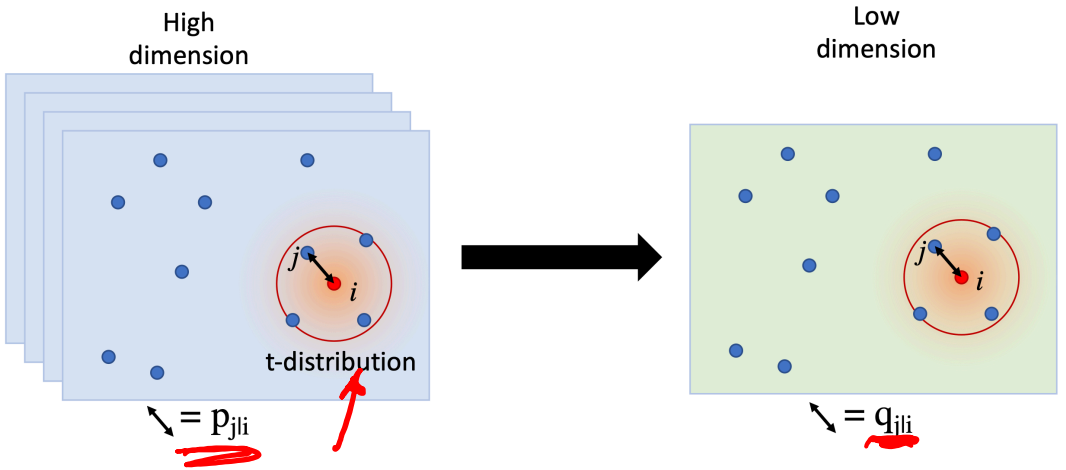
low-dim:



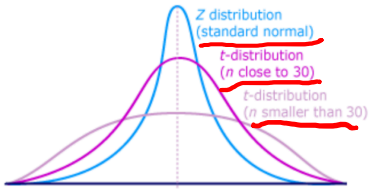
t-SNE

Crowding Problem: It is impossible to preserve distance in all Neighborhoods

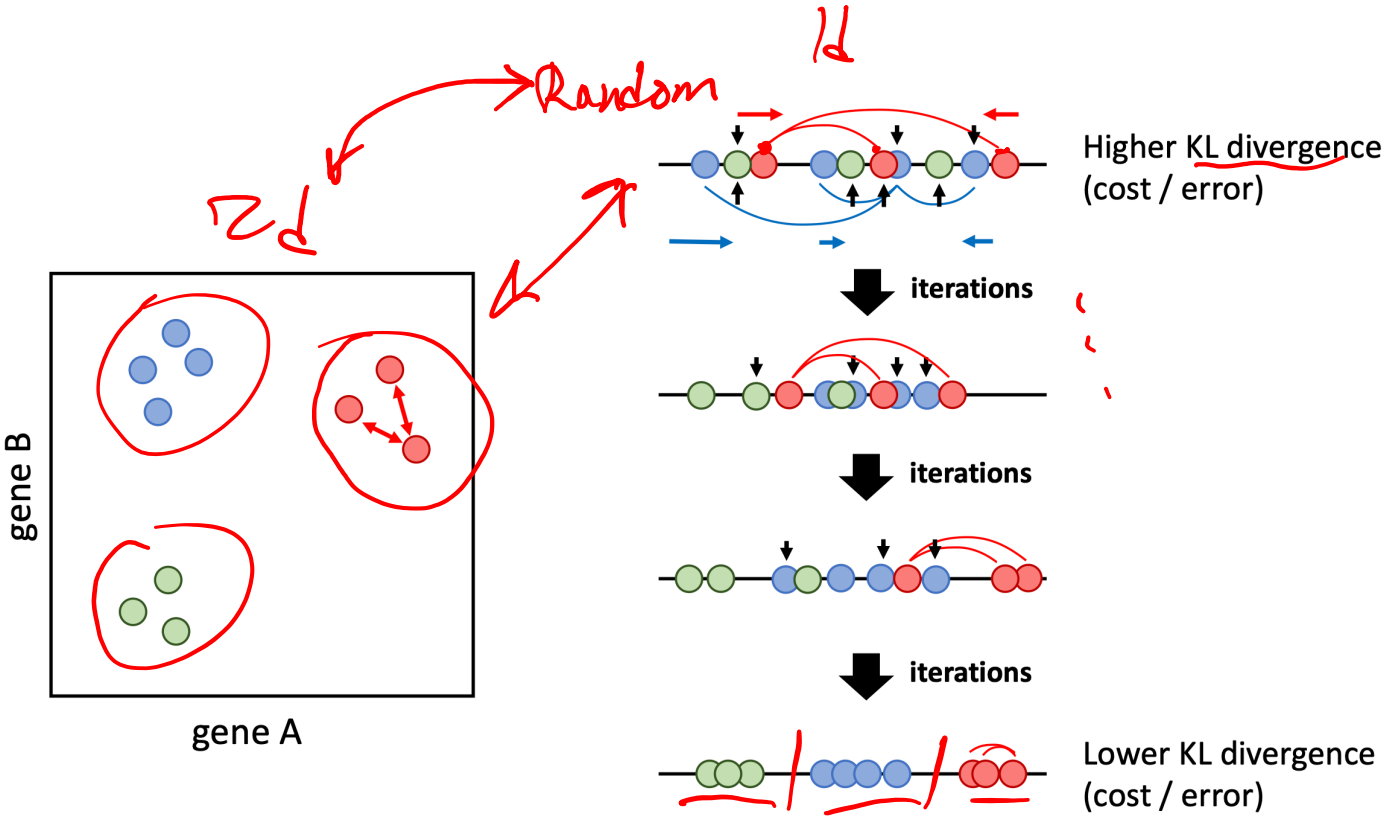
Possible Solution: t-distribution



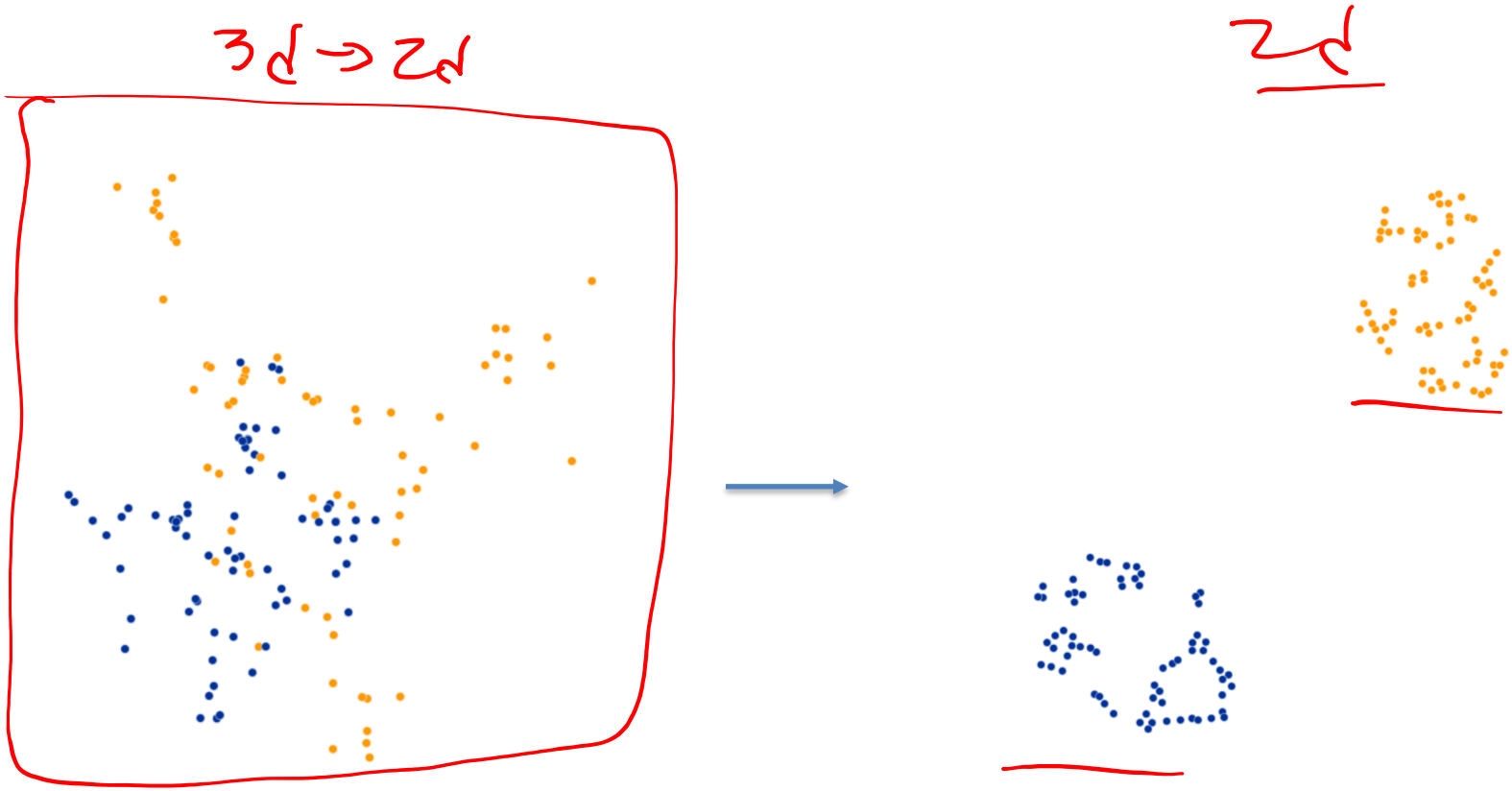
p_{jli} and q_{jli} measure the conditional probability that a point i would pick point j as it's nearest neighbor, in high (p) and low (q) dimensional space respectively.



t-SNE


Random
gene B
gene A
Higher KL divergence
(cost / error)
iterations
iterations
iterations
Lower KL divergence
(cost / error)

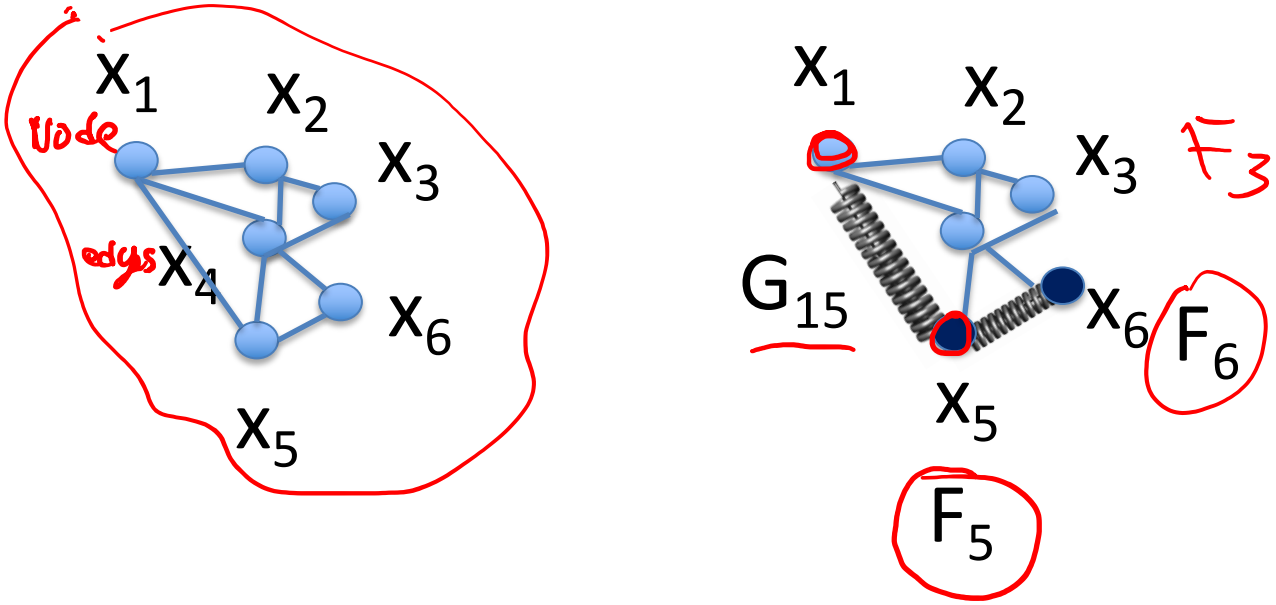
t-SNE



Blog: distill.pub

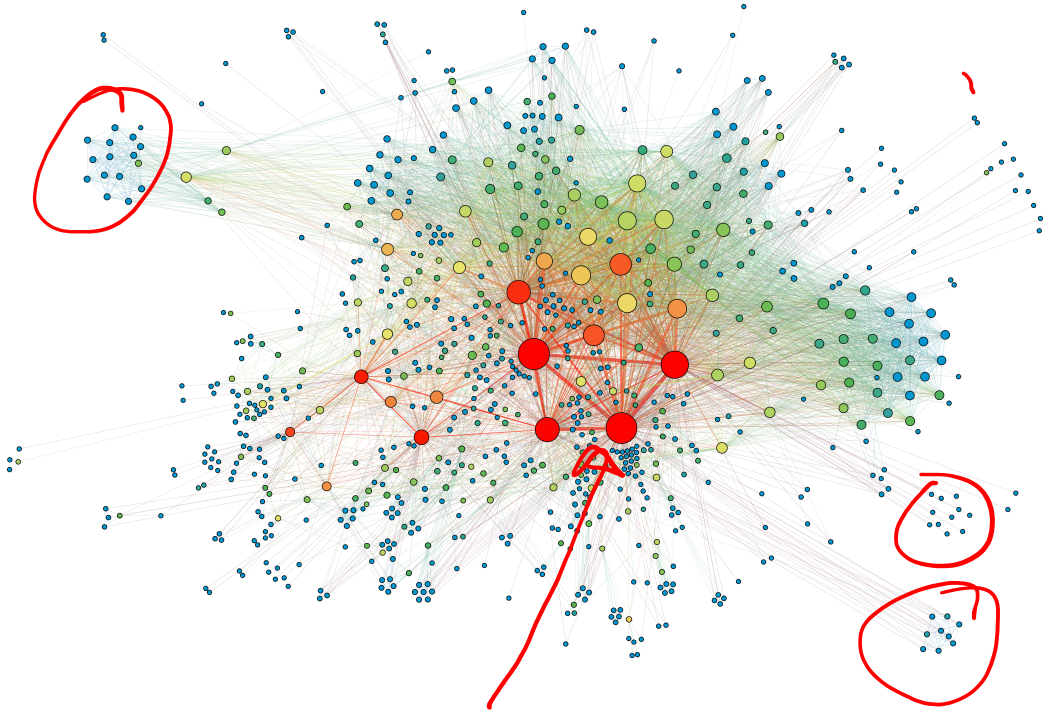
(Temporal) Force Directed Graphs

Embedding ⟶ **Visualization**



Kobourov, Stephen G. "Spring embedders and force directed graph drawing algorithms." *arXiv preprint arXiv:1201.3011*(2012).

Force Directed Graphs



D3.js