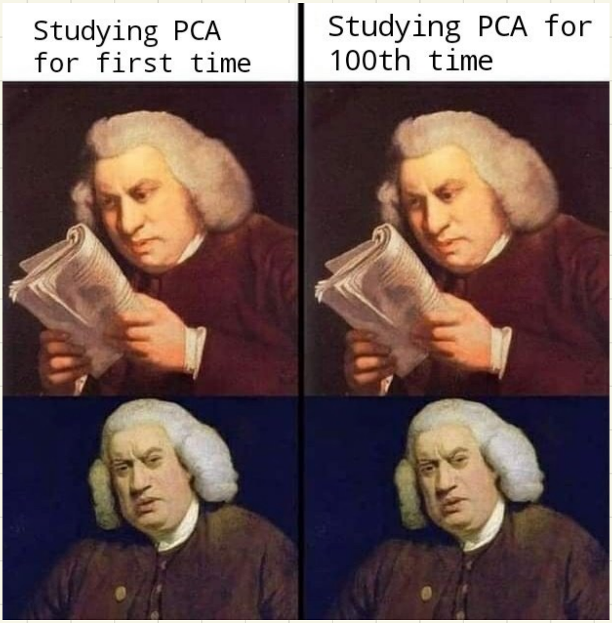
Dimension Reduction

$x_i \in \mathbb{R}^p$ Our goal

$x_i \mapsto z_i \in \mathbb{R}^q$

$q < p$



We do it by PCA

Find q vectors.

$a_1, a_2, \dots a_q \in \mathbb{R}^p$
to span q-dim space

$$\underline{A}_{p \times q} = \left(\begin{pmatrix} a_1 \end{pmatrix} \begin{pmatrix} a_2 \end{pmatrix} \cdots \begin{pmatrix} a_q \end{pmatrix} \right)$$

$$x_i \mapsto A^T x_i \in \mathbb{R}^q$$

$$\|a_j\| = 1 \qquad \langle a_j, a_{j'} \rangle = 0$$

Which q vectors?

optimal $a_1, a_2 \dots a_q$

should have least reconstruction error.

$$\frac{1}{n} \sum \|x_i - AA^T x_i\|^2$$

The problem is equivalent to $\max \mathrm{Var}(a_1^T x_i)$

$s.t.\ \|a_1\|=1$ a_1: first PC

$\max \mathrm{Var}(a_2^T x_i)$

$s.t.\ \|a_2\|=1$ $\langle a_1, a_2\rangle = 0$

a_2: 2nd PC.

Last week:

Solve $a_1, a_2 \dots a_q$

when $\mathrm{Cov}(X)$ diagonal

$\Sigma \to \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_q \end{pmatrix}$ $\lambda_1 > \lambda_2 > \cdots > \lambda_q$

$a_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $a_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $a_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$A = (a_1, a_2 \cdots a_q) = \begin{pmatrix} 1 & 0 & \cdots & \\ 0 & 1 & 0 & \vdots \\ \vdots & \vdots & 1 & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix}_{p \times q}$

$\neq I_q$

Ex: $p=5$, $q=2$

$z_i = A^T x_i$ ← Throw away the small variances.

$= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

What if Σ not diagonal?

<u>Theorem</u> (Real Spectral Thm.)

For any symmetric matrix Σ ($\Sigma^T = \Sigma$), we can find another matrix $\mathbb{P}_{p \times p}$ s.t. $\Sigma = PDP^{-1}$, where D is a $p \times p$ diagonal matrix.

Conclude: Every symmetric matrix can be diagonalized.

$$P^T P = I, \quad PP^T = I. \Rightarrow P^{-1} = P^T$$

Let $p_1, p_2, \dots, p_p \in \mathbb{R}^p$

$$P = (p_1, p_2 \cdots p_p)$$

what does $P^{-1} = P^T$ imply to $p_1, p_2, \dots, p_p$?

$$P^T P = \begin{pmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_p^T \end{pmatrix} (p_1, p_2, p_3 \cdots p_p) \Rightarrow \begin{pmatrix} p_1^T p_1 & & & \\ p_2^T p_1 & p_2^T p_2 & & \\ p_3^T p_1 & & p_3^T p_3 & \\ & & & \ddots & \\ & & & & p_p^T p_p \end{pmatrix}$$

$=1$

$=0$

Summary

$P^{-1} = P^T \quad P = (p_1, p_2, p_3) \Leftrightarrow \|p_j\| = 1$

$\langle p_j, p_{j'} \rangle = 0$ We call such matrix orthonormal

breakfast + lunch = brunch; orthogonal + normal

e.g.

$$\Sigma = \begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix}$$

We can verify:

$$\begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{3}{5} & \frac{-4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}}_{P} \times \underbrace{\begin{pmatrix} 50 & 0 \\ 0 & 25 \end{pmatrix}}_{D} \times \underbrace{\begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{-4}{5} & \frac{3}{5} \end{pmatrix}}_{P^{-1}}$$

column length = 1

$P^T = P^{-1}$

We can apply R.S.T. to solve PCA

non diagonal Σ, $\Sigma = PDP^{-1}$, D = diagonal.

e.g.

$$\max_{a_1 \in \mathbb{R}^p} Var(a_1^T X) \quad \|a_1\| = 1$$

Note: $Var(a_1^T X)$

$= a_1^T \Sigma a_1$

$= a_1^T P D P^{-1} a_1$

$$= \underbrace{(P^T a_1)^T}_{b_1^T} D \underbrace{P^T a_1}_{b_1}$$

Define $b_1 = P^T a_1$.

objective: $\max \ b_1^T D b_1$

constraint

$\|a\| = 1 \iff a^T a = 1$

$$\begin{aligned} a^T a &= a_1^T P P^{-1} a_1 \\ &= (P^T a_1)^T P^T a_1 \\ &= b^T b_1 = 1 \\ &\iff \|b_1\| = 1 \end{aligned}$$

$$\max \ b_1^T D b_1$$
$$\text{s.t. } \|b_1\| = 1$$

$$b_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$(P^{-1} = P^T)$

$$b_1 = P^T a_1 \Rightarrow a_1 = P b_1$$

Quiz:

$$\Sigma = \begin{pmatrix} 34 & 12 \\ 12 & 41 \end{pmatrix}$$

$$\Sigma = PDP^T$$

$$\mathbb{P} = \begin{pmatrix} \frac{3}{5} & \frac{-4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}, \quad D = \begin{pmatrix} 50 & 0 \\ 0 & 25 \end{pmatrix}$$

What is a_1? (first PC)

b_1?

$\Rightarrow \Sigma = PDP^T \leftarrow$ Do it yourself because I went to the washroom.

How to choose q?

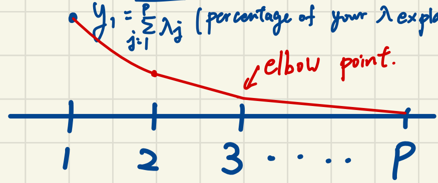
It depends.

$$\Sigma = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix} \quad \lambda_1 > \lambda_2 > \cdots > \lambda_p$$

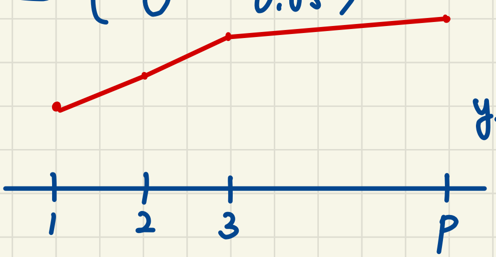
$$\hat{\Sigma} = \left(\begin{array}{cc|c} \lambda_1 & & 0 \\ & \lambda_2 & \\ \hline 0 & & \lambda_3 \end{array}\right)$$

$q = 2$, λ_3 will be your information loss.

$y_1 = \dfrac{\lambda_1}{\sum_{j=1}^{p} \lambda_j}$ (percentage of your λ explained)



$$\Sigma = \begin{pmatrix} 0.8 & & 0 \\ & 0.15 & \\ 0 & & 0.05 \end{pmatrix}$$

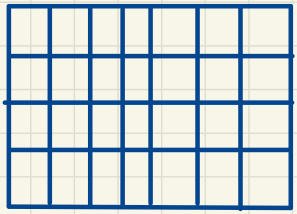


$$y_j = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_j}{\sum_{j=1}^{p} \lambda_j}$$

Application of PCA

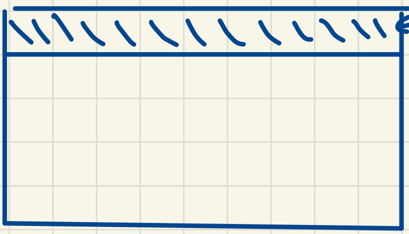
① Data Compression

e.g. image

image  $= 4096 = p$

64×64

Data of homework



← human faces

$\Rightarrow \hat{\Sigma} \Rightarrow a_1, a_2 \ldots a_q$

$q = 200$

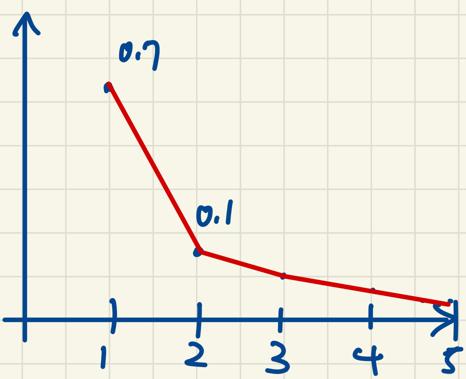
② USE PCA as "summary statistics"

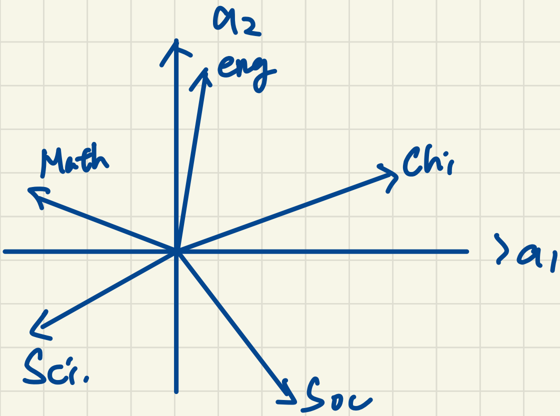
e.g. $\begin{pmatrix} \text{Chi} \\ \text{Eng} \\ \text{Math} \\ \text{Social} \\ \text{Sci} \end{pmatrix}$ $a_1 = ?$ $a_2 = ?$

$a_1 = \begin{pmatrix} 0.47 \\ -0.005 \\ -0.479 \\ 0.468 \\ -0.570 \end{pmatrix}$ $a_2 = \begin{pmatrix} 0.33 \\ 0.79 \\ 0.1 \\ -0.45 \\ -0.19 \end{pmatrix}$

Either you're good at Chi, Soc or Math, Sci

Second means that if you're good at Eng. or not.





③ PCA as feature engineering.

$X_i \to Z_i \in \mathbb{R}^q$

regression clustering

④ Factor analysis.

$$\underset{p\times 1}{\underline{X}} = \underset{p\times q}{A} \times \underset{q\times 1}{\underline{Z}} + \underset{p\times 1}{\underline{\varepsilon}}$$

What we observed — latent variable — error term.

$$\begin{pmatrix} \text{Chi} \\ \text{Eng} \\ \text{Math} \\ \text{Soc.} \\ \text{Sci.} \end{pmatrix} = \underset{p\times q}{\begin{pmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \\ 0.9 & 0.1 \\ 0.3 & 0.7 \\ 0.8 & 0.2 \end{pmatrix}} \begin{pmatrix} \text{ability in Sci} \\ \text{ability in literature} \end{pmatrix} + \varepsilon_i$$

$X = AZ + \varepsilon$

Assum1. $\varepsilon \sim N_p(0, \sigma^2 I)_{p\times p}$

Assum2 $Z \sim N_q(0, |D|)$, $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \; \lambda_q \end{pmatrix}$

Assum3. $Z \perp\!\!\!\perp \varepsilon$

$\Rightarrow X = AZ + \varepsilon$

Generative Model

distribution of data is specified.

Parameters

A, D, σ^2 ← need to be estimated

(assume q known)

e.g. In Psycology, Big five traits (大五人格特質)

我服學沒做完 也要升大五ㄌ!!

A person's personality is built by extrovert. friendliness

Conscientious -ness openess, neuroticism

PCA & factor analysis

need to estimate, A, D, σ^2

$A = (a_1, a_2 \dots a_q)$ PCs

$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 \ddots & \\ 0 & & \lambda_q \end{pmatrix}$

when

$\sigma^2 \to 0$

$\varepsilon \sim N(0, \sigma^2 I)$

$z \sim N(0, \underline{D}) \leftarrow$ Only this version would lead to PCA.

Why factor Model?

① better interpretation.

② statistical inference. on $a_1, a_2 \dots a_q$

Hypothesis testing

$H_0: \dfrac{\lambda_1 + \lambda_2}{\sum_{j=1}^{p} \lambda_j} > 0.8$ (Benefit of Generative Model)