Principle Components

$\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$ { n obs, p variables

$X = \begin{pmatrix} \underset{\sim}{x}_1^T \\ \vdots \\ \underset{\sim}{x}_n^T \end{pmatrix}$ ($n \times p$) $\xrightarrow[\text{scale}]{\text{centre}} \tilde{X}$

PCA: Transform original data to obtain p uncorrelated variables

SVD: $\tilde{X} = UDV^T$ ($p \times p$)

($n \times p$, $p \times p$ diagonal)

$(U^TU = I = V^TV)$

$$\tilde{X}V = UD = \begin{pmatrix} \underset{\sim}{u}_1 & \cdots & \underset{\sim}{u}_p \end{pmatrix} \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{pmatrix} = \begin{pmatrix} d_1\underset{\sim}{u}_1 & \cdots & d_p\underset{\sim}{u}_p \end{pmatrix} \qquad d_1 \geq d_2 \geq \cdots d_p > 0$$

($d_1\underset{\sim}{u}_1$, $d_p\underset{\sim}{u}_p$: principal component scores)

Covariance matrix

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

$$= \frac{1}{n-1}\tilde{X}^T\tilde{X}$$

$$S = V\Lambda V^T \quad (\Lambda = \frac{D^2}{n-1})$$

- 1st PC maximizes $\underset{\sim}{a}^T S \underset{\sim}{a}$ s.t. $\underset{\sim}{a}^T\underset{\sim}{a} = 1$
  $\Rightarrow \underset{\sim}{a} = \underset{\sim}{v}_1 \quad V = (\underset{\sim}{v}_1, \cdots, \underset{\sim}{v}_p)$
- kth PC maximizes $\underset{\sim}{a}^T S \underset{\sim}{a}$ s.t. $\underset{\sim}{a}^T\underset{\sim}{a} = 1$ and $\underset{\sim}{a}^T\underset{\sim}{v}_j = 0$ for $j = 1, \cdots, k-1$
- Principal component scores:
  columns of $\tilde{X}V$
- Principal components loadings:
  columns of $V = (\underset{\sim}{v}_1, \cdots, \underset{\sim}{v}_p)$

Hopefully, PC { scores, loadings } are easily interpretable

Example: Atheletics records (Blackboard)
8 records
55 countries

PCA on correlation matrix ($\tilde{X}$ centred & scaled)

$\lambda_1 = 0.79$
$\lambda_2 = 0.11$ } $\lambda_1 + \lambda_2 = 0.90$
$\lambda_3 = 0.04$ } $\lambda_1 + \lambda_2 + \lambda_3 = 0.94$

Loadings	PC # 1	PC #2
100m	-0.33	-0.54
200m	-0.34	-0.47
400m	-0.36	-0.25
800m	-0.38	0
1500m	-0.39	0.13
5000m	-0.37	0.31
10000m	-0.33	0.35
marathon	-0.35	0.47

PC#1 is a measure of overall strength
PC#2 is a measure of contrast b/w distance & shorter events.

Biplot

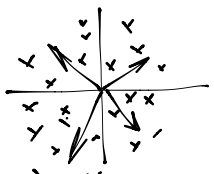
- Very useful plot. Plot 1st PC scores versus 2nd PC scores
- Biplot: adds variable information.

Rank 2 approx to $\tilde{X}$ (largest two singular values)

$$\tilde{X}_2 = (\underset{\text{scores}}{\underset{\uparrow}{\underset{n\times 2}{\uparrow}}}\ \underset{\uparrow}{\underset{\text{scores}}{\underset{\uparrow}{u_1}}}, \underset{\uparrow}{\underset{\text{scores}}{\underset{\uparrow}{u_2}}}) \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} v_1^T \\ v_2^T \end{pmatrix} \leftarrow 2\times p$$

$$(\tilde{X}v_1, \tilde{X}v_2) = (d_1 u_1 \quad d_2 u_2) \quad \text{1st two PC scores}$$

- Plot u_1 vs u_2 (plot of 1st 2 PC scores)
- $\underbrace{(d_1 v_1 \ d_2 v_2)}_{p\times 2}$ represent rows variables of this matrix by vectors
  - length of vector represent variance
  - angles between vectors represent correlation.



$\tilde{X}v = \begin{pmatrix} & \\ & \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}$ ← loadings

Multidimensional scaling

Problem: $x_1, \dots, x_n$ not necessarily observed
But ... given an $n\times n$ symmetric matrix of distances $d(x_i, x_j)$

Find $y_1, \dots, y_n$ (low dimensional) such that $d(y_i, y_j) \doteq d(x_i, x_j)$
Define $D = (d_{ij} = d(x_i, x_j))$

Special Case: Assume $\underbrace{x_1, \dots, x_n}_{\text{centred \& scaled}}$ known and $d(x_i, x_j)$ = Euclidean distance

$$d(u,v) = \left[\sum_{j=1}^{p} (u_j - v_j)^2\right]^{\frac{1}{p}}$$

Define $X = \underset{n\times p}{\begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}}$ and $B = \underbrace{XX^T}_{n\times n}$

Claim: Can recover $d_{ij} = d(x_i, x_j)$ from elements of B.

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \qquad d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$$

Moreover, if we rotate each row of X then B & D remain unchanged: $X\,OO^T\,X^T = B$ (O: orthogonal)

SVD: $X = U \Lambda^{\frac{1}{2}} V^T$

$X X^T = U \Lambda^{\frac{1}{2}} V^T V \Lambda^{\frac{1}{2}} U^T$

$V^T V = I$

$= U \Lambda U^T$

U: $n \times p$, Λ: $p \times p$, U^T: $p \times n$

$\sim$ result is $n \times n$

- suggests that best r dimensional distance preserving transformation is first r columns of U