Principal Component Analysis (PCA)

$\underset{\sim}{X} = (X_1, \ldots, X_p)'$

$$\Sigma = \mathrm{Cov}(\underline{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ & \sigma_{22} & \cdots & \sigma_{2p} \\ & & \ddots & \\ & & & \sigma_{pp} \end{pmatrix} = ((\sigma_{ij}))$$

Total variation in $\underset{\sim}{X} = \mathrm{tr}\,\Sigma = \sum_{i=1}^{p} \sigma_{ii}$

Aim of PCA

Replace $\underset{\sim}{X}$ with $\underset{\sim}{Y} = (Y_1, \ldots, Y_p)'$, where Y_i s are

are linear combinations of X_i s $\ni$

(i) Y_i s are uncorrelated, i.e.

$$\mathrm{Cov}(Y_i, Y_j) = 0 \quad \forall\, i \neq j$$

(ii) total variation in $\underset{\sim}{X}$

= total variation in $\underset{\sim}{Y}$

& (iii) total variation in $(Y_1, \ldots, Y_k)'$

$\approxeq$ total variation of $\underset{\sim}{X}$; $k \ll p$

<u>Note</u>: (iii) may not always be possible

Major uses of PCA

(i) Data dimension reduction ~~to visualization~~

(ii) Outlier mining

(iii) Detection of clusters in the multidimensional data cloud

(iv) Projection & visualization

(v) ranking of multidimensional data

(vi) Checking for multivariate normality

Defⁿ: Principal components

PCs are uncorrelated linear combinations $Y_1, \ldots, Y_p$ (derived from $X_1, \ldots, X_p$) whose variances are in decreasing order, with Y_1 explaining the maximum variability, Y_2 explaining the second maximum variability and so on.

i.e. the 1^{st} PC, Y_1, is the linear combination $\underset{\sim}{l}_1' \underset{\sim}{X}$ that maximizes $V(\underline{l}'\underline{X})$ subject to $\underline{l}'\underline{l} = 1$

the 2^{nd} PC, Y_2, is the linear combination $\underline{l}_2'\underline{X}$ that maximizes $V(\underline{l}'\underline{X})$ subject to $\underline{l}'\underline{l} = 1$ and $\text{cov}(\underline{l}_1'\underline{X}, \underline{l}'\underline{X}) = 0$

the i^{th} PC, Y_i, is the linear combination $\underline{l}_i'\underline{X}$ that maximizes $V(\underline{l}'\underline{X})$ subject to $\underline{l}'\underline{l} = 1$ and $\text{cov}(\underline{l}_k'\underline{X}, \underline{l}'\underline{X}) = 0 \;\forall\, k < i$.

Derivation of PCs

Let Σ be the covariance matrix associated with the random vector $\underset{\sim}{X}_{p\times 1}$. The eigenvalue-eigenvector (o.n). pairs of Σ are $(\lambda_1, \underset{\sim}{e}_1), \ldots, (\lambda_p, \underset{\sim}{e}_p)$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$.

The i^{th} PC is given by

$$Y_i = \underset{\sim}{e}_i' \underset{\sim}{X} \qquad i = 1(1)p$$

with $V(Y_i) = \lambda_i$ & $\text{Cov}(Y_i, Y_j) = 0 \;\forall\, i \neq j$

Pf: Consider any $\underline{l}'\underset{\sim}{X}$, $\underset{\sim}{l} \in \mathbb{R}^p$; $\ni \underline{l}'\underline{l} = 1$

$$V(\underline{l}'\underline{X}) = \underline{l}' V(\underset{\sim}{X})\,\underline{l}$$
$$= \underline{l}' \Sigma \underline{l}$$

Consider the eigen value - eigen vector decomposition of Σ as

$$\Sigma = P D_\lambda P' \quad ; \quad P = (\underset{\sim}{e}_1 : \cdots : \underset{\sim}{e}_p)$$
$$D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$$

$$\Rightarrow V(\underline{l}'\underline{X}) = \underline{l}' P D_\lambda P' \underset{\sim}{l}$$
$$= \underline{\beta}' D_\lambda \underset{\sim}{\beta}$$
$$= \sum_{i=1}^{p} \beta_i^2 \lambda_i$$

where $\underline{\beta} = P'\underset{\sim}{l} \in \mathbb{R}^p$

$$\underline{l} = P\underset{\sim}{\beta}$$

Now, $\underline{l}'\underline{l} = 1 \Rightarrow \underline{\beta}' P' P \underline{\beta} = \underline{\beta}'\underline{\beta} = 1$

Thus

$$\max_{\underset{\sim}{l} \ni \underline{l}'\underline{l} = 1} V(\underline{l}'\underline{X}) = \max_{\underset{\sim}{\beta} \ni \underline{\beta}'\underline{\beta} = 1} \underline{\beta}' D_\lambda \underline{\beta} = \max_{\underline{\beta} \ni \underline{\beta}'\underline{\beta} = 1} \sum_{i=1}^{p} \lambda_i \beta_i^2$$

Realize that,

$$\sum_{i=1}^{p} \lambda_i \beta_i^2 \leq \lambda_1 \sum \beta_i^2 = \lambda_1$$
$$\text{i.e. } V(\underline{l}'\underline{X}) \leq \lambda_1$$

and

$$V(\underset{\sim}{e}_1'\underline{X}) = \underset{\sim}{e}_1 \Sigma \underset{\sim}{e}_1$$
$$= \underset{\sim}{e}_1' P D_\lambda P' \underset{\sim}{e}_1$$
$$= \underset{\sim}{e}_1' (\underset{\sim}{e}_1 : \cdots : \underset{\sim}{e}_p) D_\lambda \begin{pmatrix} \underset{\sim}{e}_1' \\ \vdots \\ \underset{\sim}{e}_p' \end{pmatrix} \underset{\sim}{e}_1$$

i.e. $V(\underset{\sim}{e}_1' \underline{X}) = (1, 0, \dots, 0)\, D_\lambda \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$$= \lambda_1 = \max_{\underline{l} \ni \underline{l}'\underline{l} = 1} V(\underline{l}'\underline{X})$$

$\Rightarrow Y_1 = \underset{\sim}{e}_1' \underline{X}$ is the 1^{st} PC

For the second PC, we need $Y_2 = \underline{l}_2' \underset{\sim}{X} \ni Y_2$ is uncorrelated with Y_1 and $V(\underline{l}_2' \underline{X})$ to be the maximum under this constraint.

Now, $\mathrm{Cov}(\underline{l}'\underset{\sim}{X}, \underset{\sim}{e}_1'\underline{X})$

$$= E\left(\underset{\sim}{l}'(\underline{X} - \underline{\mu})\right)\left(\underset{\sim}{e}_1'(\underline{X} - \underline{\mu})\right)'$$

$$= \underline{l}' \, \Sigma \, \underset{\sim}{e}_1$$

$$= \underset{\sim}{l}' \left(\sum_{i=1}^{p} \lambda_i \underset{\sim}{e}_i \underset{\sim}{e}_i' \right) \underset{\sim}{e}_1$$

$$= \underset{\sim}{l}' (\lambda_1 \underset{\sim}{e}_1) = \lambda_1 \underline{l}' \underset{\sim}{e}_1$$

Thus $\mathrm{Cov}(\underline{l}'\underset{\sim}{X}, \underset{\sim}{e}_1'\underline{X}) = 0 \Leftrightarrow \underset{\sim}{l} \perp \underset{\sim}{e}_1$

Thus $\max_{\underline{l} \ni \underline{l}'\underline{l} = 1} V(\underline{l}'\underline{X})$ subject to $\mathrm{Cov}(\underline{l}'\underset{\sim}{X}, Y_1) = 0$ is

equivalent to $\max_{\substack{\underline{l} \ni \underline{l} \perp \underset{\sim}{e}_1 \\ \underline{l}'\underline{l} = 1}} V(\underline{l}'\underline{X})$

$$V(\underline{l}'\underline{X}) = \underline{l}' P D_\lambda P'$$

$$= \underline{l}' (\underset{\sim}{e}_1 : \dots : \underset{\sim}{e}_p) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \begin{pmatrix} \underset{\sim}{e}_1' \\ \vdots \\ \underset{\sim}{e}_p' \end{pmatrix}$$

$\forall\ \underset{\sim}{l} \perp \underset{\sim}{e}_1$

$$\underline{l}' P D_\lambda P' \underline{l} = (\underset{\sim}{0}, b_1, \ldots, b_p) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \begin{pmatrix} 0 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}$$

$$= \sum_{i=2}^{p} b_i^2 \lambda_i$$

$$\le \lambda_2$$

$$\left(\begin{array}{l} \underset{\sim}{b} = (0, b_2, \ldots, b_p)' = P' \underset{\sim}{l} \\ \text{i.e. } \underline{l} = P \underset{\sim}{b} \\ \underline{l}' \underline{l} = 1 \Rightarrow \underline{b}' \underline{b} = 1 \text{ i.e. } \sum_2^p b_i^2 = 1 \end{array} \right)$$

And $V(\underline{e}_2' \underline{X}) = \underline{e}_2' \Sigma \underset{\sim}{e}_2$

$$= \underline{e}_2' (\underset{\sim}{e}_1 : \cdots : \underline{e}_p) D_\lambda \begin{pmatrix} \underline{e}_1' \\ \vdots \\ \underline{e}_p' \end{pmatrix} \underline{e}_2$$

$$= (0, 1, 0, \ldots, 0) D_\lambda \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \lambda_2$$

$$= \max_{\substack{\underline{l} \ni \underline{l} \ni \underset{\sim}{e}_1 \\ \underline{l}'\underline{l} = 1}} V(\underline{l}' \underline{X})$$

$\Rightarrow Y_2 = \underline{e}_2' \underline{X}$ is the 2^{nd} PC

For the $(K+1)^{th}$ PC $(K+1 \le p)$, we need to

maximize $V(\underline{l}'\underline{X}) \ni Cov(\underline{l}'\underset{\sim}{X}, Y_i) = 0\ \forall\ i \le K$

i.e. maximize $V(\underline{l}'\underline{X}) \ni \underline{l} \perp \underset{\sim}{e}_i\ \forall\ i \le K$

$\forall\ \underset{\sim}{l} \perp \underset{\sim}{e}_1, \ldots, \underset{\sim}{e}_k$

$$V(\underline{l}'\underline{X}) = \underset{\sim}{l}'(\underset{\sim}{e}_1 \vdots \cdots \vdots \underset{\sim}{e}_k \vdots \underset{\sim}{e}_{k+1} \vdots \cdots \vdots \underset{\sim}{e}_p)\, D_\lambda \begin{pmatrix} \underset{\sim}{e}_1' \\ \vdots \\ \underset{\sim}{e}_k' \\ \underset{\sim}{e}_{k+1}' \\ \vdots \\ \underset{\sim}{e}_p' \end{pmatrix} \underset{\sim}{l}$$

$$= (0, \ldots 0, \underbrace{c_{k+1}}_{\underline{l}'\underset{\sim}{e}_{k+1}}, \ldots, c_p)\, D_\lambda \begin{pmatrix} 0 \\ \vdots \\ 0 \\ c_{k+1} \\ \vdots \\ c_p \end{pmatrix}$$

$$= \sum_{i=k+1}^{p} c_i^2 \lambda_i$$

$\underset{\sim}{c} = P'\underset{\sim}{l}$ with $\underset{\sim}{c} = (\underbrace{0, \ldots 0}_{k}, c_{k+1}, \ldots, c_p)'$

$\underline{l}'\underline{l} = 1 \Rightarrow \underline{c}'\underline{c} = 1$ i.e. $\sum_{k+1}^{p} c_i^2 = 1$

Thus $\forall\ \underline{l} \ni \underset{\sim}{e}_1, \ldots, \underset{\sim}{e}_k$

$$V(\underline{l}'\underline{X}) = \sum_{k+1}^{p} c_i^2 \lambda_i \le \lambda_{k+1}$$

and $\lambda_{k+1} = V(\underline{e}_{k+1}'\underline{X}) \left(= \underline{e}_{k+1}' P D_\lambda P' \underset{\sim}{e}_{k+1}\right)$

$$\Rightarrow V(\underset{\sim}{e}_{k+1}'\underline{X}) = \lambda_{k+1} = \max_{\substack{\underline{l} \ni \underline{l} \perp \underset{\sim}{e}_1 \ldots \underset{\sim}{e}_k \\ \& \ \underline{l}'\underline{l} = 1}} V(\underline{l}'\underline{X})$$

$\Rightarrow$ ~~th~~ $(k+1)^{\text{th}}$ PC is

$$Y_{k+1} = \underline{e}_{k+1}'\underline{X}$$

Note 1: $Y_1, \ldots, Y_p$ are $\ni$

$$\text{Cov}(\underset{\sim}{Y}) = D_\lambda$$

$$Y_i = \underset{\sim}{e}_i' \underset{\sim}{X}$$

$$\begin{aligned}
\text{Cov}(Y_i, Y_j) &= \text{Cov}(\underset{\sim}{e}_i' \underset{\sim}{X}, \underset{\sim}{e}_j' \underset{\sim}{X}) \\
&= \underset{\sim}{e}_i' \Sigma \underset{\sim}{e}_j \\
&= \underset{\sim}{e}_i' \left(\sum_{k=1}^{p} \lambda_k \underset{\sim}{e}_k \underset{\sim}{e}_k' \right) \underset{\sim}{e}_j \\
&= \begin{cases} \lambda_i, & \text{If } i = j \\ 0, & \text{If } i \neq j \end{cases}
\end{aligned}$$

$$\Rightarrow \text{Cov}(\underset{\sim}{Y}) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} = \Sigma_Y$$

Note 2: Total variation in $\underset{\sim}{X}$ = Total variation in $\underset{\sim}{Y}$

$$\begin{aligned}
\text{Total variation in } \underset{\sim}{X} &= \text{tr}\, \Sigma \\
&= \text{tr}(P D_\lambda P') \\
&= \text{tr}(D_\lambda P' P) \\
&= \text{tr}\, D_\lambda = \sum_{i=1}^{p} \lambda_i
\end{aligned}$$

$$\begin{aligned}
\text{Total variation in } \underset{\sim}{Y} &= \text{tr}(\text{Cov}(\underset{\sim}{Y})) \\
&= \text{tr}\, D_\lambda = \sum_{i=1}^{p} \lambda_i
\end{aligned}$$

Note 3 : If $\underset{\sim}{X} \sim N_p(\underline{\mu}, \Sigma)$, then

(i) $Y_i = \underline{e}_i' \underset{\sim}{X} \quad (\forall\, i = 1(1)p) \sim N_1$

(ii) $$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} = \begin{pmatrix} \underline{e}_1' \underline{X} \\ \vdots \\ \underline{e}_p' \underline{X} \end{pmatrix} = \begin{pmatrix} \underline{e}_1' \\ \vdots \\ \underline{e}_p' \end{pmatrix} \underset{\sim}{X}$$

i.e. $\underset{\sim}{Y} = P' \underset{\sim}{X} \sim N_p(P'\underset{\sim}{\mu}, P'\Sigma P)$

$$P'\Sigma P = P'(P D_\lambda P')P = D_\lambda$$

(iii) $Y_1, \ldots, Y_p$ are independent N_1

(cov $= 0$ for joint normal $\Rightarrow$ independence)

Note 4: $\text{Cov}(X_i, Y_j) = ?$

$Y_j = \underline{e}_j' \underset{\sim}{X}$

$\underset{\sim}{Y} = P' \underset{\sim}{X}$

$$\underset{\sim}{X} = P \underset{\sim}{Y} = (\underset{\sim}{e}_1 : \ldots : \underset{\sim}{e}_p)\, \underset{\sim}{Y}$$

$$= \begin{pmatrix} e_{11} & \ldots & e_{1p} \\ e_{21} & \ldots & e_{2p} \\ \vdots & & \vdots \\ e_{p1} & \ldots & e_{pp} \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix}$$

i.e. $$X_i = \sum_{k=1}^{p} e_{ik} Y_k$$

$$\text{Cov}(X_i, Y_j) = \text{Cov}\left(\sum_{k=1}^{p} e_{ik} Y_k, Y_j\right) = e_{ij}\lambda_j$$

e_{ij} is the i^{th} element (row) of $\underset{\sim}{e}_j$

$$\text{Corr}^n(X_i, Y_j) = e_{ij}\sqrt{\frac{\lambda_j}{\sigma_{ii}}}$$

$$V(X_i) = \sigma_{ii} \quad \& \quad V(Y_j) = \lambda_j$$

$$\Rightarrow \text{Corr}^n(X_i, Y_j) = \frac{e_{ij}\lambda_j}{(\sigma_{ii}\lambda_j)^{1/2}} = e_{ij}\sqrt{\frac{\lambda_j}{\sigma_{ii}}}$$

<u>Note 5</u>: Principal Components derived from correlation matrix

Consider the standardized variables

$$X_i \rightarrow Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \quad i = 1(1)p; \quad \sigma_{ii} > 0$$

Define

$$V = \text{diag}(\sigma_{11}, \ldots, \sigma_{pp})$$

$$\underset{\sim}{Z} = (Z_1, \ldots, Z_p)' = V^{-1/2}(\underline{X} - \underline{\mu})$$

$$E\,\underset{\sim}{Z} = \underset{\sim}{0}\ ; \quad \text{cov}(\underset{\sim}{Z}) = V^{-1/2}\,\Sigma\, V^{-1/2} = \rho$$

Let $(\eta_i, \underset{\sim}{f}_i)$ $(i = 1(1)p)$, $\eta_1 \geq \eta_2 \cdots \geq \eta_p \geq 0$ be the eigen value - eigen vector (o.n.) pairs of ρ matrix, then

(a) i^{th} PC derived from ρ is

$$Y_i = \underset{\sim}{f}_i'\,\underset{\sim}{Z} = \underset{\sim}{f}_i'\,V^{-1/2}(\underline{X} - \underline{\mu})$$

(b) total variation in $\underset{\sim}{Y} = \sum_{i=1}^{p} V(Y_i) = \underbrace{\sum_{i=1}^{p} V(Z_i)}_{\text{total variation in } \underset{\sim}{Z}} = p$

(c) $\rho_{Z_i, Y_j} = f_{ij}\sqrt{\eta_j}\ ; \quad i, j = 1(1)p$

Remark: (i) Proportion of total variation in $\underset{\sim}{X}$ explained by the k^{th} PC (based on Σ) is

$$\lambda_k \Big/ \sum_1^p \lambda_i$$

(ii) Proportion of total variation in $\underset{\sim}{X}$ explained by the first k PCs (based on Σ) is

$$\sum_{i=1}^{k} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i$$

(iii) If $\sum_{i=1}^{k} \lambda_i \approx \sum_{i=1}^{p} \lambda_i$ for some small k, then it is possible to have a meaningful data dimension reduction as it is enough in such a situation to consider only $(Y_1, \ldots, Y_k)$. Ideally $k \ll p$!

Remark: $e_{ij} \propto \rho_{X_i, Y_j}$ and hence the magnitude of e_{ij} indicates how important is X_i to Y_j, this helps in interpreting Y_j

Remark: In case the units of variables are different or the variables have widely varying ranges, we should obtain PCs from standardized variables i.e. work with ρ instead of Σ to get PCs.

Examples

(1) $\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ & 5 & 0 \\ & & 2 \end{pmatrix}$.

Soln of $|\Sigma - \lambda I| = 0$ gives $\lambda_1 = 5.83$, $\lambda_2 = 2.0$, $\lambda_3 = 0.17$

$$\underset{\sim}{e}_1' = (0.383, -.924, 0)$$

$$\underline{e}_2' = (0, 0, 1)$$

$$\underline{e}_3' = (.924, 0.383, 0)$$

PCs

$$Y_1 = .383\, X_1 - .924\, X_2$$

$$Y_2 = X_3$$

$$Y_3 = .924\, X_1 + .383\, X_2$$

1st PC explains $\frac{5.83}{8} \times 100 = 73\%$ of total variation

1st 2 PCs explain $\frac{5.83+2.0}{8} \times 100 = 98\%$. of total variation

$$\rho_{X_1, Y_1} = .383\sqrt{\frac{5.83}{1}} \approx .925 \qquad \left(\rho_{X_i, X_j} = e_{ij}\sqrt{\frac{\lambda_j}{\sigma_{ii}}}\right).$$

$$\rho_{X_2, Y_1} = (-.924)\sqrt{\frac{5.83}{5}} \approx -.998$$

- - .

Example 2

$$\Sigma = \begin{pmatrix} 1 & 4 \\ 4 & 100 \end{pmatrix} \quad X_1, X_2 \rightarrow \text{widely varying in spread}$$

$\lambda_1 = 100.16$, $\lambda_2 = 0.84$

$\underset{\sim}{e}_1' = (.04, .999)$; $\underset{\sim}{e}_2' = (.999, -0.04)$

$Y_1 = .04 X_1 + .999 X_2$; $Y_2 = .999 X_1 - .04 X_2$

Y_1 explains $\frac{100.16}{101} \times 100 \approx 99\%$ of total variation

$$\left.\begin{aligned} \rho_{X_1, Y_1} &= 0.04 \sqrt{\frac{100.16}{1}} \approx 0.4 \\ \rho_{X_2, Y_1} &= .999 \sqrt{\frac{100.16}{100}} \approx .99 \end{aligned}\right\} \begin{array}{l} X_2 \text{ is much more imp} \\ \text{to } Y_1 \text{ than } X_1 \text{ is to } Y_1 \end{array}$$

$Var(X_2)$ dominating!

Consider standardized variables

$$Z_1 = \frac{X_1 - \mu_1}{\sqrt{\sigma_{11}}} \qquad Z_2 = \frac{X_2 - \mu_2}{\sqrt{\sigma_{22}}}$$

$$\Sigma_Z = \rho_X = \begin{pmatrix} 1 & .4 \\ .4 & 1 \end{pmatrix}$$

$\eta_1 = 1.4$; $\eta_2 = 0.6$

$\underset{\sim}{f}_1' = (.707, .707)$; $\underline{f}_2' = (.707, -707)$

$Y_1 = .707 \cancel{Z}_1 + .707 Z_2$

$Y_2 = .707 Z_1 - .707 Z_2$

$$\rho_{Z_1, Y_1} = .707 \sqrt{\frac{1.4}{1}} = \rho_{Z_2, Y_1}$$

Z_1 & Z_2 are equally important

Y_1 explains $\frac{1.4}{2} \times 100 = 70\%$ of total variation

Principal components from data matrix

Observed data matrix

$$\underset{p\times n}{\mathcal{X}} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pn} \end{pmatrix}$$

$\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ n obsns from multivariate pop^n with mean vector μ and cov mat Σ

Sample mean vector : $\bar{\underset{\sim}{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix} = \frac{1}{n} \mathcal{X} \underline{1}_n$

Sample variance covariance matrix :

$$S_{n-1} = \frac{1}{n-1} \mathcal{X} \left(I_n - \frac{1}{n} \underset{\sim}{1}_n \underset{\sim}{1}_n' \right) \mathcal{X}' = \frac{1}{n-1} \Sigma (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'$$

$$= \frac{1}{n-1} \begin{pmatrix} \Sigma(x_{1j} - \bar{x}_1) & \Sigma(x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) & \cdots & \Sigma(x_{1j} - \bar{x}_1)(x_{pj} - \bar{x}_p) \\ & \ddots & & \\ & & & \Sigma (x_{pj} - \bar{x}_p)^2 \end{pmatrix}$$

From the observed data matrix, our objective is to construct uncorrelated linear combinations of the measured characteristics that account for as much of the sample variation as possible.

Suppose $\underset{\sim}{l}' \underset{\sim}{x}$ is sample linear combination

We have $\underset{\sim}{l}' \underset{\sim}{x}_1, \ldots \underset{\sim}{l}' \underset{\sim}{x}_n$: n transformed obsns

Sample mean of $(\underline{l}' \underset{\sim}{x}_1, \ldots, \underline{l}' \underset{\sim}{x}_n)$:

$$\frac{1}{n} \sum_1^n \underline{l}' \underset{\sim}{x}_j$$

$$= \frac{1}{n} \sum_{j=1}^n \left(l_1 x_{1j} + l_2 x_{2j} + \cdots + l_p x_{pj} \right)$$

$$= l_1 \bar{x}_1 + l_2 \bar{x}_2 + \cdots + l_p \bar{x}_p = \underline{l}' \bar{\underset{\sim}{x}}$$

Sample variance of $(\underset{\sim}{l}'\underset{\sim}{x}_1, \ldots, \underset{\sim}{l}'\underset{\sim}{x}_n)$:

$$\frac{1}{n-1}\left((\underset{\sim}{l}'\underset{\sim}{x}_1 - \underset{\sim}{l}'\bar{\underset{\sim}{x}})^2 + \cdots + (\underset{\sim}{l}'\underset{\sim}{x}_n - \underset{\sim}{l}'\bar{\underset{\sim}{x}})^2 \right)$$

$$= \frac{1}{n-1}\left(\left(\underset{\sim}{l}'(\underset{\sim}{x}_1 - \bar{\underset{\sim}{x}})\right)^2 + \cdots + \left(\underset{\sim}{l}'(\underset{\sim}{x}_n - \bar{\underset{\sim}{x}})\right)^2 \right)$$

$$= \frac{1}{n-1}\left(\left(\underset{\sim}{l}'(\underset{\sim}{x}_1 - \bar{\underset{\sim}{x}})\right)\left(\underset{\sim}{l}'(\underset{\sim}{x}_1 - \bar{\underset{\sim}{x}})\right)' + \cdots + \left(\underset{\sim}{l}'(\underset{\sim}{x}_n - \bar{\underset{\sim}{x}})\right)\left(\underset{\sim}{l}'(\underset{\sim}{x}_n - \bar{\underset{\sim}{x}})\right)' \right)$$

$$= \frac{1}{n-1}\left(\underset{\sim}{l}'(\underset{\sim}{x}_1 - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_1 - \bar{\underset{\sim}{x}})'\underset{\sim}{l} + \cdots + \underset{\sim}{l}'(\underset{\sim}{x}_n - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_n - \bar{\underset{\sim}{x}})'\underset{\sim}{l} \right)$$

$$= \underset{\sim}{l}'\left(\frac{1}{n-1}\sum_{j=1}^{n} (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})' \right)\underset{\sim}{l}$$

$$= \underset{\sim}{l}' S_{n-1} \underset{\sim}{l}$$

Further, suppose $\underset{\sim}{l}_i'\underset{\sim}{x}$ & $\underset{\sim}{l}_j'\underset{\sim}{x}$ be 2 linear combinations,

Sample covariance bet^n n pairs

$$(\underset{\sim}{l}_i'\underset{\sim}{x}_1, \underset{\sim}{l}_j'\underset{\sim}{x}_1), \ldots \ (\underset{\sim}{l}_i'\underset{\sim}{x}_n, \underset{\sim}{l}_j'\underset{\sim}{x}_n)$$

$$\frac{1}{n-1}\left((\underset{\sim}{l}_i'\underset{\sim}{x}_1 - \underset{\sim}{l}_i'\bar{\underset{\sim}{x}})(\underset{\sim}{l}_j'\underset{\sim}{x}_1 - \underset{\sim}{l}_j'\bar{\underset{\sim}{x}}) + \cdots + (\underset{\sim}{l}_i'\underset{\sim}{x}_n - \underset{\sim}{l}_i'\bar{\underset{\sim}{x}})(\underset{\sim}{l}_j'\underset{\sim}{x}_n - \underset{\sim}{l}_j'\bar{\underset{\sim}{x}}) \right)$$

$$= \frac{1}{n-1}\sum_{l=1}^{n} \left(\underset{\sim}{l}_i'(\underset{\sim}{x}_l - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_l - \bar{\underset{\sim}{x}})'\underset{\sim}{l}_j' \right)$$

$$= \underset{\sim}{l}_i'\left(\frac{1}{n-1}\sum_{l=1}^{n} (\underset{\sim}{x}_l - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_l - \bar{\underset{\sim}{x}})' \right)\underset{\sim}{l}_j$$

$$= \underset{\sim}{l}_i' S_{n-1} \underset{\sim}{l}_j$$

Def^n: Sample Principal Components

The sample PCs are uncorrelated linear combinations of observation vectors, have sample variances in decreasing order. The 1^{st} sample PC is the lin comb $\underset{\sim}{l}_1' \underset{\sim}{x}_j$ which maximises the sample variance of $\underset{\sim}{l}'\underset{\sim}{x}_j$ subject to $\underset{\sim}{l}'\underset{\sim}{l} = 1$ (i.e. sample variance of $(\underset{\sim}{l}_1'\underset{\sim}{x}_1, \ldots \underset{\sim}{l}_1'\underset{\sim}{x}_n)$

$$= \max_{\underset{\sim}{l} \ni \underset{\sim}{l}'\underset{\sim}{l} = 1} \underset{\sim}{l}' S_{n-1} \underset{\sim}{l} \Big).$$

The 2^{nd} sample PC is the linear comb $\underline{l}_2'\underset{\sim}{x}_j \ni$ it maximizes sample variance of $\underset{\sim}{l}'\underline{x}_j \ni \underline{l}'\underline{l} = 1$ and zero sample covariance for the pairs $(\underset{\sim}{l}_1'\underset{\sim}{x}_j, \underset{\sim}{l}'\underline{x}_j)$; $j = 1(1)n$

The i^{th} sample PC is the linear comb $\underline{l}_i'\underline{x}_j \ni$ it maximizes the sample variance of $\underline{l}'\underline{x}_j$ and having zero sample covariance for all pairs $(\underline{l}'\underline{x}_j, \underline{l}_k'\underline{x}_j)$ $\forall k < i$; $j = 1(1)n$.

S_{n-1} : Sample variance covariance matrix

$(\hat{\lambda}_1, \hat{\underset{\sim}{e}}_1), \ldots, (\hat{\lambda}_p, \hat{\underline{e}}_p)$: eigenvalue - eigenvector(o.n.) pairs of S_{n-1}

$$\max_{\underset{\sim}{l} \ni \underset{\sim}{l}'\underset{\sim}{l} = 1} \underset{\sim}{l}' S_{n-1} \underset{\sim}{l} = \hat{\lambda}_1 = \text{sample variance of } (\hat{\underset{\sim}{e}}_1'\underset{\sim}{x}_1, \ldots, \hat{\underset{\sim}{e}}_1'\underset{\sim}{x}_n)$$

$$\Big(= \hat{\underset{\sim}{e}}_1' S_{n-1} \hat{\underset{\sim}{e}}_1$$

$$= \hat{\underset{\sim}{e}}_1' \Big(\sum_1^n \hat{\lambda}_j \hat{\underline{e}}_j \hat{\underline{e}}_j'\Big) \hat{\underset{\sim}{e}}_1 = \hat{\lambda}_1 \Big)$$

1^{st} sample PC is $\hat{\underset{\sim}{e}}_1 \underset{\sim}{x}_j = \hat{y}_{1(j)}$

For 2^{nd} sample PC we need to find

$$\max_{\substack{\underset{\sim}{l} \ni \underset{\sim}{l} \perp \hat{\underset{\sim}{e}}_1 \\ \underset{\sim}{l}'\underset{\sim}{l}=1}} \underset{\sim}{l}' S_{n-1} \underset{\sim}{l} = \hat{\lambda}_2 = \text{Sample variance of}$$

$$(\hat{\underset{\sim}{e}}_2' \underset{\sim}{x}_1, \ldots, \hat{\underset{\sim}{e}}_2' \underset{\sim}{x}_n)$$

2^{nd} sample PC is $\hat{\underset{\sim}{e}}_2' \underset{\sim}{x}_j = \hat{y}_{2(j)}$

k^{th} $(k \le p)$ sample PC is $\hat{\underset{\sim}{e}}_k' \underset{\sim}{x}_j = \hat{y}_{k(j)}$

<u>Note</u> (i) Sample variance of $\hat{y}_k = \hat{\lambda}_k$

(ii) Sample covariance of $\hat{y}_i$ & $\hat{y}_k = 0 \quad \forall\, i \neq k$

(Sample covariance of $(\hat{\underset{\sim}{e}}_i' \underset{\sim}{x}_j, \hat{\underset{\sim}{e}}_k' \underset{\sim}{x}_j)$ pairs

$$= \hat{\underset{\sim}{e}}_i' S_{n-1} \hat{\underset{\sim}{e}}_k$$

$$= \hat{\underset{\sim}{e}}_i' \left(\sum_{j=1}^{n} \hat{\lambda}_j \hat{\underset{\sim}{e}}_j \hat{\underset{\sim}{e}}_j' \right) \hat{\underset{\sim}{e}}_k = 0$$

(iii) total sample variation $= \sum^{p} s_{ii}$

$=$ total sample variation in sample PCs

$$= \sum_{i=1}^{p} \hat{\lambda}_i$$

(iv) $$r_{x_i, \hat{y}_k} = \hat{e}_{ik} \sqrt{\frac{\hat{\lambda}_k}{s_{kk}}}$$

. . . .

Uses of PCA

Sample PCs $\hat{y}_i = \hat{\underset{\sim}{e}}_i' \underset{\sim}{x}$; $i = 1(1)p$

$$\underset{\sim}{x}_1 \longrightarrow \left(\hat{y}_1^{(1)}, \hat{y}_2^{(1)}, \ldots, \hat{y}_p^{(1)}\right)$$

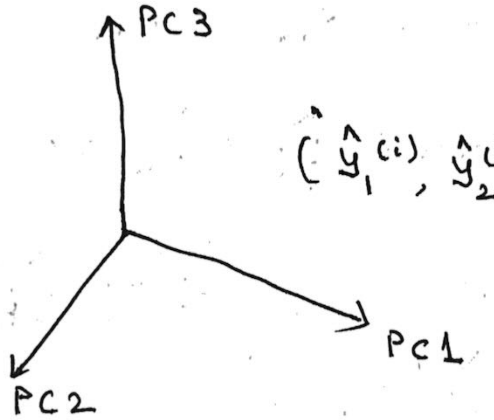
$$\underset{\sim}{x}_2 \longrightarrow \left(\hat{y}_1^{(2)}, \hat{y}_2^{(2)}, \ldots, \hat{y}_p^{(2)}\right)$$

— — — — —

$$\underset{\sim}{x}_n \longrightarrow \left(\hat{y}_1^{(n)}, \hat{y}_2^{(n)}, \ldots, \hat{y}_p^{(n)}\right)$$

(i) Data projection & visualization

Projection onto PC plane



$\left(\hat{y}_1^{(i)}, \hat{y}_2^{(i)}, \hat{y}_3^{(i)}\right)$ $i = 1(1)n$

- Meaningful when $(\hat{y}_1, \hat{y}_2, \hat{y}_3)$ collectively explains a "significant" amount of total sample variation
- a similar 2d plot on (PC1, PC2) plane, provided the first 2PC collectively captures "significant" proportion of total sample variation

(ii) Outlier detection : from 2d/3d plot (PC plane)

(iii) Cluster detection : from above plots

(iv) Data dimension reduction

through $(\hat{y}_1, \ldots, \hat{y}_K)$ for $K < p$

(v) Ranking of cases

$$\left.\begin{array}{l} \underset{\sim}{x}_1 \to \hat{y}_1^{(1)} \\ \underset{\sim}{x}_2 \to \hat{y}_1^{(2)} \\ \vdots \\ \underset{\sim}{x}_n \to \hat{y}_1^{(n)} \end{array}\right\}$$

rank based on $\hat{y}_1^{(i)}$

Coeffts of variables (original) play a crucial role

(vi) variable clustering

- $$r_{\hat{y}_K, x_i} = \hat{e}_{iK} \sqrt{\frac{\hat{\lambda}_K}{s_{ii}}}$$
- Represent each of the original p variables by $(\hat{r}_{\hat{y}_1, x_i}, \hat{r}_{\hat{y}_2, x_i}, \hat{r}_{\hat{y}_3, x_i})$ $i = 1(1)p$
- Identify variable clusters from this plot

(vii) Checking for multivariate normality

- Check for univariate normality of each of the p PCs
- For ith PC use $(\hat{y}_i^{(1)}, \ldots, \hat{y}_i^{(n)})$ to check for univariate normality
- If any of the p checks for univariate normality fails, N_p assumption will be rejected