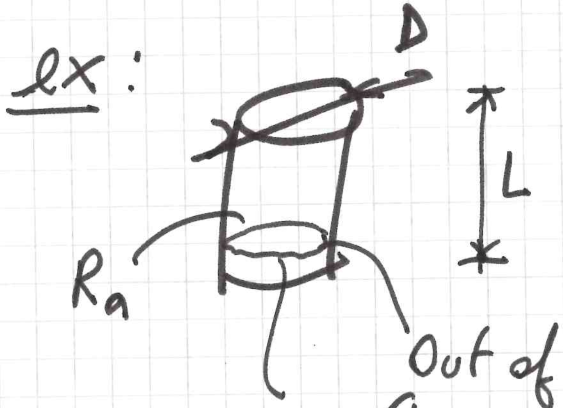
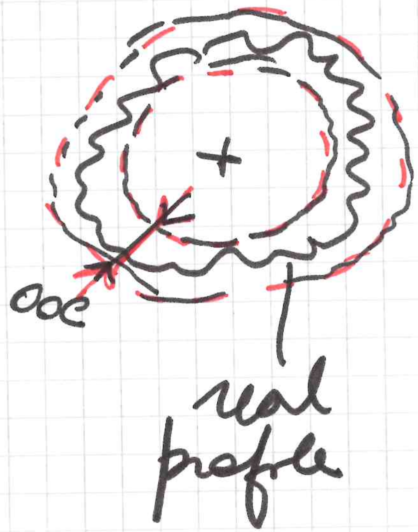
PCA

PRINCIPAL COMPONENT ANALYSIS

ex:



Out of circularity OOC
(how far we are from a perfect circle



Quality of this PIN depends on

$$\underline{X}' = [x_1 \; x_2 \; x_3 \; x_4]$$

$x_1 = D$
$x_2 = L$
$x_3 = R_a$
$x_4 = OOC$

$\downarrow$
vector of quality features

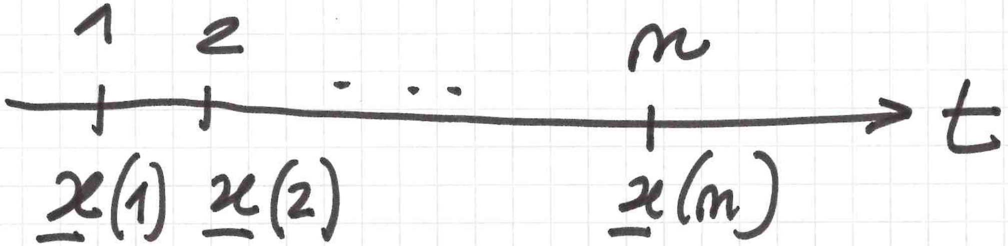
$$\underline{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \qquad \underline{X}' = [x_1 \; \dots \; x_p]$$

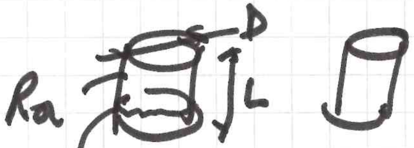
$\underline{X}$

p-variate vector

$p = \#$ variables

$m = \#$ observations





ooc

$$\underset{1\times1 = 1\times p}{c} = \underset{}{\underline{a}'} \underset{p\times1}{\underline{X}} = [a_1 \; a_2 \; \dots \; a_p] \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$$

$\underline{a}$: vector of constant numbers

$\underline{X}$: p-variate random variable

→ linear combination of x_i's

can represent most of the "content" in the p-variate $\underline{X}$ random variable

$V(\underline{a}'\underline{X})$? $E(\underline{a}'\underline{X})$?

Assume $\underline{X} \sim N_p(\underline{\mu}, \underline{\Sigma})$ $X \sim N(\mu, \sigma^2)$ $(p=1)$

p-variate normal

$$\underline{\mu} = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

$\underline{\Sigma}$ = VARIANCE COVARIANCE MATRIX $p \times p$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdots \\ \cdots & & & \sigma_p^2 \end{bmatrix}$$

$\sigma_{ji} = \sigma_{ij} = Cov(X_i, X_j)$ $i \neq j$

$\sigma_i^2 = Var(X_i)$ $i = 1 \ldots p$

(6)

$$E(\underline{a}'\underline{X}) = \underline{a}'E(\underline{X}) = \underline{a}'\underline{\mu} = [a_1 \dots a_p]\begin{bmatrix}\mu_1\\ \vdots \\ \mu_p\end{bmatrix} =$$

$$= a_1\mu_1 + a_2\mu_2 + \dots + a_p\mu_p$$

$$V(\underline{a}'\underline{X}) \underset{\text{by def}}{\overset{\Delta}{=}} E\left[(\underline{a}'\underline{X} - E(\underline{a}'\underline{X}))(\underline{a}'\underline{X} - E(\underline{a}'\underline{X}))'\right] =$$

$$= E\left[(\underline{a}'\underline{X} - \underline{a}'\underbrace{E(\underline{X})}_{\underline{\mu}})(\underline{a}'\underline{X} - \underline{a}'\underbrace{E(\underline{X})}_{\underline{\mu}})'\right] =$$

$$= E\left[(\underline{a}'(\underline{X}-\underline{\mu}))(\underline{a}'(\underline{X}-\underline{\mu}))'\right] = \quad \text{for } p=1 \quad E[(X-\mu)^2]$$

$$(\underline{A}\underline{B}')' = (\underline{B}\underline{A}') \qquad = E\left[\underline{a}'(\underline{X}-\underline{\mu})((\underline{X}-\underline{\mu})'\underbrace{(\underline{a}')'}_{\underline{a}})\right]$$

$$= E\left[\underline{a}'(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})'\underline{a}\right] = \underline{a}'\underbrace{E\left[(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})'\right]}_{\underline{\Sigma}}\underline{a}$$

$$= \underset{1\times p}{\underline{a}'}\ \underset{p\times p}{\underline{\Sigma}}\ \underset{p\times 1}{\underline{a}} = 1\times 1$$

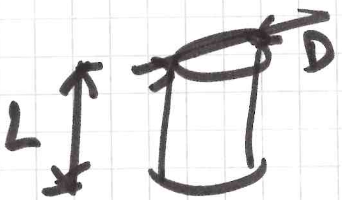
$p=1 \quad V(aX) = a^2 V(X) = a^2\sigma^2 = a\,\sigma^2 a = a'\sigma^2 a$

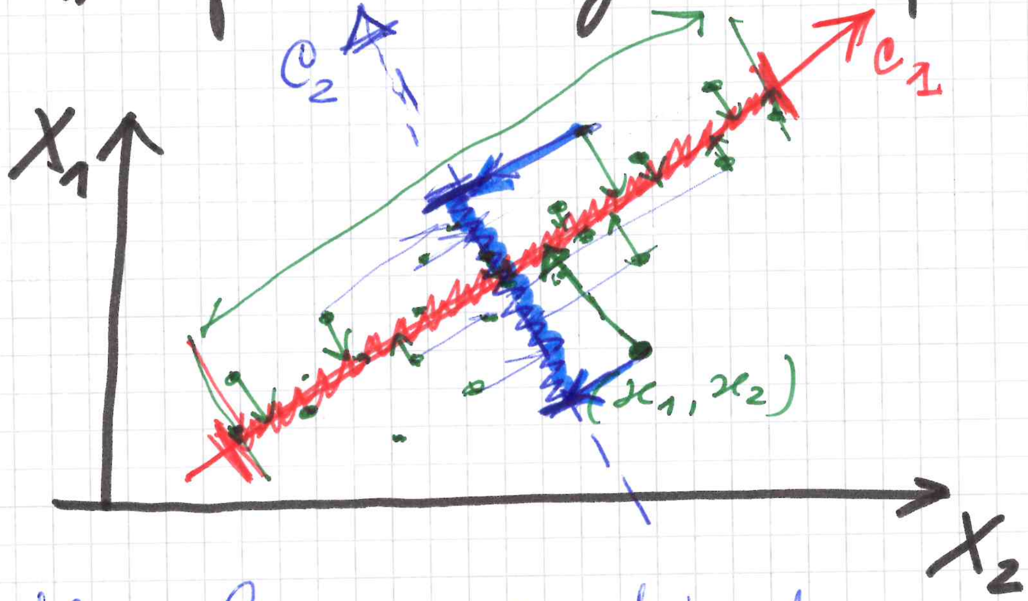
$a' = a$

$p \neq 1 \quad V(\underline{a}'\underline{X}) = \underline{a}'\underline{\Sigma}\,\underline{a}$

What linear combination ? to "represent" most of the "content" of the original p-variate random vector

$p = 2$

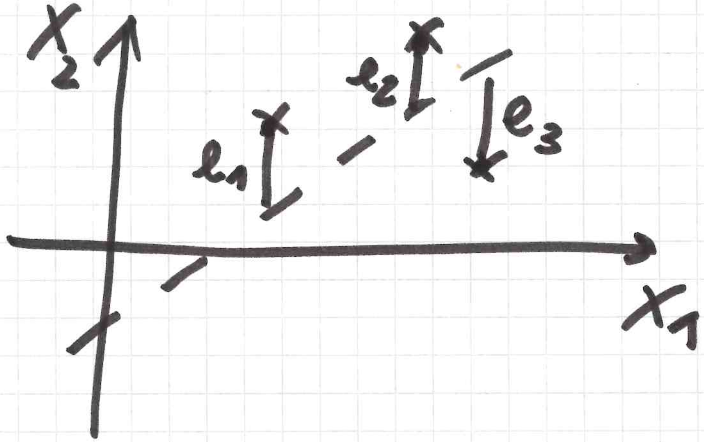




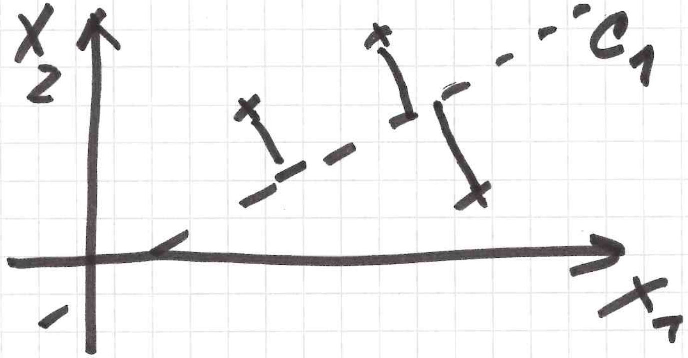
$a_1 X_1 + a_2 X_2 = c$

Let's find the linear combination that can represent most of the variability observed in this point cloud

(6)



MIN $\sum e_i^2$



OLS

$X_2 = \beta_0 + \beta_1 X_1$

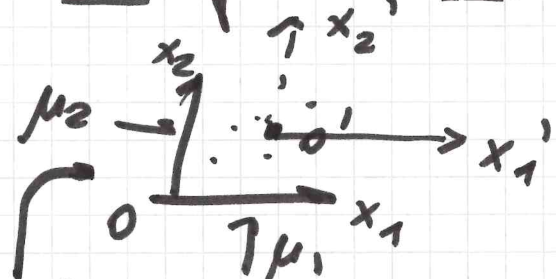
R.V. — regressor

$X_2 | X_1 = x_1 = \beta_0 + \beta_1 x_1$

MIN SSE

PCA

MIN sum of squared distances orthogonal to the direction of the PC



$E(\underline{X}') = \underline{0}'$

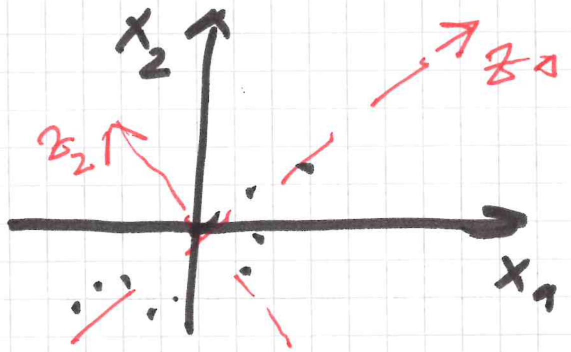
from slide 10 on $\rightarrow$ $\underline{X}' = (\underline{X} - \underline{\mu})'$ MOVING THE ORIGIN TO THE POINT GIVEN by $\underline{\mu}' = [\mu_1 \dots \mu_p]$

$$\underline{X}_p = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \longrightarrow \underline{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix}$$

(after centering in $\underline{\mu}$)

after subtracting the μ_i to x_i for $i = 1 \dots p$

PRINCIPAL COMPONENTS



① $z_1 = \underline{\alpha}_1' \underline{x} \ni$

$z_1 \ni$ Max variance

$Var(z_1) = Var(\underline{\alpha}_1' \underline{x})$ MAX

② then $z_2 = \underline{\alpha}_2' \underline{x} \ni$

1) uncorrelated with the previous PC's

2) Max variance

⋮ → ⓟ steps

8) 1^{st} PC $\rightarrow \max_{\underline{\alpha}_1'} Var(z_1) = \max Var(\underline{\alpha}_1' \underline{x})$

s.t. $\underline{\alpha}_1' \underline{\alpha}_1 = 1$ → subject to

$\alpha_{11}^2 + \alpha_{12}^2 + \dots \alpha_{1p}^2 = 1$

$\underline{\alpha}_1' \underline{\alpha}_1 - 1 = 0$

Lagrange multiplier

$$\max \; Var(\underline{\alpha}_1' \underline{x}) - \lambda (\underline{\alpha}_1' \underline{\alpha}_1 - 1)$$

$$\underline{\max} \; \cancel{\underline{\alpha}_1'} \underline{\Sigma} \underline{\alpha}_1 - \lambda (\underline{\alpha}_1' \underline{\alpha}_1 - 1) = f$$

(univariate $\alpha^2 \sigma^2$)

$$\frac{\partial f}{\partial \underline{\alpha}_1'} = 0 \Rightarrow \boxed{\underline{\Sigma} \underline{\alpha}_1 - \lambda \underline{\alpha}_1 = 0} \rightarrow (\underline{\Sigma} - \lambda \underline{I}) \underline{\alpha}_1 = 0$$

$$(\underline{\Sigma} - \lambda \underline{I}) \underline{\alpha}_1 = 0$$

EIGENVALUE λ

$\underline{\alpha}_1$: EIGENVECTOR of $\underline{\Sigma}$

$$\underline{\underline{MAX}}\ \underbrace{Var(z_1)} = Var(\underline{\alpha}_1' \underline{x}) = \dots = \lambda_1$$

$\uparrow$ EIGENVECTOR of $\underline{\Sigma}$

$\uparrow$ **CORRECT!**

$\uparrow$ CORRESPONDING EIGENVALUE

First PRINCIPAL COMPONENT is defined by the eigenvector which corresponds to the $\underline{\underline{\text{max}}}$ eigenvalue λ_1

$\vdots$

Second PC .... eigenvector $\underline{\alpha}_2'$ corresponding to λ_2 which is the max after λ_1

$$\begin{array}{llll}
MAX & \lambda_1 & \underline{\alpha}_1' & \longrightarrow z_1 = \underline{\alpha}_1' \underline{x} \\
 & > & & \\
\downarrow & \lambda_2 & \underline{\alpha}_2' & \longrightarrow z_2 = \underline{\alpha}_2' \underline{x} \\
 & > & \vdots & \\
MIN & \lambda_3 & \underline{\alpha}_3' & \longrightarrow z_3 = \underline{\alpha}_3' \underline{x}
\end{array}$$

$\underline{X} \rightarrow \underline{Z}$ $\qquad$ $\underline{z}_1 = \underline{\alpha}_1' \underline{x}$

VARIABLES $\qquad$ PC's $\qquad$ $\boxed{\text{LOADINGS}}$

data observed

$\underline{x}_i$

time i

$i = 1 \dots n$

~~$\underline{z}_1 = \underline{\alpha}_1' \underline{x}$~~

$\underline{z}_i \leftarrow \underline{x}_i$ OBSERVAT

SCORE

(Realizations of PC's)