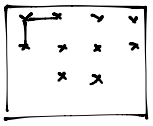
Multidimensional scaling

Distance matrix (not necessarily observed)

$D = (d_{ij})$ where $d_{ij} = d(\underset{\sim}{x}_i, \underset{\sim}{x}_j)$ $i,j = 1, \cdots, n$

$n \times n$ symmetric

- want to represent (possibly unobserved) data in low dimensions



distances in 2D plot close to those in distance matrix

Observe $X = \begin{pmatrix} \underset{\sim}{x}_1^T \\ \vdots \\ \underset{\sim}{x}_n^T \end{pmatrix}$ ($n \times p$)

Euclidean distances $d_{ij} = d(\underset{\sim}{x}_i, \underset{\sim}{x}_j) = [(\underset{\sim}{x}_i - \underset{\sim}{x}_j)^T (\underset{\sim}{x}_i - \underset{\sim}{x}_j)]^{\frac{1}{2}}$

Recover $\{d_{ij}\}$ from $B = XX^T = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix}$ ($n \times n$)

$d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$

SVD: $X = U \Lambda^{\frac{1}{2}} V^T$ ($n \times p$)

$XX^T = U \Lambda U^T$ $\qquad U = \begin{pmatrix} \underset{\sim}{u}_1 & \cdots & \underset{\sim}{u}_p \end{pmatrix}$

- optimal k dimensional representation of points $(\lambda_1^{\frac{1}{2}} \underset{\sim}{u}_1 \ \lambda_2^{\frac{1}{2}} \underset{\sim}{u}_2 \cdots \lambda_k^{\frac{1}{2}} \underset{\sim}{u}_k)$

Now the real problem: Given distance (similarity) matrix, how to to get low dimensional representation for data?

Idea: Use $\{d_{ij}\}$ to construct matrix B.

$b_{ij} = -\frac{1}{2}\{d_{ij}^2 - d_{i\cdot}^2 - d_{\cdot j}^2 + d_{\cdot\cdot}^2\}$

where $d_{i\cdot}^2 = \frac{1}{n}\sum_{j=1}^n d_{ij}^2$

$d_{\cdot j}^2 = \frac{1}{n}\sum_{i=1}^n d_{ij}^2$

$d_{\cdot\cdot}^2 = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n d_{ij}^2$

B ($n \times n$) is symmetric and so we can write

$B = U \Lambda U^T$ ($n \times n$, $n \times n$, $n \times n$) $\qquad \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ eigenvalues

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ ordered

- if all eigenvalues: $\lambda_1, \cdots, \lambda_n$ are non-negative then $\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_n}$ gives an idea of how good the k dim representation is.

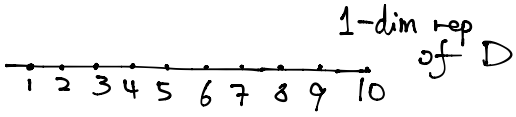
$k=2$ Plot $\lambda_1^{\frac{1}{2}} u_1$ vs $\lambda_2^{\frac{1}{2}} u_2$

Example

Start with very simple distance matrix

$n=10$ $D=(d_{ij})$ $d_{ij}=|i-j|$

$$=\begin{pmatrix} 0 & 1 & \cdots & 9 \\ 1 & 0 & 1 & \cdots & 8 \\ \vdots & & \ddots & & \vdots \\ 9 & \cdots & & & 0 \end{pmatrix}$$



What happens to MDS using D? R cmdscale

- intuition suggests that $\lambda_1>0$, $\lambda_2=\cdots=\lambda_{10}=0$

Output: $\lambda_1=82.5$, $\lambda_2=\cdots=\lambda_{10}=0$

$$\lambda_1^{\frac{1}{2}}\underset{\sim}{u}_1=\begin{pmatrix} -4.5 \\ -3.5 \\ \vdots \\ 4.5 \end{pmatrix}$$

Now add noise to D ← symmetric elements 0+1

$D_1=D+N$

- order of distances remaining unchanged for each row (& column.)

cmdscale: $\lambda_1=92.5$, $\lambda_2=5.87$, $\lambda_3=2.54$

$\lambda_4=1.69$, $\lambda_5=1.35$, $\lambda_6,\cdots,\lambda_{10}<6$

$$\lambda_1^{\frac{1}{2}}\underset{\sim}{u}=\begin{pmatrix} -4.79 \\ -3.63 \\ -2.64 \\ -1.66 \\ -0.68 \\ 0.66 \\ 1.71 \\ 2.62 \\ 3.64 \\ 4.76 \end{pmatrix} \qquad \lambda_2^{\frac{1}{2}}\underset{\sim}{u}=\begin{pmatrix} 0.96 \\ 0.58 \\ 0.18 \\ -0.62 \\ -1.16 \\ -1.26 \\ -0.18 \\ -1.06 \\ 0.71 \\ 1.35 \end{pmatrix}$$

Independent Components Analysis ICA

- Cousin of PCA
- Theory of PCA, $\underset{\sim}{X}$ random vector with covariance matrix C, mean vector $\underset{\sim}{\mu}$
- Write $\underset{\sim}{X}=\underset{\sim}{\mu}+A\underset{\sim}{Y}$ where the components of $\underset{\sim}{Y}(Y_1,\cdots Y_p)$ are uncorrelated $\text{Cov}(Y_i,Y_j)=0$ with mean 0 & variance 1.

But $\underset{\sim}{Y}$ is not uniquely determined:

$\text{Cov}(\underset{\sim}{Y})=I \Rightarrow$ if Q is orthogonal then $\text{Cov}(Q\underset{\sim}{Y})=I$

Look at C and choose A based on eigenvalues and eigenvectors of C.

$\Rightarrow$ not invariant to choice of scale.

Data analysis: Use data to estimate C (or R) and use these to estimate PC loadings

ICA: $\underset{\sim}{X}$ mean vector $\underset{\sim}{\mu}$, covariance matrix C

Model: Write $\underset{\sim}{X}=\underset{\sim}{\mu}+A\underset{\sim}{Y}$ where the components of $\underset{\sim}{Y}(Y_1,\cdots,Y_p)$ are mutually independent

↑ mixing matrix

with $E(Y_i)=0$, $\text{Var}(Y_i)=1$.

Model is invariant to linear transformations

$$\underset{\sim}{X}' = \underset{\sim}{a} + B\underset{\sim}{X} = \underset{\sim}{a} + B\underset{\sim}{\mu} + BA\underset{\sim}{Y}$$

(A: mixing matrix)

Key assumption: $\underset{\sim}{X}$ is not multivariate normal and at most one component of $\underset{\sim}{Y}$ is normal.

- non-unique of mixing matrix A.

If $\underset{\sim}{X}$ is multivariate normal then ICA = PCA

$$\text{ICA: } \underset{\sim}{X} = \underset{\sim}{\mu} + A\underset{\sim}{Y} \longrightarrow \text{indpdt components}$$

(A: mixing matrix)

Given data $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$, how to estimate A?

Two steps

① "Prewhitening": Transformed centred data $\underset{\sim}{x}_1 - \bar{\underset{\sim}{x}}, \cdots, \underset{\sim}{x}_n - \bar{\underset{\sim}{x}}$ to make variables uncorrelated (with variance 1)

$$\underset{\sim}{z}_i = L(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}}) \quad \text{with } \frac{1}{n-1}\sum_{i=1}^{n} \underset{\sim}{z}_i \underset{\sim}{z}_i^T = I \quad \text{e.g. PCs.}$$

(L: $p \times p$)

② Component extraction

Find one-dimensional projection $\underset{\sim}{w}_1, \cdots, \underset{\sim}{w}_p$ with $\underset{\sim}{w}_i^T \underset{\sim}{w}_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$

and define $W = \begin{pmatrix} \underset{\sim}{w}_1^T \\ \vdots \\ \underset{\sim}{w}_p^T \end{pmatrix}$

want components to be independent $\longrightarrow$

$$\underset{\sim}{y}_i = W\underset{\sim}{z}_i = WL(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})$$

$$\underset{\sim}{x}_i = \bar{\underset{\sim}{x}} + (WL)^{-1}\underset{\sim}{y}_i = \bar{\underset{\sim}{x}} + L^{-1}W^{-1}\underset{\sim}{y}_i$$

How to do steps ① + ②:

Look Fast ICA algorithm:

① can be done using PCA:

$$\hat{C} = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})^T = V\Lambda V^T$$

$$\underset{\sim}{z}_i = \Lambda^{-\frac{1}{2}} V^T(\underset{\sim}{x} - \bar{\underset{\sim}{x}})$$

Easy to verify $\frac{1}{n-1}\sum_{i=1}^{n} \underset{\sim}{z}_i \underset{\sim}{z}_i^T = I$

② Find independent components using prewhitened data.

- find projections of prewhitened data that are as non-normal as possible

$\Rightarrow$ projection pursuit

Need indices of non-normality for a given projection $\underset{\sim}{w} \rightarrow J(\underset{\sim}{w})$

Examples ① $H(x) = \ln[\cosh(x)] = \ln\left[\frac{\exp(x)+\exp(-x)}{2}\right]$

$$H'(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$$

$$H''(x) = 1 - [H'(x)]^2$$

Define $J(\underset{\sim}{w}) = \frac{1}{n}\sum_{i=1}^{n} H(\underset{\sim}{w}^T \underset{\sim}{z}_i)$

② Kurtosis: Maximize $J(\underset{\sim}{w}) = \left|\frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{w}^T \underset{\sim}{z}_i)^4 - 3\right|$ ← Kurtosis for normal dist'n

Why try to maximize non-normality?

Look at ICA model

$\underset{\sim}{X} = \underset{\sim}{\mu} + A\underset{\sim}{Y}$ where $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix}$

$\underset{\sim}{Y} = A^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$

$\mathrm{Cov}(\underset{\sim}{Y}) = I$ and $\mathrm{Cov}(Q\underset{\sim}{Y}) = I$ for any orthogonal matrix $Q = \begin{pmatrix} q_{11} & \cdots & q_{1p} \\ \vdots & & \vdots \\ q_{p1} & \cdots & q_{pp} \end{pmatrix}$ ($p \times p$)

$\underset{\sim}{S} = Q\underset{\sim}{Y} = \begin{pmatrix} S_1 \\ \vdots \\ S_p \end{pmatrix}$

What do we know?

① $S_1, \cdots, S_p$ are uncorrelated with mean 0 & variance 1

② $S_i = \underbrace{q_{i1}Y_1 + q_{i2}Y_2 + \cdots + q_{ip}Y_p}$ for $i \in [1, p]$

addition of $k \geq 2$ independent r.v.s

Central Limit Thm: "Sums of independent r.v.s are closer to normal than each of the summands."

$S_1, \cdots, S_p$ closer to normal than $Y_1, \cdots, Y_p$

$\Rightarrow$ maximizing non-normality of projections is reasonable.