Mahalanobis transformation

Univariate $y = \frac{x - E[x]}{\sqrt{Var[x]}} = \frac{x}{\sqrt{Var[x]}} - \frac{E[x]}{\sqrt{Var[x]}} \Rightarrow E[y] = 0,\ Var[y] = 1$

$\underline{x}$ multivariate. $E[\underline{x}] = \underline{\mu}$, $Cov[\underline{x}] = \underline{\Sigma}$

Can we find $\underline{y} = \underline{A}\,\underline{x} + \underline{b}$ such that $E[\underline{y}] = 0$

and $Cov[\underline{y}] = \underline{I}$

Must have: 1. $\underline{A}\,E[\underline{x}] + \underline{b} = \underline{A}\,\underline{\mu} + \underline{b} = 0$

2. $Cov[\underline{y}] = \underline{A}\,\underline{\Sigma}\,\underline{A}^T = \underline{I}$. (assume positive definite)

Start with 2. Define $\underline{\Sigma}^{\frac{1}{2}} = \underline{P}\,\underline{\Lambda}^{\frac{1}{2}}\underline{P}^T$ and $\underline{\Sigma}^{-\frac{1}{2}} = \underline{P}\,\underline{\Lambda}^{-\frac{1}{2}}\underline{P}^T$

where $\underline{\Lambda}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_p} \end{bmatrix}$ and $\underline{\Lambda}^{-\frac{1}{2}} = \begin{bmatrix} 1/\sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\lambda_p} \end{bmatrix}$

Then $\underline{\Sigma}^{\frac{1}{2}}\underline{\Sigma}^{-\frac{1}{2}} = \underline{P}\,\underline{\Lambda}^{\frac{1}{2}}\underbrace{\underline{P}^T\underline{P}}_{I}\,\underline{\Lambda}^{-\frac{1}{2}}\underline{P}^T = \underline{P}\,\underline{P}^T = \underline{I}$

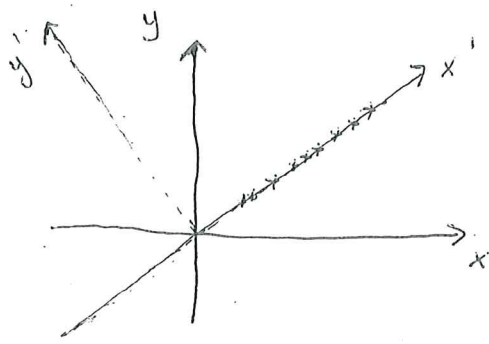
With $\underline{A} = \underline{\Sigma}^{-\frac{1}{2}}$ we get $Cov[\underline{y}] = \underline{A}\,\underline{\Sigma}\,\underline{A}^T$

$= \underline{P}\,\underline{\Lambda}^{-\frac{1}{2}}\underbrace{\underline{P}^T\underline{P}}_{I}\,\underline{\Lambda}\,\underbrace{\underline{P}^T\underline{P}}_{I}\,\underline{\Lambda}^{-\frac{1}{2}}\underline{P}^T = \underline{P}\,\underline{P}^T = \underline{I}$

and $E[\underline{y}] = \underline{\Sigma}^{-\frac{1}{2}}\underline{\mu} + \underline{b} = 0 \Rightarrow \underline{b} = -\underline{\Sigma}^{-\frac{1}{2}}\underline{\mu}$

Hence the transformation is: $\underline{y} = \underline{\Sigma}^{-\frac{1}{2}}(\underline{x} - \underline{\mu})$

Principal components



Principal component analysis is a method for re-expressing multivariate data. It allows the researcher to reorient the data such that the first few dimensions account for as much of the variation in the original variables as possible. Most often it is used for dimension reduction and data interpretation.

Let $\underline{x}^T = (x_1, x_2, \ldots, x_p)$ and let $\underline{\Sigma} = \text{cov}(\underline{x})$

A linear combination of $\underline{x}$ is given by $y = \underline{a}^T \underline{x}$

and $\text{Var}[y] = E[\underline{a}^T(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^T \underline{a}] = \underline{a}^T \underline{\Sigma}\, \underline{a}$

Principal components are uncorrelated linear combinations of $\underline{x}$ constructed such that

1. The first one maximizes $\text{Var}[\underline{a}_1^T \underline{x}]$

$$\underline{a}_1^T \underline{a}_1 = 1$$

2. The second one maximizes $\text{Var}[\underline{a}_2^T \underline{x}]$ under the

$$\underline{a}_1^T \underline{a}_2 = 1$$

constraint $\text{cov}(\underline{a}_1^T \underline{x}, \underline{a}_2^T \underline{x}) = \underline{a}_1^T \underline{\Sigma}\, \underline{a}_2 = 0$

3. The third one maximizes $\text{Var}[\underline{a}_3^T \underline{x}]$ under the

$$\underline{a}_3^T \underline{a}_3 = 1$$

constraint $\underline{a}_i^T \underline{\Sigma}\, \underline{a}_3 = 0, \quad i < 3$

and so on.

positive definite

RP1. Let $\underline{X}$ have covariance matrix $\underline{\Sigma}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and eigenvectors $\underline{e}_1, \underline{e}_2, \cdots, \underline{e}_p$. The i-th principal component is given by $Y_i = \underline{e}_i^T \underline{X}$, $i = 1, 2, \cdots, p$.
$\text{Var}[Y_i] = \lambda_i$ and $\text{Cov}[Y_i, Y_j] = 0$, $i \neq j$.

Proof $\underline{Y} = \underline{P}^T \underline{X} \Rightarrow \text{Cov}[\underline{Y}] = \underline{P}^T \underline{\Sigma} \underline{P} = \underbrace{\underline{P}^T \underline{P}}_{\underline{I}} \underline{\Lambda} \underbrace{\underline{P}^T \underline{P}}_{\underline{I}} = \underline{\Lambda}$

$$= \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \lambda_p \end{bmatrix}$$

Further: $\max_{\underline{a}^T \underline{a} = 1} \text{Var}[\underline{a}^T \underline{X}] = \max_{\underline{a}^T \underline{a} = 1} \frac{\text{Var}[\underline{a}^T \underline{X}]}{\underline{a}^T \underline{a} = 1} = \max_{\underline{a}^T \underline{a} = 1} \frac{\underline{a}^T \underline{\Sigma} \underline{a}}{\underline{a}^T \underline{a}}$

$$= \max_{\underline{a}^T \underline{a} = 1} \frac{\underline{a}^T \underline{P} \underline{\Lambda} \underline{P}^T \underline{a}}{\underline{a}^T \underbrace{\underline{P} \underline{P}^T}_{\underline{I}} \underline{a}} \quad \max_{(\underline{a}^T \underline{a} = 1)} \frac{\underline{y}^T \underline{\Lambda} \underline{y}}{\underline{y}^T \underline{y}} = \max_{(\underline{a}^T \underline{a} = 1)} \frac{\sum_{i=1}^{p} \lambda_i y_i^2}{\sum_{i=1}^{p} y_i^2} \leq \lambda_1$$

But $\underline{a} = \underline{e}_1$ gives $\frac{\underline{a}^T \underline{\Sigma} \underline{a}}{\underline{a}^T \underline{a}} = \frac{\underline{e}_1^T \underline{\Sigma} \underline{e}_1}{\underline{e}_1^T \underline{e}_1} = \lambda_1 \Rightarrow \max_{\underline{a}^T \underline{a} = 1} \text{Var}[\underline{a}^T \underline{X}] = \lambda_1$

$\Rightarrow Y_1 = \underline{e}_1^T \underline{X}$.

For $Y_2 = \underline{a}_2^T \underline{X}$ we shall have ($\underline{e}_1^T \underline{\Sigma} \underline{a}_2 = \underline{a}_2^T \underline{\Sigma} \underline{e}_1 = 0 \Rightarrow \underline{a}_2^T \underline{e}_1 = 0$)

and $\underline{y}^T = [0, y_2, \cdots, y_p]$

This implies $\max_{\underline{a}_2^T \underline{a}_2 = 1} \frac{\sum_{i=2}^{p} \lambda_i y_i^2}{\sum_{i=2}^{p} y_i^2} \leq \lambda_2$

But $\underline{a}_2 = \underline{e}_2$ gives $\frac{\underline{a}_2^T \underline{\Sigma} \underline{a}_2}{\underline{a}^T \underline{a}} = \frac{\underline{e}_2^T \underline{\Sigma} \underline{e}_2}{\underline{e}_2^T \underline{e}_2} = \lambda_2 \Rightarrow \max_{\substack{\underline{a}_2^T \underline{a}_2 \\ \underline{a}_2^T \underline{\Sigma} \underline{e}_1 = 0}} \text{Var}[\underline{a}_2^T \underline{X}] = \lambda_2$

and $Y_2 = \underline{e}_2^T \underline{X}$

and so on.

RP2 Let $\underline{X}$ have covariance matrix $\mathrm{Cov}[\underline{X}] = \underline{\Sigma}$ (pos def) and let $Y_i = \underline{e}_i^T \underline{X}$, $i = 1, 2, \ldots, p$ be the principal components. Then $\mathrm{tr}(\underline{\Sigma}) = \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$

Proof. $\mathrm{tr}(\underline{\Sigma}) = \mathrm{tr}(\underline{P}\,\underline{\Lambda}\,\underline{P}^T) = \mathrm{tr}(\underline{P}^T\underline{P}\,\underline{\Lambda}) = \mathrm{tr}(\underline{\Lambda})$