L11. PCA : problem solving :)

Hi !! :) Check !!

4_ST3188_2020.pdf → 4a

	cr 1	cr 2	...	cr 10
	0-10	0-10	0-10	0-10
appl 1				
appl 2				
⋮				
appl n				

$\frac{\lambda_1+\lambda_2+\ldots+\lambda_i}{\lambda_1+\lambda_2+\ldots+\lambda_k}$

$\frac{\lambda_i}{\lambda_1+\lambda_2+\ldots+\lambda_k}$

i λ_i

Barlett's test

Communalities

k - number of vars

n —//— of obs.

component	total	% of var	cum
1	5.3	53 %	53 %
2	1.6	16 %	69 %
3	1.1	11 %	81 %
. . .	. . .	6 %	. . .
		3 %	
		⋮	

eigen values of C

? what is it ?

? how may we use it ?

$$\underset{10\times 10}{C} = \begin{bmatrix} 1 & & & \cdots \\ & 1 & & \\ & & 1 & \\ \vdots & & & 1 \end{bmatrix}$$ ← sample corr. matrix.

$\lambda_1 = 5.3 \quad \lambda_2 = 1.6 \quad \lambda_3 = 1.1 \ldots$

eigenvector v_i

$C \cdot v_i = \lambda_i \cdot v_i$

scaling factor for eigenvector.

λ_i – eigenvalue of C

$\lambda_i = \mathrm{sVar}(p_i)$ sample variance

[standard-d]

$$X = U \cdot \Sigma \cdot V^T$$

Σ – diagonal $= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ \vdots & \sigma_2 & & \vdots \\ 0 & \cdots & 0 & \ddots \; \sigma_k \\ 0 & \cdots & & 0 \end{bmatrix}$

σ_i – singular value of X matrix

$$\lambda_i = \frac{\sigma_i^2}{n-1} \qquad \lambda_i \text{ – eigenvalue of } C \text{ matrix}$$

$$\sigma_i^2 = \|p_i\|^2 \leftarrow \text{squared length of } p_i$$

R-squared-view of PCA:

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_i}{\sum_{i=1}^{k} \lambda_i} = \frac{\sigma_1^2 + \ldots + \sigma_i^2}{\sum_{i=1}^{k} \sigma_i^2} =$$

$$= \frac{R_i^2(x_1) + R_i^2(x_2) + \ldots + R_i^2(x_{10})}{10}$$

i	λ_i	$\frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i}$	$\frac{\lambda_1 + \ldots + \lambda_i}{\sum_{i=1}^{k} \lambda_i}$
1	..	53%	. . .
2	..	16%	69%
...		..	. . .

53% for p_1

$p_1 = \begin{bmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{n1} \end{bmatrix}$

x_1 on p_1 $\quad R_1^2(x_1)$

x_2 on p_1 $\quad R_1^2(x_2)$

$\vdots$

x_{10} on p_1 $\quad R_1^2(x_{10})$

$$0.53 = \frac{R_1^2(x_1) + \ldots + R_1^2(x_{10})}{10}$$

16% gain

$$0{,}53 = \frac{s\,Corr^2(x_1, p_1) + s\,Corr^2(x_2, p_1) + \dots s\,Corr^2(x_{10}, p_1)}{10}$$

$$0{,}16$$

$$0.69 = \frac{R_2^2(x_1) + R_2^2(x_2) + \dots + R_2^2(x_{10})}{10}$$

R^2 in regression of x_1 on p_1 and p_2

$$\hat{x}_{i1} = \hat{\beta}_1 \cdot p_{i1} + \hat{\beta}_2 \cdot p_{i2}$$

$$0{,}69 = \frac{s\,Corr^2(x_1, \hat{x}_1) + s\,Corr^2(x_2, \hat{x}_2) + \dots + s\,Corr^2(x_{10}, \hat{x}_{10})}{10}$$

where predictions of original variables are made, using two pr. comp-s.

$$0{,}16 = \frac{R_{p_2}^2(x_1) + \dots + R_{p_2}^2(x_{10})}{10}$$

where $R_{p_2}^2(x_1)$ is the R^2 in regression of x_1 on p_2 (only)

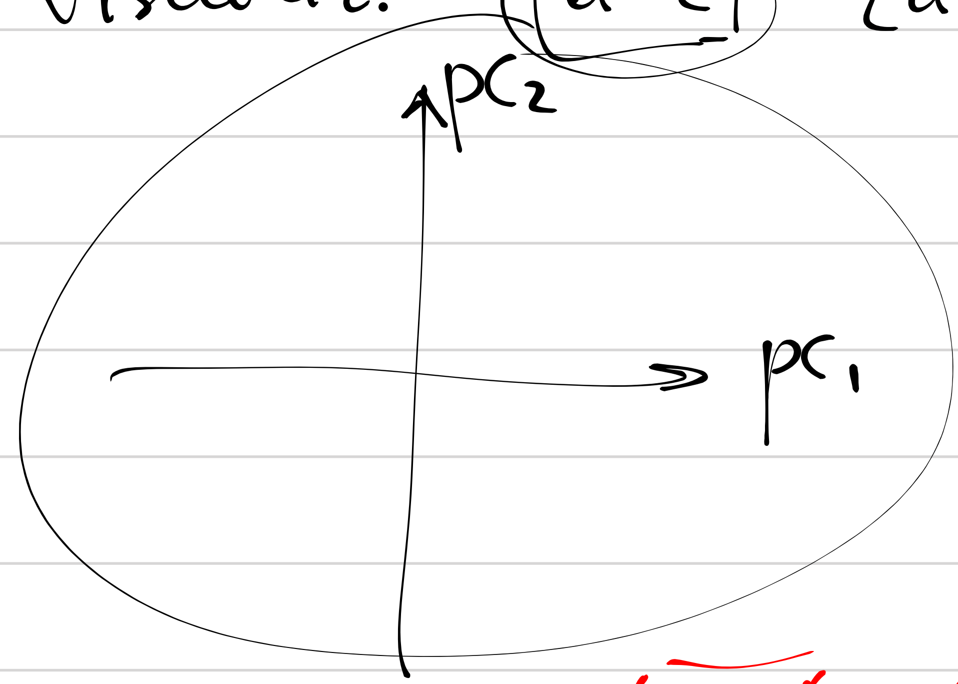
$$\boxed{\hat{x}_{i1} = \hat{\beta}_1 \cdot p_{i2}}$$

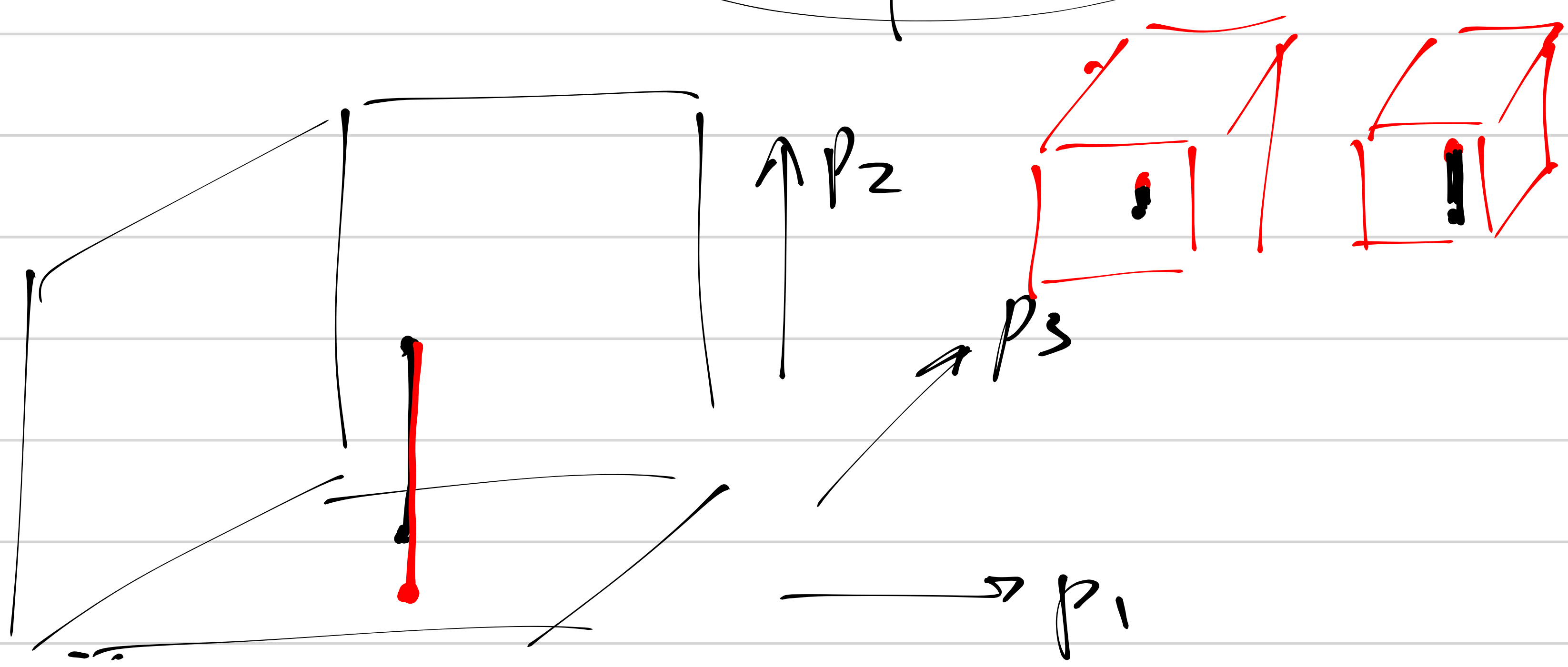
$$0{,}16 = \frac{R^2(x_1 \text{ on } p_2) + \dots + R^2(x_{10} \text{ on } p_2)}{10}$$

$$0{,}69 = \frac{R^2(x_1 \text{ on } p_1, p_2) + \dots + R^2(x_{10} \text{ on } p_1, p_2)}{10}$$

How we determine the number of factors?

$p_1\ p_2 \dots p_d \rightarrow$ visualiz. $[d=2]$ $[d=3]$





Never use 3d plots !! ☺

$$p_1 = X \cdot \underset{\text{weights}}{v_1}$$

n observ.

$$\begin{bmatrix} p_1 \end{bmatrix} = \begin{bmatrix} x_1 \end{bmatrix} \cdot v_{11} + \begin{bmatrix} x_2 \end{bmatrix} \cdot v_{21} + \dots \begin{bmatrix} x_{10} \end{bmatrix} \cdot v_{10,1}$$

$$\begin{bmatrix} p_2 \end{bmatrix} = \begin{bmatrix} x_1 \end{bmatrix} \cdot v_{12} + \begin{bmatrix} x_2 \end{bmatrix} \cdot v_{22} + \dots \begin{bmatrix} x_{10} \end{bmatrix} \cdot v_{10,2}$$

Rotated comp matrix. [later]

Reproduced Correlations table

	$\widehat{App}$	$\widehat{Likel}$	......	$\widehat{Suit}$
App	0.432	0.502		
Like.				
⋮				
Suitability				0.864

$p_1 \quad p_2 \quad p_3$

x_1 on $p_1 \; p_2 \; p_3$

$$\hat{x}_{i1} = \hat{\beta}_1 \cdot p_{i1} + \hat{\beta}_2 \cdot p_{i2} + \hat{\beta}_3 \cdot p_{i3}$$

$0.432 = \text{sCorr}(App, \widehat{App})$

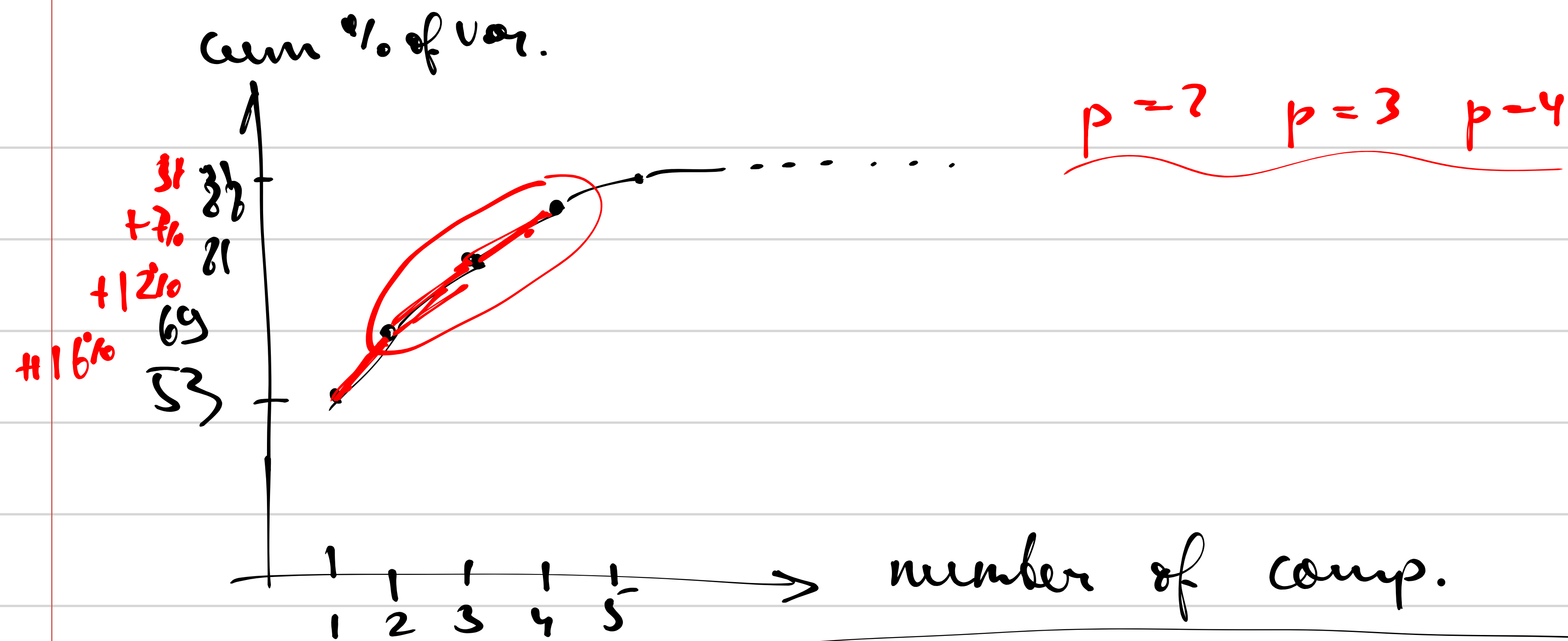
$0.910 = \text{sCorr}(Amb, \widehat{Amb})$

Comp Score Coeff Matrix.

	p_1	p_2	p_3
x_1 Appear	0.04	0.18	0.023
	-0.18	⋮	⋮
⋮			
x_{10} Suitability	-0.050		

$$p_1 = 0.04 \cdot x_1 - 0.18 \cdot x_2 + \ldots - 0.050 \cdot x_{10}$$

$$\|V_1\|^2 = 1 \qquad \Sigma V_{j1}^2 = 1$$



Ex. sample corr-matrix.

$$C = \begin{bmatrix} 1 & 0{,}3 \\ 0{,}3 & 1 \end{bmatrix}$$

two orig: x_1 x_2

a) $p_1 = ?\, x_1 + ?\, x_2$?

0.65 → b) how much variance is explained by p_1? [first weight > 0]

c) for obs N5 $p_{51} = 0{,}2$ what are $\hat{x}_{51}$ and $\hat{x}_{52}$?

Eigenvalues of C?

$$\det \begin{pmatrix} 1-\lambda & 0.3 \\ 0.3 & 1-\lambda \end{pmatrix} = 0$$

$$(1-\lambda)^2 - 0.3^2 = 0$$

$1-\lambda = 0{,}3$ $\qquad$ $1-\lambda = -0{,}3$

$\lambda_2 = 0{,}7$ $\qquad$ $\lambda_1 = 1{,}3$

$$\boxed{\frac{R^2(x_1 \text{ on } p_1) + R^2(x_2 \text{ on } p_1)}{2}} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

$$\frac{R^2(x_1 \text{ on } p_1) + R^2(x_2 \text{ on } p_1)}{2} = \frac{1{,}3}{1{,}3 + 0{,}7} = 0.65$$

$$(C-\lambda_1 I)\cdot v_1 = 0$$

$$\left[\begin{pmatrix}1 & 0.3\\ 0.3 & 1\end{pmatrix}-\begin{pmatrix}1.3 & 0\\ 0 & 1.3\end{pmatrix}\right]\cdot\begin{pmatrix}v_{11}\\ v_{21}\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$$

$$\begin{bmatrix}-0.3 & 0.3\\ 0.3 & -0.3\end{bmatrix}\cdot\begin{bmatrix}v_{11}\\ v_{21}\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$$

$$\|v_1\|=1$$

$$-v_{11}+v_{21}=0 \qquad v_{11}^2+v_{21}^2=1$$

$$v_{11}=\frac{1}{\sqrt{2}} \quad v_{21}=-\frac{1}{\sqrt{2}} \quad \text{or} \quad v_{11}=-\frac{1}{\sqrt{2}} \quad v_{21}=+\frac{1}{\sqrt{2}}$$

$$p_1=\frac{1}{\sqrt{2}}\cdot x_1-\frac{1}{\sqrt{2}}\cdot x_2$$

a) !!

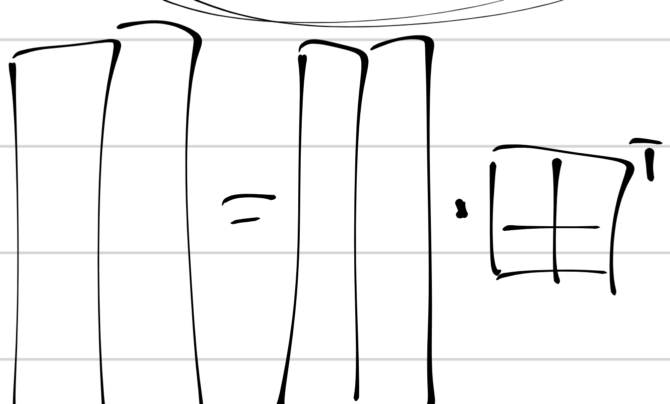
c) $$X=U\cdot\Sigma\cdot V^T$$

[SVD] orth. matr! $U^TU=I$, $V^TV=I$

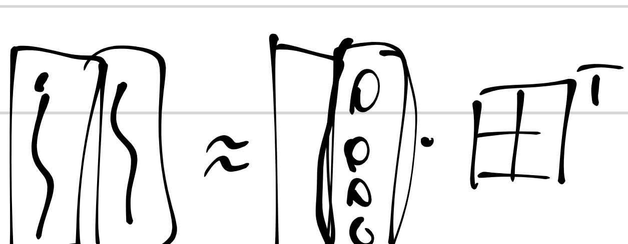
$$X\cdot V=U\cdot\Sigma=P$$

$$X=P\cdot V^T$$

perfect reconstr of orig. varibles from pc-s.



$$\Sigma=\begin{pmatrix}\sigma_1 & 0\\ 0 & \sigma_2\\ 0 & 0\\ 0 & 0\end{pmatrix}$$



approximate reconst.

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \approx \begin{bmatrix} p_1 & \begin{matrix}0\\0\\0\\ \vdots \\0\end{matrix} \end{bmatrix} \cdot \begin{bmatrix} v_1 & v_2 \end{bmatrix}^T$$

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

v_2 ?

$$\lambda_2 = 0{,}7$$

$$\begin{pmatrix} 1-0{,}7 & 0{,}3 \\ 0{,}3 & 1-0{,}7 \end{pmatrix} \cdot \begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$v_{12}^2 + v_{22}^2 = 1$$

$$v_{12} = v_{22}$$

$$v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \rightarrow v_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\begin{bmatrix} X_1 & X_2 \end{bmatrix} \approx \begin{bmatrix} p_1 & \begin{matrix}0\\0\\0\\0\\0\\0\\0\end{matrix} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T$$

$$\begin{bmatrix} p_{51} & 0 \\ 0{,}2 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \hat{x}_{51} & \hat{x}_{52} \\ \frac{0{,}2}{\sqrt{2}} & \frac{-0{,}2}{\sqrt{2}} \end{bmatrix}$$

HA1 HA2 comp

↳ Cluster

5 pr.