Lecture 10 Part2 - Math 143
ISOMAP

Prof. Weiqing Gu

ISOMAP is an extension of PCA

PCA

(Use Euclidean distance)

↓

MDS (multi dimensional scaling,

Use distance not necessary from L_2)

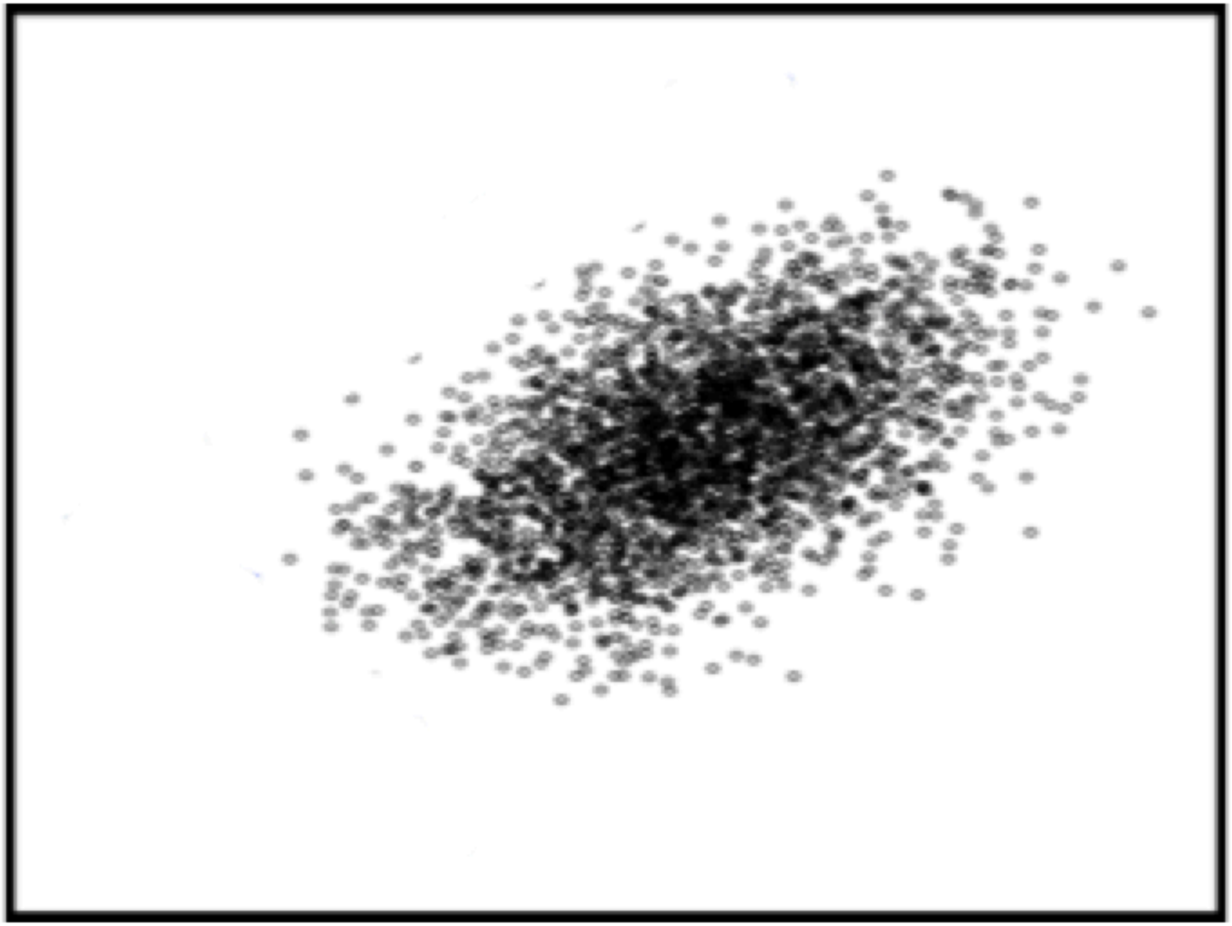
↓

ISOMAP

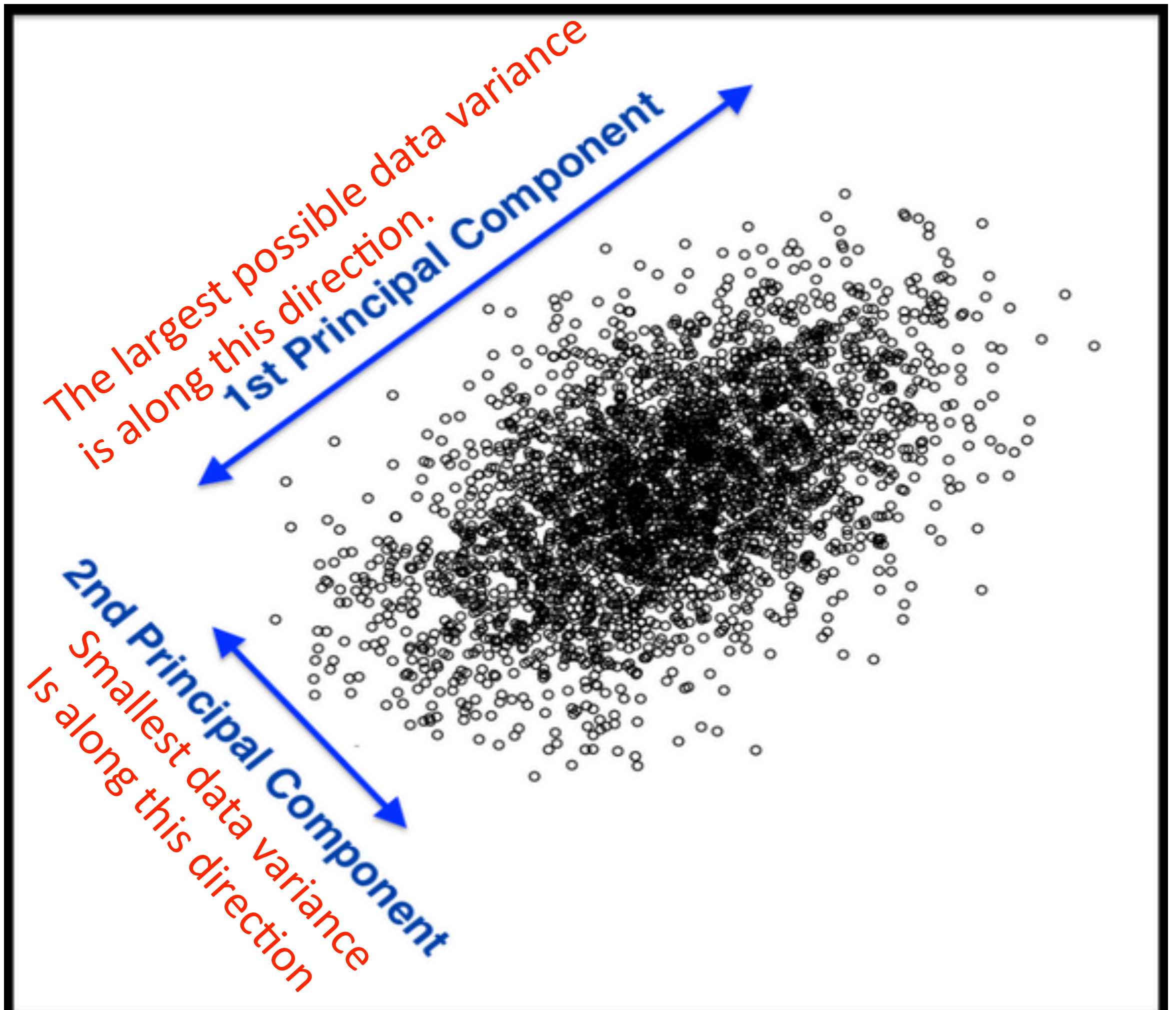
(Replace Euclidean distance by Geodesic distance)

Recall: Principal Component Analysis (PCA)

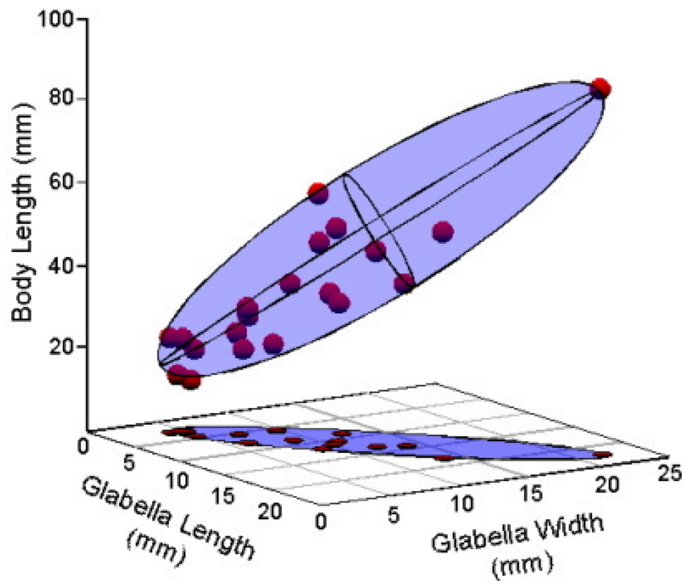
An intuitive example

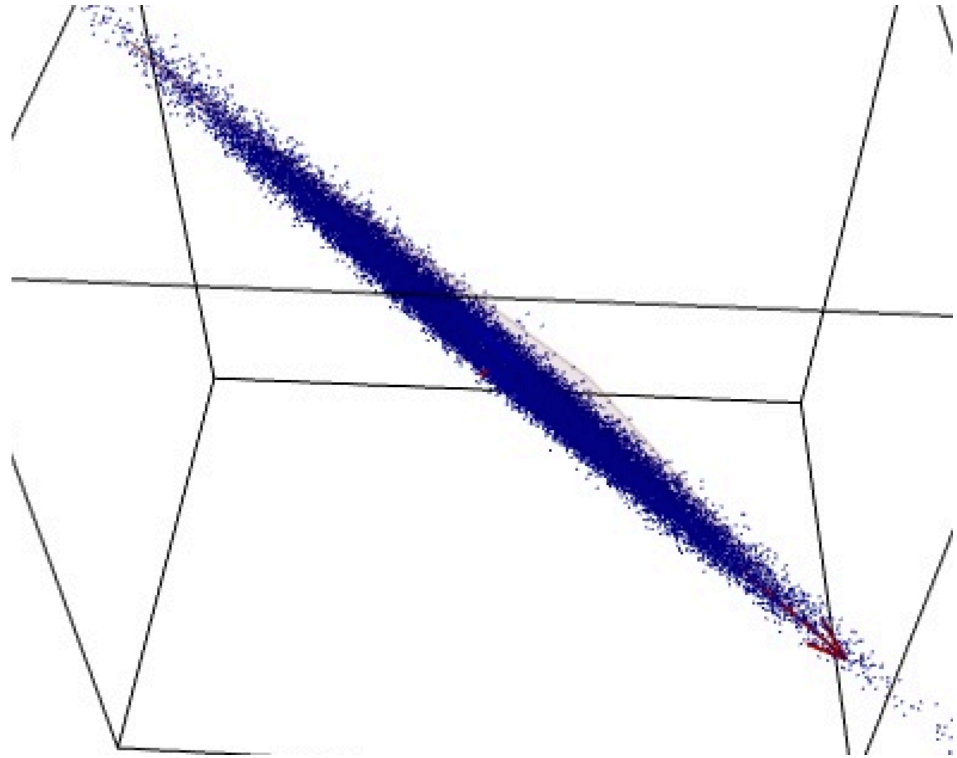


Degree of Data Variation in different directions



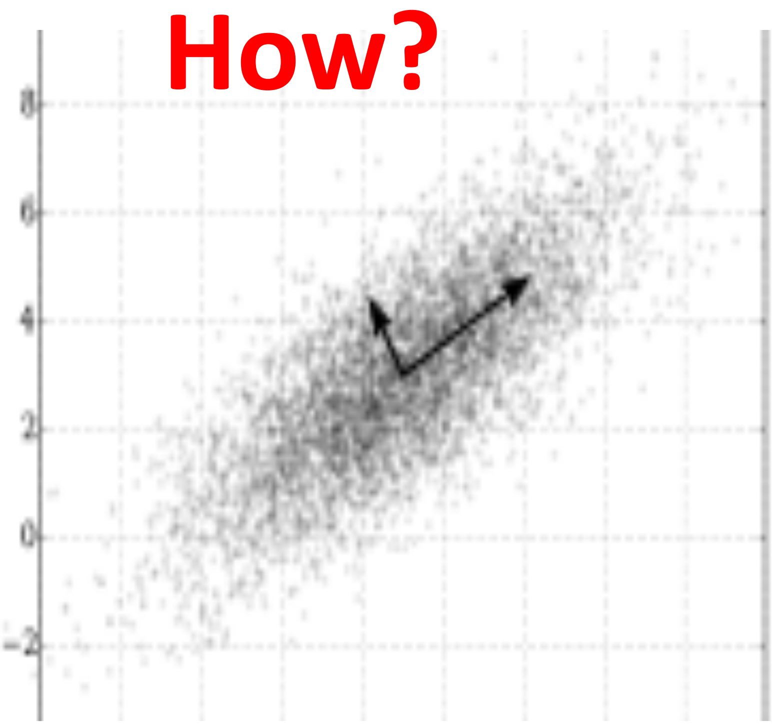
Similarly in Higher Dimensional





All the data can be projected to 1-dimensional line!

Geometrically, we want to find two axis directions of the elliptic curve. They are called **principal axes.**

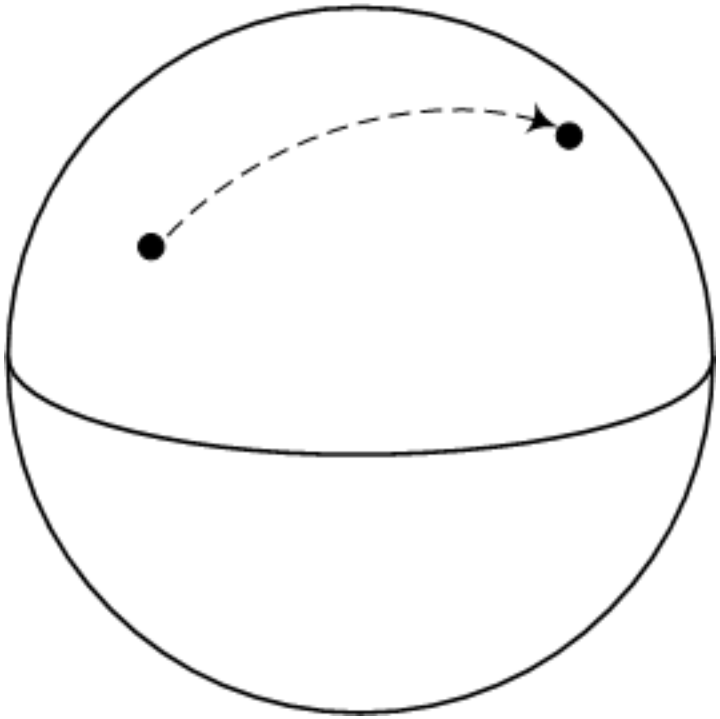


Note: the x-value and y-value of the data are correlated.
Their correlation are reflected by the covariance matrix of the data.

- **Principal component analysis** (**PCA**) is a procedure to find an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated.
- **Procedure of find the principal directions:**
- Step 1: Find the covariance matrix of the data directly (Note: one can first standardize data:
  - Find the mean μ_1 of the x-value & the mean μ_2 of the y-value.
  - Subtract all x-value by μ_1 and subtract all y-value by μ_2. *Geometrically, move the x-axis and y-axis to the data center*.)
- Step 2: Find eigenvalues and eigenvectors of the covariance matrix of the data.
- Step 3: Order the eigenvalues from largest to smallest. The eigenvector corresponding to the largest eigenvalue is called the 1st principal axis. So on and so forth.
- Step 4: Form the rotation matrix using the corresponding eigenvectors.

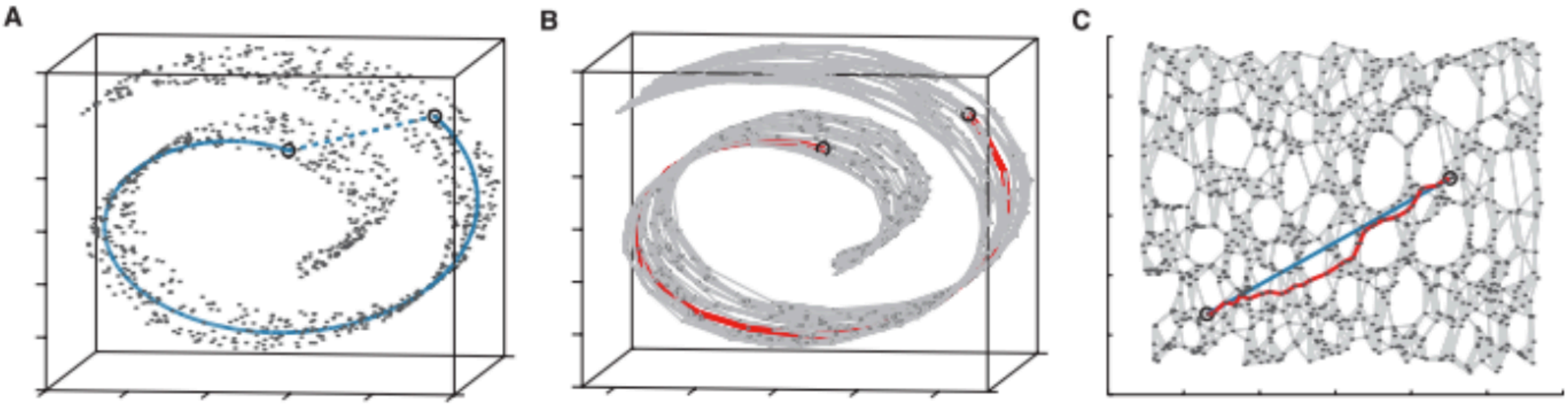
ISOMAP

- In ISOMAP, we replace Euclidean distance by Geodesic distance!



Note: A Geodesic on a manifold is usually not a linear segment. But if we cut a geodesic to very small pieces, each piece can be viewed as a line segment.

ISOMAP



Key Steps in ISOMAP

- **Step 1: Construct a k-nearest neighbor graph on the given n data points.**

(i.e. Choose a k, for each point in the data set, find the k nearest neighbors and then connect them using edges).

- **Step 2: Compute shortest path between all points as estimation of geodesic distance**. Denoted by $S_{G.}$

- **Step 3: Form $M = -1/2\, HS_GH$. Find largest p eigenvalues of M and their corresponding eigenvectors** $\{v_1, \ldots, v_p\}$.

Where H is the centering matrix,

$H = I - 1/n\, ee^T$ with $e^T = (1, 1, \ldots 1)$.

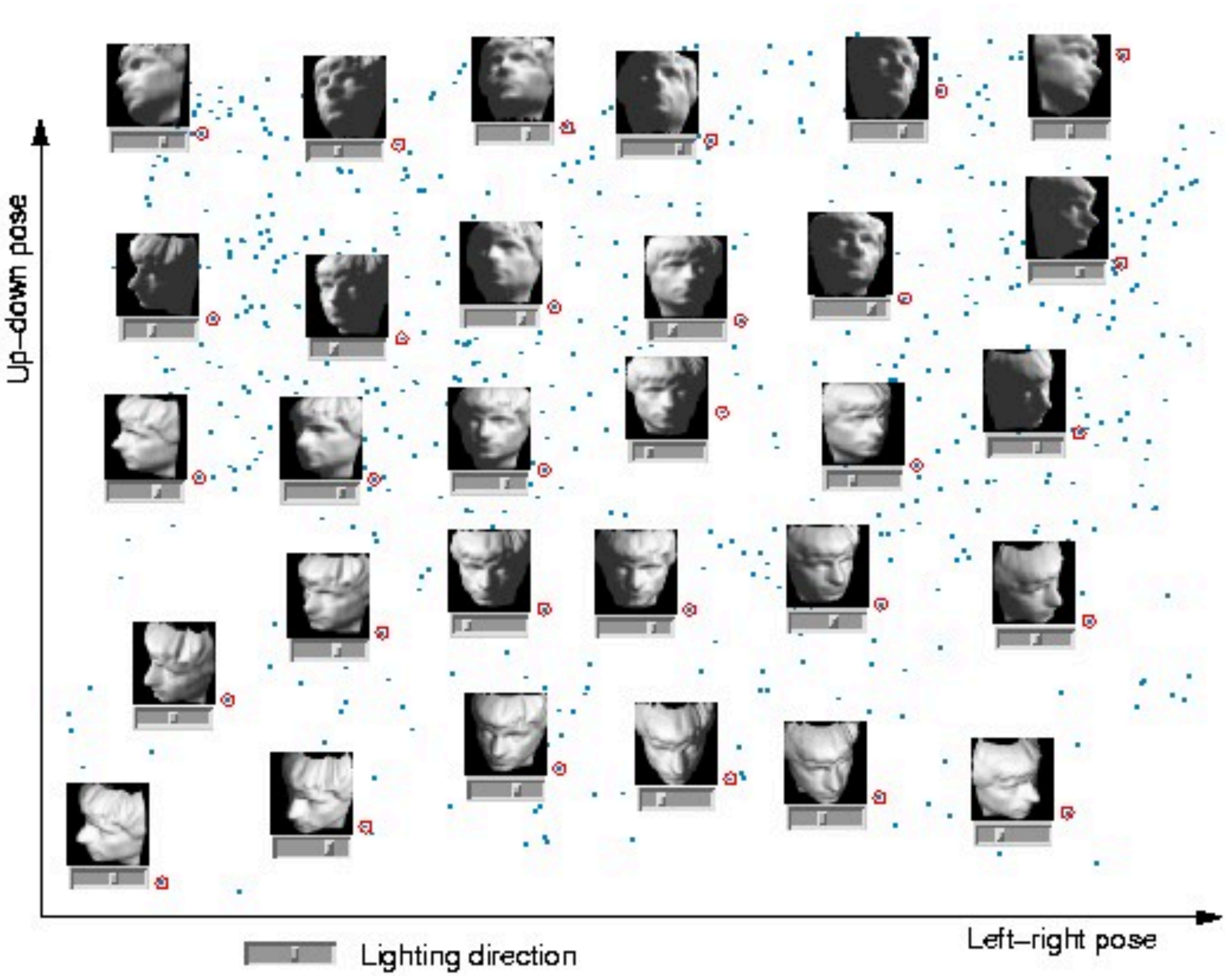
- **Step 4: Output Solution $Y = D^{1/2} V^{T}$,** where D is the diagonal matrix with the top p eigenvalues of M and $V = (v_1, \ldots, v_p)$ is the matrix of the corresponding eigenvectors.

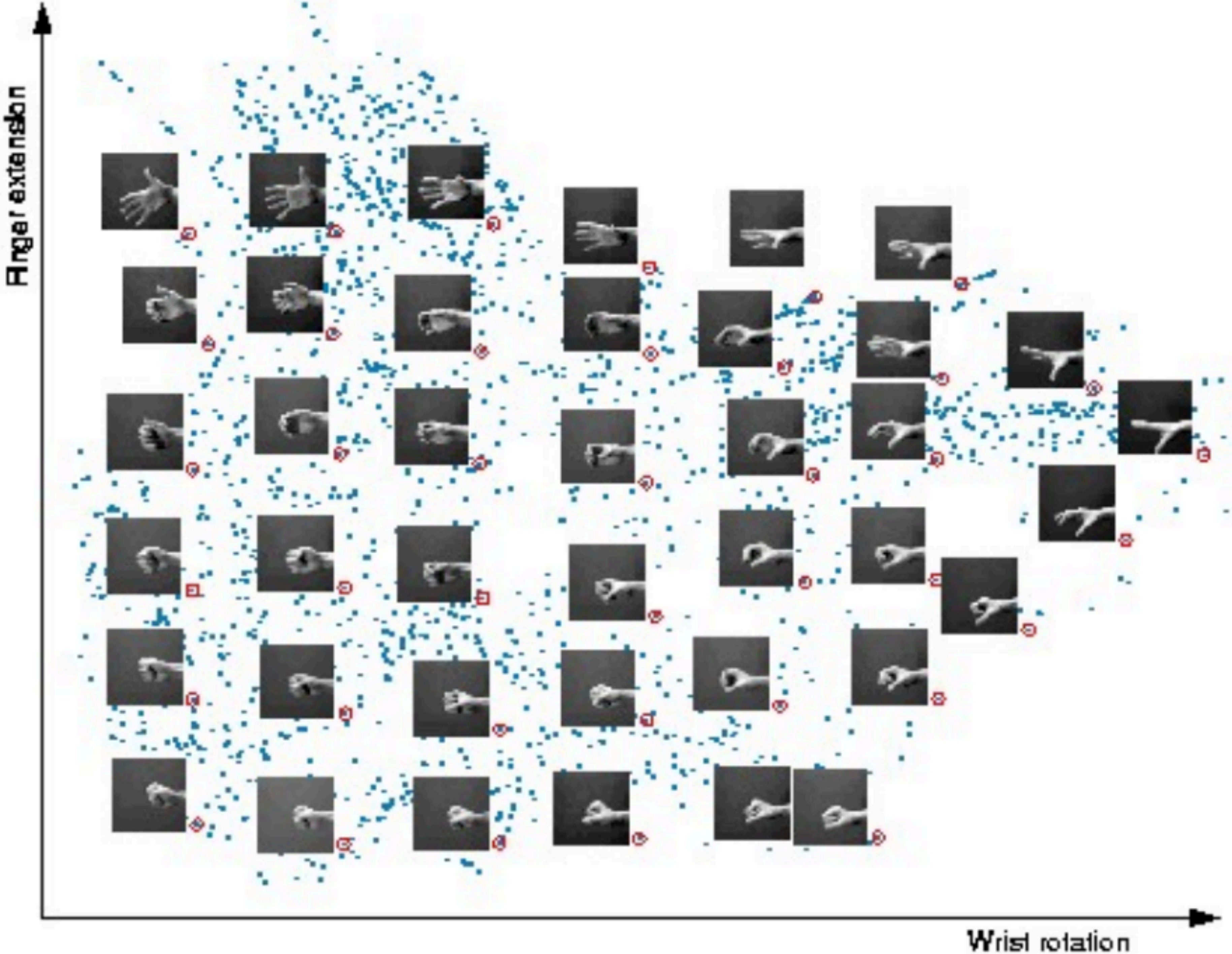
Subtlety

- You always can form the matrix M, but M may not be semi-definite matrix.
- What's wrong?
- We can not decompose M into $X^T X$.
- Means there does not exist such configuration approximate the data.
- Means there does not exist such an embedding. i.e. you can not un-roll the data to a low dimensional manifold.)

Note: People try to find a semi-definite matrix as close as possible to M (e.g. just get rid off the negative eigenvalues).

Images data: Taken using 3 degree of freedom. Unrolled to 2-D manifold: Rotating parallel to the xy-plane, also along z-axis




Finger extension
Wrist rotation

- Back up slides

- PCA
- MDS (multi dimensional scaling)
- ISOMAP
- https://www.youtube.com/watch?v=RPjPLlGefz