Random vectors

$\underline{x} \in \mathbb{R}^p$

Distribution of $\underline{x}$: $\mathbb{P}[\underline{x} \in B] = \nu_{\underline{x}}(B) \quad \forall B \in \mathcal{B}(\mathbb{R}^p)$ Borel in $\mathbb{R}^p$

$\hookrightarrow$ distribution of $\underline{x}$ (measure on $\mathcal{B}$)

Special case : $\nu_{\underline{x}}(B) = \int_B f(\underline{x})\,d\underline{x} \qquad \forall B \in \mathcal{B}$

$\hookrightarrow$ density of $\underline{x}$

MEAN $\qquad \mathbb{E}(\underline{x}) = \begin{bmatrix} \mathbb{E}[x_1] \\ \vdots \\ \mathbb{E}[x_p] \end{bmatrix} := \boxed{\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}}$

COVARIANCE $\qquad \boxed{\Sigma = [\sigma_{ij}] = \operatorname{cov}(\underline{x})} \qquad i,j = 1,\dots,p$

$$\sigma_{ij} = \operatorname{Cov}(x_i, x_j) = \begin{cases} \sigma_{ii} = \operatorname{Var}(x_i) & i=j \\ \sigma_{ij} = \operatorname{Cov}(x_i,x_j) & i \neq j \end{cases}$$

$= \mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]$

CORRELATION $\qquad \Sigma = \mathbb{E}[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})']$

$$V = \operatorname{diag}(\Sigma) = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \ddots & \vdots \\ 0 & \dots & \sigma_{pp} \end{bmatrix} \qquad p \times p$$

$\boxed{V^{-1/2} \Sigma V^{-1/2} = \rho = [\rho_{ij}]}$ correlation matrix

Remark

1. x_1, x_2 r.v.s. in $\mathbb{R}$ $\quad c_1, c_2 \in \mathbb{R}$

$\mathbb{E}[c_1 x_1 + c_2 x_2] = c_1 \mathbb{E}[x_1] + c_2 \mathbb{E}[x_2] = c_1\mu_1 + c_2\mu_2$

$\operatorname{Var}[c_1 x_1 + c_2 x_2] = c_1^2 \operatorname{Var}(x_1) + c_2^2 \operatorname{Var}(x_2) + 2c_1c_2 \operatorname{Cov}(x_1, x_2)$

$\underline{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

$\mathbb{E}[\underline{c}'\underline{x}] = \underline{c}\underline{\mu} \qquad \operatorname{Var}[\underline{c}'\underline{x}] = \underline{c}'\Sigma\underline{c}$

2. if $\underline{x} \in \mathbb{R}^p \quad \underline{c} \in \mathbb{R}^p$

$\mathbb{E}[\underline{x}] = \underline{\mu} \qquad \operatorname{Cov}(\underline{x}) = \Sigma$

$\mathbb{E}[\underline{c}'\underline{x}] = \mathbb{E}[c_1 x_1 + \dots + c_p x_p] = c'\underline{\mu}$

$\operatorname{Cov}[\underline{c}'\underline{x}] = \underline{c}'\Sigma\underline{c}$

3. if c $k \times p$ matrix

$c \cdot \underline{x} \in$ row vector in $\mathbb{R}^k$

$\mathbb{E}[c \cdot \underline{x}] = c\underline{\mu}$

$\operatorname{Cov}[c\underline{x}] = c\Sigma c'$

Data frame

$$\mathbb{X} = \begin{array}{c} \underline{x} \\ \begin{bmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \end{bmatrix} \end{array} \begin{array}{l} \leftarrow \underline{x}_1 \\ \\ \leftarrow \underline{x}_n \end{array}$$

Basic assumption: $\underline{x}_1, \dots, \underline{x}_n$ iid $\sim \underline{x}$

$\Sigma = \operatorname{cov}(\underline{x}) = \operatorname{Cov}(\underline{x}_i) \qquad i = 1,\dots,n$

$\mu = \mathbb{E}[\underline{x}] = \mathbb{E}[x_i] \qquad i = 1,\dots,n$

Data $\longmapsto$ estimate of $(\underline{\mu}, \Sigma)$

$$\mathbb{X} = \begin{bmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \end{bmatrix} \quad \xrightarrow{\text{for } \mu} \quad \boxed{\bar{\underline{x}} = \frac{1}{n} \sum_{i=1}^{n} \underline{x}_i}$$

$$\xrightarrow{\text{for } \Sigma} \quad \boxed{[S_{ij}] = \frac{1}{n} \sum_{i=1}^{n} (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'} \qquad S_{ij} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

Proposition

1. $\mathbb{E}[\bar{\underline{x}}] = \underline{\mu}$ ⟵ estimator $\bar{\underline{x}}$ is UNBIASED of the reality, in fact it's not always good to have unb. estimators. (if the variability is high we have a distorted comprension)
2. $Cov(\bar{\underline{x}}) = \frac{\Sigma}{n}$
3. $\mathbb{E}[S_n] = \frac{n-1}{n} \Sigma$

Corollary: $\mathbb{E}\left[\frac{n}{n-1} S_n\right] = \Sigma$ $\qquad \boxed{S = \frac{n}{n-1} S_n = \frac{1}{n-1} \sum_{i=1}^{n} (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'}$

Proof: 1.

$$\mathbb{E}[\bar{\underline{x}}] = \mathbb{E}\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\bar{x}_1] \\ \mathbb{E}[\bar{x}_2] \\ \vdots \\ \mathbb{E}[\bar{x}_p] \end{bmatrix} \qquad \mathbb{E}[\bar{x}_k] = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{E}[x_{ik}]}_{\mu_k \text{ identic. distribution}} = \mu_k \qquad k = 1, \dots, p$$

2. $$Cov(\bar{\underline{x}}) = \mathbb{E}[(\bar{\underline{x}} - \underline{\mu})(\bar{\underline{x}} - \underline{\mu})'] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(\underline{x}_i - \underline{\mu})\right)\left(\frac{1}{n}\sum_{j=1}^{n}(\underline{x}_j - \underline{\mu})\right)'\right] =$$
$$= \frac{1}{n^2} \sum_i \sum_j \mathbb{E}[(\underline{x}_i - \underline{\mu})(\underline{x}_j - \underline{\mu})'] = *$$

$$\mathbb{E}[(\underline{x}_i - \underline{\mu})(\underline{x}_j - \underline{\mu})'] \begin{cases} 0 & i \neq j \quad \text{independent} \\ \Sigma & i = j \quad \text{id. distributed} \end{cases}$$

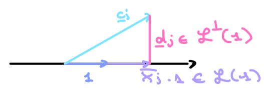
$$* = \frac{1}{n^2} \cdot n \cdot \Sigma = \frac{\Sigma}{n}$$

3. To do at home

Transposed Data set

$$\mathbb{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \qquad \underline{c}_1 \quad \underline{c}_2 \quad \underline{c}_p$$

$$\underline{d}_j = \underline{c}_j - \bar{x}_j \cdot \underline{1} \qquad j = 1, \dots, p$$
$$= \underline{c}_j - \frac{\underline{1}\underline{1}'}{\underline{1}'\underline{1}} \underline{c}_j = \underbrace{\left[I - \frac{\underline{1}\underline{1}'}{\underline{1}'\underline{1}}\right]}_{\text{orthogonal proj on } \mathcal{L}^\perp(\underline{1})} \underline{c}_j$$



$$\underline{d} = [\underline{d}_1, \dots, \underline{d}_p] = \left[I - \frac{\underline{1}\underline{1}'}{\underline{1}'\underline{1}}\right] \mathbb{X}$$

$$S = \frac{1}{n-1} \begin{bmatrix} \underline{d}_1'\underline{d}_1 & \underline{d}_1'\underline{d}_2 & \dots & \underline{d}_1'\underline{d}_p \\ & \ddots & & \\ & & & \underline{d}_p'\underline{d}_p \end{bmatrix} = \frac{1}{n-1} \underline{d}'\underline{d} = \frac{1}{n-1} \mathbb{X}' \left[I - \frac{\underline{1}\underline{1}'}{\underline{1}'\underline{1}}\right]\left[I - \frac{\underline{1}\underline{1}'}{\underline{1}'\underline{1}}\right] \mathbb{X}$$

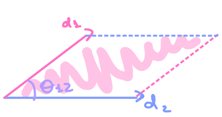
Multivariate variability?

Generalized variance: $Det(S)$

Total variance: $tr(S) = s_{11} + s_{22} + \dots + s_{pp}$

Why $Det(S)$ to assess variab?

$p = 2$



$$S = \begin{bmatrix} \underline{d}_1'\underline{d}_1 & \underline{d}_1'\underline{d}_2 \\ \underline{d}_2'\underline{d}_1 & \underline{d}_2'\underline{d}_2 \end{bmatrix} = \frac{1}{n-1} \begin{bmatrix} \|\underline{d}_1\|^2 & \|\underline{d}_1\|\|\underline{d}_2\| \cos\theta_{12} \\ \|\underline{d}_1\|\|\underline{d}_2\| \cos\theta_{12} & \|\underline{d}_2\|^2 \end{bmatrix}$$

$$Det(S) = \frac{1}{(n-1)^2} \|\underline{d}_1\|^2 \|\underline{d}_2\|^2 - \|\underline{d}_1\|^2 \|\underline{d}_2\|^2 \cos^2\theta_{12} = \boxed{\frac{1}{(n-1)^2} \|\underline{d}_1\|^2 \|\underline{d}_2\|^2 \sin^2\theta_{12}} \propto \text{Area}^2 \ (\text{parallelogram}(\underline{d}_1, \underline{d}_2))$$

$\theta_{12} = 0 \longrightarrow Det(S) = 0$

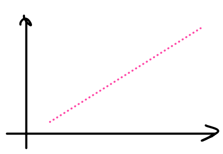
$\theta_{12} = \frac{\pi}{2} \longrightarrow Det(S)$ is MAX, keeping s_{11} and s_{22} ⟶ that's because they are uncorrelated

↳ if I change it the perimeter remains the same

$Det(S) = 0 \implies \underline{d}_1, \dots, \underline{d}_p$

$\implies \sum c_i \underline{d}_i = \underline{0}$ but $\underline{c} \neq \underline{0}$

WLOG $c_1 \neq 0 \implies \underline{c}_1 - \bar{x}_1 \underline{1} = \sum_{i=2}^{p} \frac{c_i}{c_1} (\underline{c}_j - \bar{x}_j \underline{1})$



Basically we collecting some thing with different unity of measure.

Corollary

If $p \geq n \implies \underline{d}_1, \dots, \underline{d}_p$ are linearly dep
$\iff \det(S') = 0$

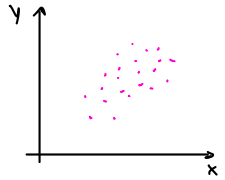
Proof: $p \geq n$ $\quad \underline{d}_1, \dots, \underline{d}_p \quad p$ vectors in $\mathbb{R}^n$

$\underline{d}_i \in \mathcal{L}^\perp(1) \quad \dim(\mathcal{L}^\perp(1)) \leq n-1$

$\implies \underline{d}_1, \dots, \underline{d}_p$ lin. dep. $\implies \det(S') = 0$

Example

Data $\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}$



Now let's construct another data frame $\longrightarrow$

$$\begin{array}{ccccccc} 1 & x & x^2 & \dots & x^p & y \end{array}$$

$$X = \left[\begin{array}{ccccc|c} 1 & x_1 & x_1^2 & \dots & x_1^p & y_1 \\ 1 & x_2 & x_2^2 & \dots & x_2^p & y_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^p & y_n \end{array}\right]$$

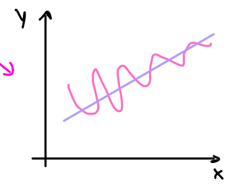
$\underbrace{\qquad\qquad}_{p+1 \text{ coef}}$

$p+1 = n-1$

$p+2 = n$

$\implies \mathrm{Det}(S) = 0$

$\implies y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + x_i^p$



You can always find a polynomial that can interpolate your data!

S' $p \times p$ symmetric entries are real numbers

Spectral decomposition thm

$\exists (\lambda_i, \underline{e}_i) \quad i = 1, \dots, p \quad$ s.t. $\quad \underline{e}_i \in \mathbb{R}^p \quad$ s.t. $\quad \underline{e}_i \underline{e}_j < \begin{matrix} 0 & i \neq j \\ 1 & i = j \end{matrix}$

$$\boxed{S' = \sum_{i=1}^{p} \lambda_i \underline{e}_i \underline{e}_i'}$$

Notation

$P = [\underline{e}_1 \dots \underline{e}_p]$

$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}$

$$\boxed{S = P \Lambda P'}$$

(Note since S is positive semidef $\lambda_i \geq 0$)

Proposition

S' is positive semi def. If $\mathrm{Det}(S') \neq 0 \implies$ positive def

Proof: $\forall \underline{c} \in \mathbb{R}^p \quad \underline{c}' S \underline{c} \geq 0 \quad$ (def of +- semidef)

but $\underline{c} \in \mathbb{R}^p \quad \underline{c}' S \underline{c} = \frac{1}{n-1} \underline{c}' d' d \underline{c} = \frac{1}{n-1} \| d \underline{c} \|^2$ are + semidef

If $\mathrm{Det}(S) \neq 0$ suppose $\exists \underline{c} \neq \underline{0}$ s.t. $\underline{c}' S' \underline{c} = 0 \implies \| d \underline{c} \| = 0 \implies d \underline{c} = \underline{e}$

$\implies d_1, \dots, d_n$ linearly indep $\implies \mathrm{Det}(S') = 0 \implies S'$ is pos. def

Convention

For now on $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ (with $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$ above)

Suppose $\mathrm{Det}(S') \neq 0$

Definition (Mahalanobis distance)

$\forall \underline{x}, \underline{y} \in \mathbb{R}^p. \quad d^2_{S}(\underline{x}, \underline{y}) = (\underline{x} - \underline{y})' S'^{-1} (\underline{x} - \underline{y})$

$S' = \sum_i^p \lambda_i \underline{e}_i \underline{e}_i'$ since $\mathrm{Det}(S') \neq 0 \implies \lambda_i > 0 \quad \forall i$

$S'^{-1} = \sum_i^p \frac{1}{\lambda_i} \underline{e}_i \underline{e}_i'$

Exercise

Check that $d_{S^{-1}}$ is a distance

Hint $S^{-1/2} = \sum_{i=1}^{p} \frac{1}{\sqrt{\lambda_i}} \underline{e}_i \underline{e}_i'$

$d^2(\underline{x}, \underline{y}) = [S^{-1/2} (\underline{x} - \underline{y})]' [S^{-1/2} (\underline{x} - \underline{y})]$

possible exam question

Example

$\mathcal{E}_r(\bar{x}) = \{ x \in \mathbb{R}^p : d^2_{S^{-1}}(x, \bar{x}) \le r^2 \}$ $\quad r > 0$ sphere

$S^{-1} = \sum \frac{1}{\lambda_i} e_i e_i'$

B $p \times p$ pos def

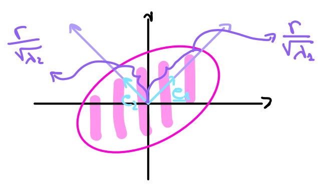
$\{ x : x' B x \le r^2 \}$ $\quad r > 0$

$\Rightarrow$ ellips in $\mathbb{R}^p$ $\quad B = \sum_{i=1}^{p} \lambda_i e_i e_i'$ $\quad$ center 0

Volume elipse $\propto r^p \frac{1}{\sqrt{\lambda_1}\sqrt{\lambda_2}\cdots\sqrt{\lambda_p}}$

$Vol(\mathcal{E}_r(\bar{x})) \propto r^p \prod_{i=1}^{p} \sqrt{\lambda_i}$

$\propto r^p \sqrt{\prod_{i=1}^{p} \lambda_i} \rightarrow Det(S)$



EXAMPLE

X data

$\bar{x}$ sample mean

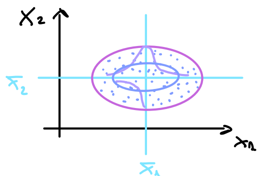
S sample covarience

Assume $det(S) \neq 0$

$d_{S^{-1}} : \mathbb{R}^p \times \mathbb{R}^p \longrightarrow [0, +\infty)$

$d^2_{S^{-1}}(x, y) = (x - y)' S^{-1} (x - y) \quad \forall x, y \in \mathbb{R}^p$ $\quad$ Mahalanobis dist

$p = 2 \qquad S = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix}$



$\mathcal{E}_{r=1} = \{ (x - \bar{x})' S^{-1} (x - \bar{x}) \le 1 \}$

$Volume(\mathcal{E}_{r=1}(\bar{x})) \propto \sqrt{Det(S)} = \sqrt{\lambda_1 \cdot \lambda_2} = \sqrt{S_{11} S_{22}}$

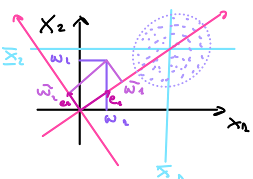
$\hookrightarrow$ Generale variance

Different data set: $\quad$ same $\bar{x}$

Different S: $\quad S^* = \begin{bmatrix} S_{11}^* & 0 \\ 0 & S_{22}^* \end{bmatrix} \qquad \begin{matrix} S_{11}^* > S_{11} \\ S_{22}^* > S_{22} \end{matrix}$

If S_{11} increases or $S_2 \uparrow \implies Volume(\mathcal{E}_{r=1}(\bar{x})) \uparrow$

Generale case $\mathbb{R}^p$



Axes of the system: $\quad S = \sum_{i=1}^{p} \lambda_i e_i e_i' \qquad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$

$\frac{e_1 e_1'}{e_1' e_1} w = (e_1' w_1) e_1$

OLD SYSTEM $\qquad\qquad$ NEW SYSTEM

$w = \begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix} \longleftrightarrow \tilde{w} = \begin{pmatrix} e_1' w \\ \vdots \\ e_p' w \end{pmatrix} = \boxed{P' w}$

$X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} \longleftrightarrow \tilde{X} = \begin{bmatrix} (e_1' x_1)' \\ \vdots \\ (e_1' x_n)' \end{bmatrix} = \begin{bmatrix} x_1' P \\ \vdots \\ x_n' P \end{bmatrix} = \boxed{X P}$

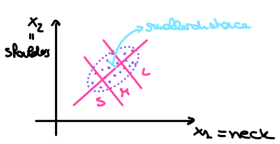
$S = \frac{1}{n-1} X' \left(I - \frac{11'}{1'1}\right) X \longleftrightarrow \tilde{S} = \frac{1}{n-1} \tilde{X}' \left(I - \frac{11'}{1'1}\right) \tilde{X} = \frac{1}{n-1} P' X' \left(I - \frac{11'}{1'1}\right) X P = P' S P = P' P \Lambda P' P = \boxed{\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}}$

General Variance: $\quad Det(S) = \prod_{i=1}^{p} \lambda_i = $ Gen Variation: $Det(\tilde{S}) = \prod_{i=1}^{p} \lambda_i$

Total Variance: $\quad tr(S) = \sum_{i=1}^{p} \lambda_i = tr(\tilde{S})$

EXAMPLE

T-shirt sizes



Basically I want to not depend on 2 components, but only one, I need something to have both informations

PRINCIPAL COMPONENT ANALYSIS (PCA)

$\underline{x}$ random vector in $\mathbb{R}^p$

$\underline{x} \sim \underline{\mu}, \Sigma$

Q_1: Find $\underline{a} \in \mathbb{R}^p$ $\quad \|\underline{a}\| = 1$ s.t. $Var(\underline{a}'\underline{x})$ is max $\longrightarrow \underline{a}_1$

Q_2: Find $\underline{a} \in \mathbb{R}^p$ $\quad \|\underline{a}\| = 1$ s.t. $Var(\underline{a}'\underline{x})$ is max and $Cov(\underline{a}_1'\underline{x}, \underline{a}\underline{x}) = 0 \longrightarrow \underline{a}_2$

$\vdots$

$i = 1, \dots, p-1$

Q_p: Find $\underline{a} \in \mathbb{R}^p$ $\quad \|\underline{a}\| = 1$ s.t. $Var(\underline{a}'\underline{x})$ is max and $Cov(\underline{a}_i'\underline{x}, \underline{a}\underline{x}) = 0 \longrightarrow \underline{a}_p$

Lemma

B $p \times p$ symmetric, pos def matrix s.t. $B = \sum_{i=1}^{p} \lambda_i \underline{e}_i \underline{e}_i'$

1. $\max_{\underline{x} \in \mathbb{R}^p} \frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}} = \lambda_1$, $\quad$ argmax $= \underline{e}_1$

2. $\max_{\substack{\underline{x} \in \mathbb{R}^p \\ \underline{x} \perp \underline{e}_1}} \frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}} = \lambda_2$, $\quad$ arg max $= \underline{e}_2$

$\vdots$

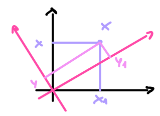
3. $\max_{\substack{\underline{x} \in \mathbb{R}^p \\ \underline{x} \perp \underline{e}_1 \\ \vdots \\ \underline{x} \perp \underline{e}_{p-1}}} \frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}} = \lambda_p = \min_{\underline{x} \in \mathbb{R}^p} \frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}}$, $\quad$ argmax $= \underline{e}_p$ $\quad$ argmin $= \underline{e}_p$

Proof: 1. $\underline{x} \neq 0$

$$\frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}} = \frac{\underline{x}'P\Lambda P'\underline{x}}{\underline{x}'\underline{x}} = \frac{\underline{x}'P\Lambda P'\underline{x}}{\underline{x}'PP'\underline{x}}$$

$\underline{y} = P'\underline{x}$

$$= \frac{\underline{y}'\Lambda\underline{y}}{\underline{y}'\underline{y}} = \sum_{i=1}^{p} \frac{\lambda_i y_i^2}{\sum y_i^2} \leq \lambda_1 \frac{\sum y_i^2}{\sum y_i^2} = \lambda_1$$



Take $\underline{x} = \underline{e}_1$ $\quad \frac{\underline{e}_1'B\underline{e}_1}{\underline{e}_1'\underline{e}_1} = \lambda_1 \frac{\underline{e}_1'\underline{e}_1}{\underline{e}_1'\underline{e}_1} = \lambda_1 \implies \max_{\underline{x} \in \mathbb{R}^p} \frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}} = \lambda_1$ $\quad$ argmax $= \underline{e}_1$

2. $\underline{x} \neq 0$, $\underline{x} \perp \underline{e}_1$

$$\frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}} = \frac{\underline{x}'P\Lambda P'\underline{x}}{\underline{x}'\underline{x}} = \frac{\underline{x}'P\Lambda P'\underline{x}}{\underline{x}'PP'\underline{x}} = \frac{\underline{y}'\Lambda\underline{y}}{\underline{y}'\underline{y}} = \sum_{i=1}^{p} \frac{\lambda_i y_i^2}{\sum y_i^2} \leq \lambda_2$$

Take $\underline{x} = \underline{e}_2$ $\quad \frac{\underline{e}_2'B\underline{e}_2}{\underline{e}_2'\underline{e}_2} = \lambda_2 \frac{\underline{e}_2'\underline{e}_2}{\underline{e}_2'\underline{e}_2} = \lambda_2 \implies \max_{\underline{x} \in \mathbb{R}^p} \frac{\underline{x}'B\underline{x}}{\underline{x}'\underline{x}} = \lambda_2$ $\quad$ argmax $= \underline{e}_2$

A_1: $\max_{\substack{\underline{a} \in \mathbb{R}^p \\ \|\underline{a}\| = 1}} Var(\underline{a}'\underline{x}) = \max_{\substack{\underline{a} \in \mathbb{R}^p \\ \|\underline{a}\| = 1}} \underline{a}'\Sigma\underline{a} = \max_{\underline{a} \in \mathbb{R}^p} \frac{\underline{a}'\Sigma\underline{a}}{\|\underline{a}\|\,\|\underline{a}\|} = \max_{\underline{a} \in \mathbb{R}^p} \frac{\underline{a}'\Sigma\underline{a}}{\underline{a}'\underline{a}} = \lambda_1$ (lemma)

argmax $= \underline{e}_1$

A_2: $\max_{\substack{\underline{a} \in \mathbb{R}^p \\ \|\underline{a}\| = 1 \\ Cov(\underline{e}_1'\underline{x}_1, \underline{a}\underline{x}) = 0}} Var(\underline{a}'\underline{x}) = \max_{\substack{\underline{a} \in \mathbb{R}^p \\ Cov(\,) = 0}} \frac{\underline{a}'\Sigma\underline{a}}{\underline{a}'\underline{a}}$

$= \max_{\substack{\underline{a} \in \mathbb{R}^p \\ \underline{a} \perp \underline{e}_1}} \frac{\underline{a}'\Sigma\underline{a}}{\underline{a}'\underline{a}} = \lambda_2$

argmax $= \underline{e}_2$

$Cov(\underline{e}_1'\underline{x}, \underline{a}'\underline{x})$

$C = \begin{bmatrix} \underline{e}_1' \\ \underline{a}' \end{bmatrix}$ $\quad 2 \times p$

$$Cov(C\underline{x}) = C\Sigma C' = \begin{bmatrix} \underline{e}_1' \\ \underline{a}' \end{bmatrix} \Sigma \begin{bmatrix} \underline{e}_1' & \underline{a}' \end{bmatrix} = \begin{bmatrix} \underline{e}_1'\Sigma\underline{e}_1 & \underline{e}_1'\Sigma\underline{a} \\ \underline{a}'\Sigma\underline{e}_1 & \underline{a}'\Sigma\underline{a} \end{bmatrix}$$

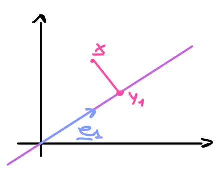
$$0 = Cov(\underline{a}'\underline{x}, \underline{e}_1'\underline{x}) = \underline{e}_1'\Sigma\underline{a} = \underline{a}'\Sigma\underline{e}_1 = \lambda_1 \underline{a}'\underline{e}_1$$

$\Updownarrow$

$\underline{a} \perp \underline{e}_1$

By product A_p: $\quad \min_{\substack{\underline{a} \in \mathbb{R}^p \\ \|\underline{a}\| = 1}} Var(\underline{a}'\underline{x}) = \lambda_p$

argmin $= \underline{e}_p$



Def

Σ, $\Sigma = \sum_{i=1}^{p} \lambda_i \underline{e}_i \underline{e}_i'$ $\qquad y_1 = \underline{e}_1'\underline{x}$ $\quad$ score of 1-st PC (PC1)

$\underline{e}_1$ $\quad$ loading of 1-st PC (PC1)

(y_1, e_1) PC1

For $i = 1, \dots, p$: $y_i = \underline{e}_i' \underline{x}$ is the score on the i-th PC (PCi) (or $y = \underline{e}_i'(\underline{x} - \underline{\mu})$)

$\underline{e}_i$ is the loading of the i-th PC

$y_i = e_{1i} x_1 + e_{2i} x_2 + \dots + e_{pi} x_p$

Corollary

$$Cov(y_i, y_j) = \lambda_i \delta_{ij} = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}$$

Proof: $Cov(\underline{e}_i' \underline{x}, \underline{e}_j' \underline{x}) = \underline{e}_i' \Sigma \underline{e}_j = \lambda_j \underline{e}_i' \underline{e}_j = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}$

PCA (again)

$\underline{x}$ r. vect. $\in \mathbb{R}^p$

$\mathbb{E}[\underline{x}] = \underline{\mu}$

$Cov(\underline{x}) = \Sigma = P \Lambda P'$

$\det(\Sigma) \neq 0$

PC's : $\underline{y} = P'(\underline{x} - \underline{\mu}) = \begin{bmatrix} \underline{e}_1'(\underline{x} - \underline{\mu}) \\ \vdots \\ \underline{e}_p'(\underline{x} - \underline{\mu}) \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$

$\mathbb{E}[\underline{y}] = \underline{0}$

$Cov(\underline{y}) = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}$

$\lambda_1 \geq \lambda_2 \geq \dots$ | k $\geq \lambda_p$

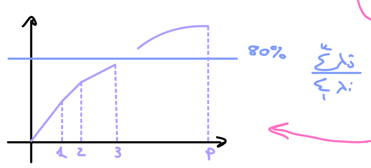
$Var(y_1)$ $Var(y_2)$ $Var(y_p)$

$\frac{\lambda_1}{\sum \lambda_i}$ $\frac{\lambda_2}{\sum \lambda_i}$ $\cancel{\frac{\lambda_p}{\sum \lambda_i}}$

$\frac{\lambda_1}{\sum \lambda_i}$ $\frac{\lambda_1 + \lambda_2}{\sum \lambda_i}$

choose k s.t. $\frac{\sum_{j=1}^{k} \lambda_j}{\sum \lambda_i} \geq \tau$ treshold (80%, 90%)

prop. of total variance ($tr(\Lambda) = tr(\Sigma) = \sum \lambda_j$) explained



We describe the fenomenon with k elements, not all of them, so we have a different representation of the same fenomenon

Interpretation? For $i = 1, \dots, p$ y_i i-th princ. component

$y_i = \underline{e}_i'(\underline{x} - \underline{\mu}) = e_{1i}(x_1 - \mu_1) + e_{2i}(x_2 - \mu_2) + \dots + e_{pi}(x_p - \mu_p)$

Proposition

For $i, k = 1, \dots, p$ $Corr(y_i, X) = e_{ki} \frac{\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$

Proof: $Corr(y_i, x_k) = \frac{Cov(y_i, x_k)}{\sqrt{\lambda_i} \sqrt{\sigma_{kk}}}$ $\underline{u}_k = \frac{\lambda_i e_{ki}}{\sqrt{\lambda_i}\sqrt{\sigma_{kk}}} = \frac{\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} e_{ki}$

$\underline{u}_k = (0 \dots \overset{k}{1} \dots 0)'$

$Cov(\underline{e}_i' \underline{x}, \underline{u}_k' \underline{x}) - \underbrace{\cancel{Cov(\underline{e}_i' \underline{\mu}, \underline{u}_k' \underline{x})}}_{0} = \underline{e}_i' \Sigma \underline{u}_k = \underline{u}_k' \Sigma \underline{e}_i = \lambda_i \underline{u}_k' \underline{e}_i = \lambda_i e_{ki}$

Hence: $corr(y_i, x_k) \propto \frac{e_{ki}}{\sqrt{\sigma_{kk}}}$

So if $\sigma_{11}, \sigma_{22}, \dots, \sigma_{kk}$ are of the same "order" of magnitude $\Longrightarrow$ $Corr(y_i, x_k) \propto e_{ki}$

Else if $\sigma_{11}, \sigma_{22}, \dots, \sigma_{kk}$ are VERY different $\Longrightarrow$ standardize $\underline{x}$ first

$\underline{z} = V^{1/2}(\underline{x} - \underline{\mu})$ $V = \begin{bmatrix} \sigma_{11} & & \\ & \ddots & \\ & & \sigma_{pp} \end{bmatrix}$ $\mathbb{E}[\underline{z}] = \underline{0}$

$Cov(\underline{z}) = V^{-1/2} \underset{\overset{\parallel}{\rho}}{\Sigma} V^{-1/2} = \underline{\rho} \Lambda \underline{\rho}$ for $\overset{\rho}{\overset{\parallel}{Cor(\underline{z})}}$ (not for Σ)

$\rho = \sum_{i=1}^{p} \lambda_i \underline{e}_i \underline{e}_i'$ $\underline{y} = \underline{\rho}' \underline{z} = \underline{\rho}' V^{-1/2}(\underline{x} - \underline{\mu})$

→ note $\rho = P \Lambda P'$

$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix}$ $\quad tr(\Lambda) = tr(\rho) = p = \sum \lambda_i$

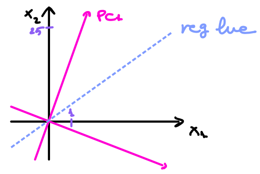
Ave value of λ_i's : $\frac{\sum \lambda_i}{p} = 1$

A simple rule of thumb for selecting components :

choose $Y_1, \dots, Y_k$ if $\lambda_1, \dots \lambda_k \geq 1 = \frac{\sum \lambda_i}{p}$ discard the other components

Example

$\mu = 0 \qquad \Sigma = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix} = cov(\underline{X}) = cov\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$



PC1 line: $x_2 = \frac{0,999}{0,04} x_1$

$x_2 \simeq 25 x_1$

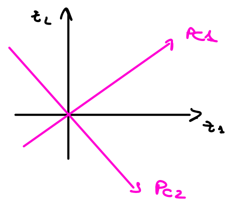
$\Sigma = \lambda_1 \underline{e}_1 \underline{e}_1' + \lambda_2 \underline{e}_2 \underline{e}_2'$ $\qquad \lambda_1 = 100,16 \qquad \underline{e}_1 = (0,04 \quad 0,999)'$

$\lambda_2 = 0,84 \qquad \underline{e}_2 = (0,999 \quad -0,04)'$

$\rho = \begin{bmatrix} 1 & \frac{4}{1\cdot 10} \\ \frac{4}{10} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0,4 \\ 0,4 & 1 \end{bmatrix}$ $\qquad$ Regression line $\frac{x_2 - 0}{\sqrt{100}} = 0,4 \frac{x_1 - 0}{1} \iff X_2 = 4 X_1$

$\lambda_1 = 1,4 \qquad \underline{e}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)'$

$\lambda_2 = 0,6 \qquad \underline{e}_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)'$



Regression line: $z_1 = 0,4 z_2$

$x_1 = 4 x_2$

Common situation

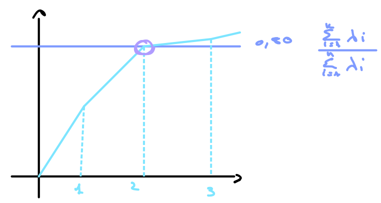
μ, Σ unknown $\qquad$ Estimate with

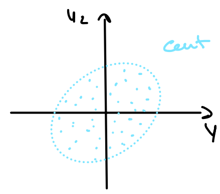
$\mathbb{X}$ $n \times p$ Data $\qquad \Sigma \quad S \leftarrow$ PCA

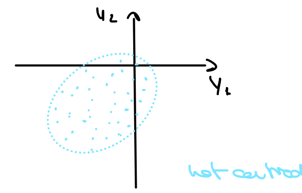
$\mu \quad \bar{\underline{x}}$

$S = \sum_{i=1}^{p} \lambda_i \underline{e}_i \underline{e}_i'$ $\qquad \mathbb{X} = \begin{array}{c} \begin{matrix} x_1 & x_2 & \cdots & x_p \end{matrix} \\ \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \end{array} = \begin{array}{c} \underline{x} \\ \begin{bmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_n' \end{bmatrix} \end{array} \implies \mathbb{Y} = \begin{array}{c} \underline{y} \\ \begin{bmatrix} (\underline{P}'\underline{x}_1)' \\ \vdots \\ (\underline{P}'\underline{x}_n)' \end{bmatrix} \end{array} = \mathbb{X}\underline{P}$

$\mathbb{Y} = \begin{array}{c} \begin{matrix} y_1 & y_2 & \cdots & y_k & \cdots & \cancel{y_p} \end{matrix} \\ \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} & \cdots & \cancel{y_{1p}} \\ \vdots & & & & & \\ y_{n1} & y_{n2} & \cdots & y_{nk} & \cdots & \cancel{y_{np}} \end{bmatrix} \end{array}$







If $\lambda_p \simeq 0 \qquad Var(Y_p) \approx 0$

$Y_p = e_{1p}(X_1 - \mu_1) + \dots + e_{pp}(X_p - \mu_p) = 0 \longrightarrow$ Totally deterministic values, no variability

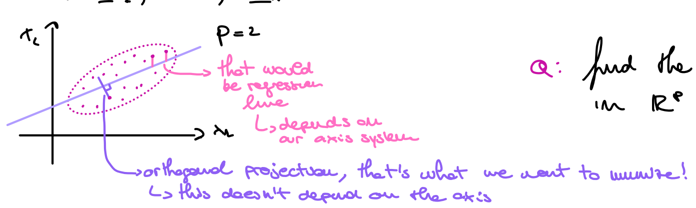
$\mathbb{E}[Y_p] = 0$

Observations:

1. PCA for categorical variables is called "correspondence analysis" and works on contingency tables.

		x_2 Color of eyes			
		Blu	Brown	Green	
x_1 Gender	M	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
	F	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
					n

2. Data $\underline{X}_1, \dots, \underline{X}_n \in \mathbb{R}^p$



Q: find the linear subspace ($\mathcal{L}$) of dimension k $(k < p)$ in $\mathbb{R}^p$ closest to the Data.

Q: Find $\underline{q}_1, \dots, \underline{q}_k$ orthonormal s.t. $\mathcal{L} = \text{span}(\underline{q}_1, \dots, \underline{q}_k)$ is closest to $\underline{x}_1, \dots, \underline{x}_n$. Op center data, $\underline{x}_1 - \overline{\underline{x}}, \dots, \underline{x}_n - \overline{\underline{x}}$

1. Projection $\underline{x}_i - \overline{\underline{x}}$ on $\mathcal{L}$: $\pi_{\underline{x}_i \mid \mathcal{L}} = \sum_{j=1}^{n} \underline{q}_j \underline{q}_j'(\underline{x}_i - \overline{\underline{x}})$

Find $\underline{q}_1, \dots, \underline{q}_k$ orthon in $\mathbb{R}^p$ s.t. $\boxed{\sum_{i=1}^{n} \| (\underline{x}_i - \overline{\underline{x}}) - \sum_{j=1}^{n} \underline{q}_j \underline{q}_j'(\underline{x}_i - \overline{\underline{x}}) \|^2 \text{ is minimize}}$ optimization problem

$\underline{v}_i = \underline{x}_i - \overline{\underline{x}} \qquad i = 1, \dots, n$

$$\sum_{i=1}^{n} \| \underline{v}_i - \sum_{j=1}^{n} \underline{q}_j \underline{q}_j' \underline{v}_i \|^2 = \sum_{i=1}^{n} (\underline{v}_i - \sum_j \underline{q}_j \underline{q}_j' \underline{v}_i)'(\underline{v}_i - \sum_j \underline{q}_j \underline{q}_j' \underline{v}_i) = \underline{v}_i'\underline{v}_i - 2\sum_j \underline{v}_i' \underline{q}_j \underline{q}_j' \underline{v}_i + (\sum_j \underline{q}_j \underline{q}_j' \underline{v}_i)'(\sum_t \underline{q}_t \underline{q}_t' \underline{v}_i)$$

$$= \underline{v}_i'\underline{v}_i - \sum_j \underline{v}_i' \underline{q}_j \underline{q}_j' \underline{v}_i \qquad + \quad \underline{q}_j'\underline{q}_t = \begin{cases} 0 & j \neq t \\ 1 & j = t \end{cases}$$

$$= \sum_{i=1}^{n} \underline{v}_i'\underline{v}_i - \sum_{i=1}^{n}\sum_{j=1}^{k} \underline{v}_i' \underline{q}_j \underline{q}_j' \underline{v}_i = \sum_{i=1}^{n} \underline{v}_i'\underline{v}_i - \sum_{i=1}^{n}\sum_{j=1}^{k} \underline{q}_j' \underline{v}_i \underline{v}_i' \underline{q}_j =$$

$$= \sum_{i=1}^{n} \underline{v}_i'\underline{v}_i - \sum_{j=1}^{k} \underline{q}_j' \sum_{i=1}^{n} \underline{v}_i \underline{v}_i' \underline{q}_j = \sum_{i=1}^{n} \underline{v}_i'\underline{v}_i - \underbrace{(n-1)\sum_{j=1}^{k} \underline{q}_j' S \underline{q}_j}_{\circledast}$$

$\circledast$ $k = 1$ $\quad \max_{\substack{\underline{q}_1 \in \mathbb{R}^p \\ \|\underline{q}_1\| = 1}} (n-1)\underline{q}_1' S \underline{q}_1 = (n-1)\lambda_1 \qquad \text{argmax} = \underline{e}_1 \Longrightarrow \underline{q}_1 = \underline{e}_1 \qquad S = \sum_{i=1}^{p} \lambda_i \underline{e}_i \underline{e}_i'$

By induction $\underline{q}_1 = \underline{e}_1, \dots, \underline{q}_k = \underline{e}_k$

$\mathcal{L} = \text{span}(\underline{e}_1, \dots, \underline{e}_k)$

Approx error: $\sum_{i=1}^{n} \| (\underline{x}_i - \overline{\underline{x}}) - \sum_{j=1}^{k} \underline{e}_j \underline{e}_j' (\underline{x}_i - \overline{\underline{x}}) \|^2 = (n-1) \sum_{k=1}^{p} \lambda_i$