# MDS and PCA


Lecture 1.

Yuan Yao

Peking Univesity

# Geometric Embedding

- A Fundamental Problem in Data Representation
- Unstructured data -> Euclidean Space
- a.k.a. `feature' learning (e.g. deep learning)
- speech, text, image, video…

# Multidimensional Scaling

- Given pairwise distances between data points, can we find a system of Euclidean coordinates for those points whose pairwise distances meet given constraints?

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | BOST | NY | DC | MIAM | CHIC | SEAT | SF | LA | DENV |
| 1 | BOSTON | 0 | 206 | 429 | 1504 | 963 | 2976 | 3095 | 2979 | 1949 |
| 2 | NY | 206 | 0 | 233 | 1308 | 802 | 2815 | 2934 | 2786 | 1771 |
| 3 | DC | 429 | 233 | 0 | 1075 | 671 | 2684 | 2799 | 2631 | 1616 |
| 4 | MIAMI | 1504 | 1308 | 1075 | 0 | 1329 | 3273 | 3053 | 2687 | 2037 |
| 5 | CHICAGO | 963 | 802 | 671 | 1329 | 0 | 2013 | 2142 | 2054 | 996 |
| 6 | SEATTLE | 2976 | 2815 | 2684 | 3273 | 2013 | 0 | 808 | 1131 | 1307 |
| 7 | SF | 3095 | 2934 | 2799 | 3053 | 2142 | 808 | 0 | 379 | 1235 |
| 8 | LA | 2979 | 2786 | 2631 | 2687 | 2054 | 1131 | 379 | 0 | 1059 |
| 9 | DENVER | 1949 | 1771 | 1616 | 2037 | 996 | 1307 | 1235 | 1059 | 0 |

# Isometric Euclidean Embedding

Find an Euclidean embedding of low total metric distortion:

$$\min_{Y_i \in \mathbb{R}^k} \sum_{i,j} (\|Y_i - Y_j\|^2 - d_{ij}^2)^2$$

take the derivative w.r.t $Y_i \in \mathbb{R}^k$:

$$\sum_{i,j} (\|Y_i\|^2 + \|Y_j\|^2 - 2Y_i^T Y_j - d_{ij}^2)(Y_i - Y_j) = 0$$

which implies $\sum_i Y_i = \sum_j Y_j$. For simplicity set $\sum_i Y_i = 0$, *i.e.*putting the origin as data center.

Use a linear transformation to move the sample mean to be the origin of the coordinates, *i.e.* define a matrix $B_{ij} = -\frac{1}{2}HDH$ where $D = (d_{ij}^2)$, $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$, then, the minimization (1) is equivalent to find $Y_i \in \mathbb{R}^k$:

$$\min \|Y^T Y - B\|_F^2$$

then the row vectors of matrix $Y$ are the eigenvectors corresponding to $k$ largest eigenvalues of $B = \widetilde{X}^T \widetilde{X}$, or equivalently the top $k$ *right singular vectors* of $\widetilde{X} = USV^T$.

B is Gram matrix or kernel matrix

**Algorithm 1:** Classical MDS Algorithm

**Input**: A squared distance matrix $D^{n \times n}$ with $D_{ij} = d_{ij}^2$.

**Output**: Euclidean $k$-dimensional coordinates $\widetilde{X}_k \in \mathbb{R}^{k \times n}$ of data.

1 Compute $B = -\dfrac{1}{2} H \cdot D \cdot H^T$, where H is a centering matrix.

2 Compute Eigenvalue decomposition $B = U \Lambda U^T$ with $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$;

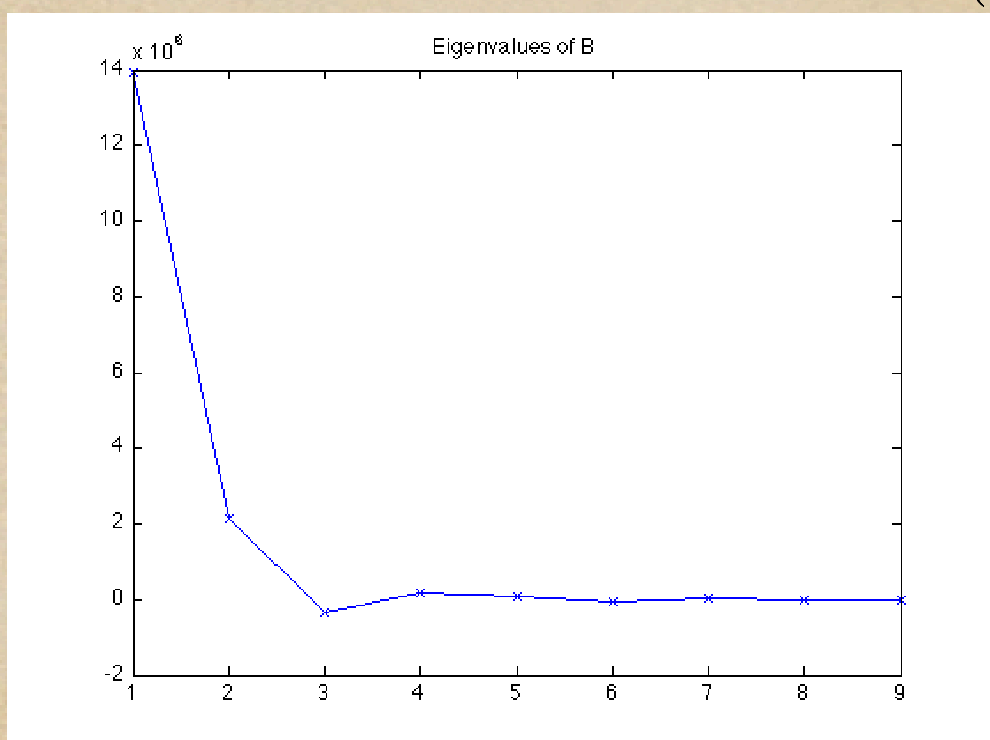
3 Choose top $k$ nonzero eigenvalues and corresponding eigenvectors, $\widetilde{X}_k = U_k \Lambda_k{}^{\frac{1}{2}}$ where

$$U_k = [u_1, \ldots, u_k], \quad u_k \in \mathbb{R}^n,$$
$$\Lambda_k = \text{diag}(\lambda_1, \ldots, \lambda_k)$$

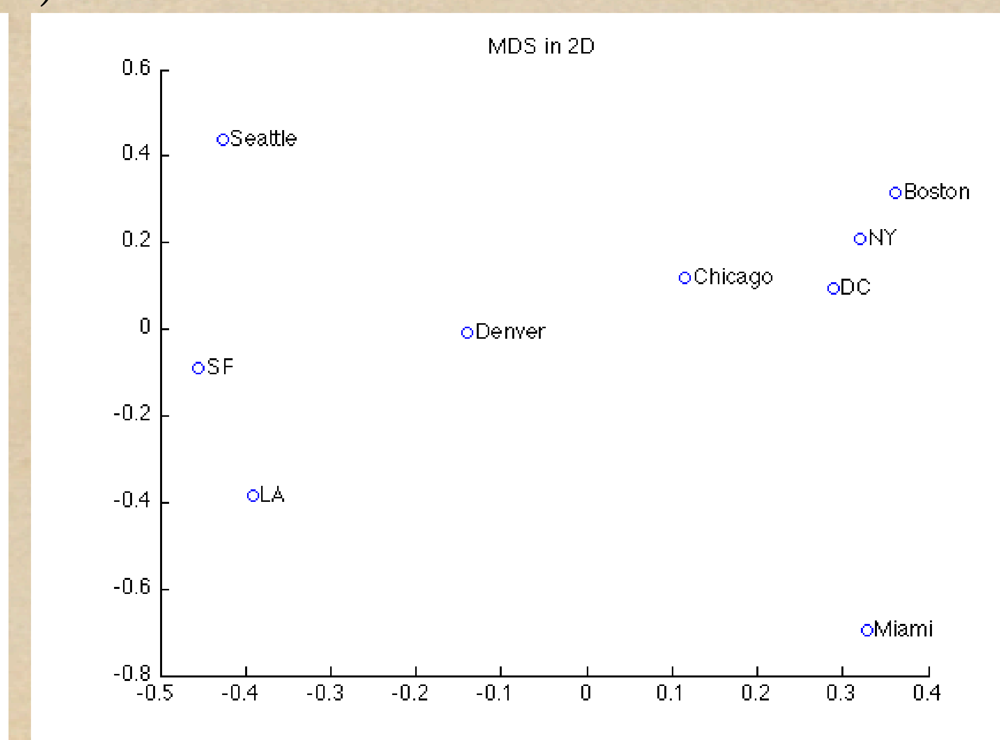
with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$.

| | | 1 BOST | 2 NY | 3 DC | 4 MIAM | 5 CHIC | 6 SEAT | 7 SF | 8 LA | 9 DENV |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | BOSTON | 0 | 206 | 429 | 1504 | 963 | 2976 | 3095 | 2979 | 1949 |
| 2 | NY | 206 | 0 | 233 | 1308 | 802 | 2815 | 2934 | 2786 | 1771 |
| 3 | DC | 429 | 233 | 0 | 1075 | 671 | 2684 | 2799 | 2631 | 1616 |
| 4 | MIAMI | 1504 | 1308 | 1075 | 0 | 1329 | 3273 | 3053 | 2687 | 2037 |
| 5 | CHICAGO | 963 | 802 | 671 | 1329 | 0 | 2013 | 2142 | 2054 | 996 |
| 6 | SEATTLE | 2976 | 2815 | 2684 | 3273 | 2013 | 0 | 808 | 1131 | 1307 |
| 7 | SF | 3095 | 2934 | 2799 | 3053 | 2142 | 808 | 0 | 379 | 1235 |
| 8 | LA | 2979 | 2786 | 2631 | 2687 | 2054 | 1131 | 379 | 0 | 1059 |
| 9 | DENVER | 1949 | 1771 | 1616 | 2037 | 996 | 1307 | 1235 | 1059 | 0 |

(a)




(b)




(c)

# Inverse problem:when it is an Exact Isometry Embedding?

Consider a forward problem: given a set of points $x_1, x_2, ..., x_n \in \mathbb{R}^p$, let

$$X = [x_1, x_2, ..., x_n]^{p \times n}.$$

The distance between point $x_i$ and $x_j$ satisfies

$$d_{ij}^2 = \|x_i - x_j\|^2 = (x_i - x_j)^T (x_i - x_j) = {x_i}^T x_i + {x_j}^T x_j - 2{x_i}^T x_j.$$

# Inner Product (kernel) Matrix vs. Squared Distance Matrix

Let $K$ be the inner product matrix

$$K = X^T X,$$

with $k = \mathrm{diag}(K_{ii}) \in \mathbb{R}^n$. So

$$D = (d_{ij}^2) = k \cdot \mathbf{1}^T + \mathbf{1} \cdot k^T - 2K.$$

where $\mathbf{1} = (1, 1, ..., 1)^T \in \mathbb{R}^n$.

# Centered the data

Define the mean and the centered data

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\cdot X \cdot \mathbf{1},$$

$$\widetilde{x}_i = x_i - \widehat{\mu}_n = x_i - \frac{1}{n}\cdot X \cdot \mathbf{1},$$

$$\widetilde{X} = X - \frac{1}{n} X \cdot \mathbf{1} \cdot \mathbf{1}^T.$$

$$\begin{aligned}
\tilde{K} &\triangleq \tilde{X}^T \tilde{X} \\
&= \left(X - \frac{1}{n} X \cdot \mathbf{1} \cdot \mathbf{1}^T\right)^T \left(X - \frac{1}{n} X \cdot \mathbf{1} \cdot \mathbf{1}^T\right) \\
&= K - \frac{1}{n} K \cdot \mathbf{1} \cdot \mathbf{1}^T - \frac{1}{n} \mathbf{1} \cdot \mathbf{1}^T \cdot K + \frac{1}{n^2} \cdot \mathbf{1} \cdot \mathbf{1}^T \cdot K \cdot \mathbf{1} \ \cdot \mathbf{1}^T
\end{aligned}$$

Let

$$B = -\frac{1}{2} H \cdot D \cdot H^T$$

where $H = I - \frac{1}{n} \cdot \mathbf{1} \cdot \mathbf{1}^T$. H is called as a *centering matrix.*

$$B = -\frac{1}{2} H \cdot (k \cdot \mathbf{1}^T + \mathbf{1} \cdot k^T - 2K) \cdot H^T$$

Since $k \cdot \mathbf{1}^T \cdot H^T = k \cdot \mathbf{1}(I - \frac{1}{n} \cdot \mathbf{1} \cdot \mathbf{1}^T) = k \cdot \mathbf{1} - k(\frac{\mathbf{1}^T \cdot \mathbf{1}}{n}) \cdot \mathbf{1} = 0$, we have $H \cdot k\ \mathbf{1} \cdot H^T = H \cdot \mathbf{1} \cdot k^T \cdot H^T = 0$.

Therefore,

$$\begin{aligned} B &= H \cdot K \cdot H^T = (I - \frac{1}{n} \cdot \mathbf{1} \cdot \mathbf{1}^T) \cdot K \cdot (I - \frac{1}{n} \cdot \mathbf{1} \cdot \mathbf{1}^T) \\ &= K - \frac{1}{n} \cdot \mathbf{1} \cdot \mathbf{1} \cdot K - \frac{1}{n} \cdot K \cdot \mathbf{1} \cdot \mathbf{1}^T + \frac{1}{n^2} \cdot \mathbf{1}(\mathbf{1}^T \cdot K\mathbf{1}) \cdot \mathbf{1}^T \\ &= \tilde{K}. \end{aligned}$$

# Gram/Kernel matrix

$$B = -\frac{1}{2} H \cdot D \cdot H^T = \tilde{X}^T \tilde{X}.$$

Note that often we define the covariance matrix

$$\widehat{\Sigma}_n \triangleq \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \widehat{\mu}_n)(x_i - \widehat{\mu}_n)^T = \frac{1}{n-1} \widetilde{X} \widetilde{X}^T.$$

# P.S.D. as inner product

**Definition** (Positive Semi-definite)**.** Suppose $A^{n\times n}$ is a real symetric marix,then: $A$ is p.s.d.(positive semi-definite)($A \succeq 0$) $\iff \forall v \in \mathbb{R}^n, v^T Av \geq 0 \iff A = Y^T Y$

**Property.** Suppose $A^{n\times n}$, $B^{n\times n}$ are real symetric marix, $A \succeq 0$, $B \succeq 0$. Then we have:

$$\begin{aligned}&(1)\quad A + B \succeq 0;\\&(2)\quad A \circ B \succeq 0;\end{aligned}$$

where $A \circ B$ is called Hadamard product and $(A \circ B)_{i,j} := A_{i,j} \times B_{i,j}$.

# C.N.D.

**Definition** (Conditionally Negative Definite)**.** Suppose $A^{n\times n}$ is a real symetric marix,then:
$A$ is c.n.d.(conditionally negative definite) $\iff \forall v \in \mathbb{R}^n, \mathbf{1}v^T = \sum_{i=1}^n v_i = 0$, we have $v^T A v \leq 0$

**Lemma 2.1** (Young/Househölder-Schoenberg '1938)**.** For any signed probability measure $\alpha$ $(\alpha \in \mathbb{R}^n, \sum_{i=1}^n \alpha_i = 1)$,

$$B_\alpha = -\frac{1}{2} H_\alpha C H_\alpha^T \succeq 0 \iff C \text{ is c.n.d.}$$

where $H_\alpha$ is Householder centering matrix: $H_\alpha = \mathbf{I} - \mathbf{1} \cdot \alpha^T$.