Multidimensional Scaling

consider: data matrix $X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$

$$\begin{array}{ccc} & X & \\ \swarrow & & \searrow^{PCA} \\ \text{Proximity matrix } D & & Y \end{array}$$

Proximity matrix D.

represents "distance".

similarity or dissimilarity.

Rmk: Proximity matrix is invariant in

i) Translation ii) rotation iii) reflection.

So we can't recover X completely.

Def: A proximity matrix is

i) distance-like if $d_{ij} \geq 0$. $d_{ii} = 0$. $d_{ij} = d_{ji}$

ii) metric if it's distance-like and $d_{ij} \leq d_{ik} + d_{jk}$.

iii) Euclidean if exists a configuration of points in Euclidean space with distance (Point i, Point j) $= d_{ij}$.

(1) Classical MDS:

suppose proximity matrix $D_{n\times n}$ is of squared Euclidean distance from $X_{n\times q}$. data matrix.

<u>Denote:</u> $B = XX^T$. $b_{ij} = \sum_{k=1}^{q} x_{ik} x_{jk}$.

$\Rightarrow d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}. = (x_{(i)} - x_{(j)})^T (x_{(i)} - x_{(j)})$

<u>Ideal:</u> If b_{ij}'s can be found in term of d_{ij}. Then. we can derive X from B.

To obtain B from D. We need some location constraint for unique solution.

suppose: $\sum_{i=1}^{n} x_{ik} = 0$. $\forall\, 1 \le k \le q$. (center of $c(x)$ at 0)

$\Rightarrow \sum_{j=1}^{n} b_{ij} = 0$. i.e. row of B have sum 0.

Notice: $\sum_{i=1}^{n} d_{ij}^2 = tr(B) + n b_{jj}$ $\quad \sum_{j=1}^{n} d_{ij}^2 = tr(B) + n b_{ii}$

$$\sum_i \sum_j d_{ij}^2 = 2n\,Tr(B).$$

$\Rightarrow b_{ij} = -\frac{1}{2} [d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2]$. where $d_{i.}^2 = \sum_j d_{ij}^2/n$. $d_{.j}^2 = \sum_i d_{ij}^2/n$. $d_{..}^2 = \sum_i \sum_j d_{ij}^2/n^2$.

Then. we can obtain X from B:

by spectral decomposition: $B = V \Lambda V^T$.

1°) when $q < n$. $B = (V^* \; *) \begin{pmatrix} \Lambda^* & \\ & 0 \end{pmatrix} \begin{pmatrix} V^{*T} \\ * \end{pmatrix}$

so $X = V^* \Lambda^{*\frac{1}{2}}$.

2) When $q \geq n$.

Choose the first k components. s.t. $\frac{\sum_1^k \lambda_i}{\sum_1^n \lambda_i}$ is large enough.

Rmk: When proximity matrix isn't Euclidean.

Then B may not be positive definite.

Consider criteria: $\sum_1^k |\lambda_i| / \sum_1^n |\lambda_i|$.

(二) Other methods:

i) Metric Scaling: Define loss func. for D and distance matrix based on X. called stress. then find X to minimize stress.

ii) Non-metric Scaling: Only trust orders of D rather than its values.

iii) Asymmetric proximity matrix: $d_{ij} \neq d_{ji}$.