Result 2.

Let X : data matrix of dim. $(n \times p)$

Σ : variance-covariance matrix of X; dim. $(p \times p)$

Let $(\lambda_1, \underset{\sim}{e_1}), (\lambda_2, \underset{\sim}{e_2}), \ldots, (\lambda_p, \underset{\sim}{e_p})$ be the eigenvalue-eignvector pairs of Σ

From result 1: if $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p$ then

$\underset{\sim}{e_1}$ = 1 PC loading

$\underset{\sim}{e_2}$ = 2 PC loading,

$\vdots$

$\underset{\sim}{e_i}$ = i PC loading ...

p

P_1	P_2	...	P_p

Result 2: λ_i, eigenvalue associated with the ith eigenvector $\underset{\sim}{e_i}$, is the measure of variance of the ith PC.

$var(P_1) = \lambda_1, \quad var(P_2) = \lambda_2, \ldots, var(P_p) = \lambda_p.$

Result 3

X

x_1	x_2	$\cdots$	x_p

Σ = var-cor (X)

$(\lambda_1, \underset{\sim}{e_1}), \ldots, (\lambda_p, \underset{\sim}{e_p})$ = eigenvalue-eigenvector pairs of Σ.

$\sigma_{11} = \text{var}(x_1)$

$\sigma_{22} = \text{var}(x_2)$

$\vdots$

$\sigma_{pp} = \text{var}(x_p)$

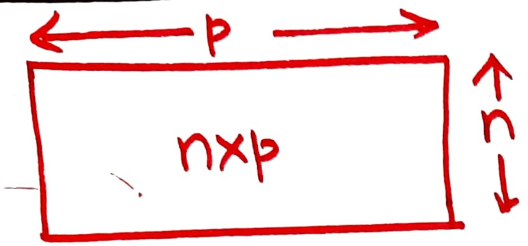
$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}$$

$$= \lambda_1 + \lambda_2 + \cdots + \lambda_p .$$

i.e. $\boxed{\sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i}$

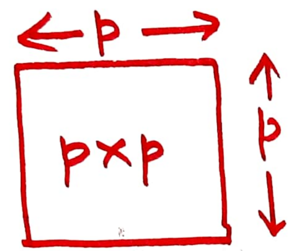
PCA for $p > n$

X =

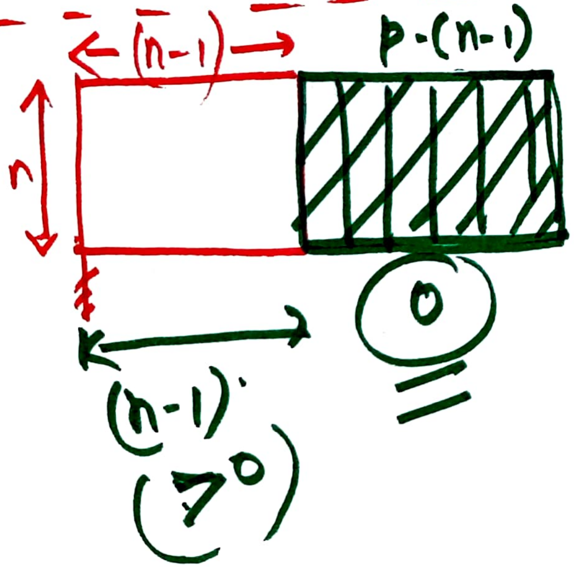


$p >> n$

Σ = var-cor(x) =



P =

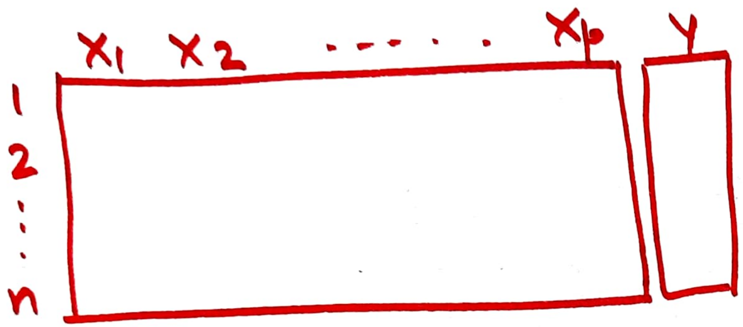


A full-rank var-cov matrix Σ cannot be estimated using a number of observations n smaller than or equal to the no. of ~~p~~ variables p.

In such a case, the rank of Σ is ~~(p-1)~~ $(n-1)$. In these cases, the eigen-decomposition of Σ produces $(n-1)$ real and $p-(n-1)$ null eigenvalues.

Linear Regression

$p > n$



$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_p x_{pi} + \varepsilon_i$$

$i = 1(1)n.$

$(p+1)$

$n = 100$

$p = 500.$

$$y_1 = b_0 + b_1 x_{11} + \dots + b_p x_{p1} + \varepsilon_1 \quad - (1)$$

$$y_2 = b_0 + b_1 x_{12} + \dots + b_p x_{p2} + \varepsilon_2 \quad - (2)$$

$$\vdots$$

$$y_{100} = b_0 + b_1 x_{1\,100} + \dots + b_p x_{p\,100} + \varepsilon_{100} \quad - (100)$$

$$20 = 3x_1 + 5x_2 - 3x_3 \quad - (1)$$

$$35 = 5x_1 + 13x_2 + x_3 \quad - (2)$$

$- (3)$

3 unknowns
2 equation.

501 unknowns
100 equation.

501 equation

Rank(matrix) = # of Linearly independent columns in the matrix

(OR) = # " " " rows in the matrix.

$$A = \begin{array}{ccc} C_1 & C_2 & C_3 \\ \end{array}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$C_1 = p\,C_2 + q\,C_3$

Rank = 2

Rank = 3 (Full-rank matrix)

$\Sigma = \boxed{p \times p}$

$n << p$. Rank $\neq p$. $= (n-1)$

$n > p$ Rank$(\Sigma) = p$

$X = \boxed{n \times p}$, $n > p$

$P = \boxed{n \times p}$

$\Sigma = \boxed{p \times p}$

$\text{Rank}(\Sigma) = p$

p eigenvalues > 0

$X = \boxed{n \times p}$, $n < p$

$P = \boxed{n \times (n-1)}$ $\leftarrow p-(n-1) \rightarrow$

$n-1$ $\quad (p-n-1)$

$\Sigma = \boxed{p \times p}$

$(\lambda_1, \underset{\sim}{e_1}), (\lambda_2, \underset{\sim}{e_2}), \dots, (\lambda_p, \underset{\sim}{e_p})$

$\text{Rank}(\Sigma) \neq p$

$= (n-1)$

First $(n-1)$ eigenvalues > 0

$p-(n-1)$ eigenvalues $= 0$

$p-(n-1)$ columns can be expressed as a linear combination of other cols.