Dimension Reduction:

Set up: Data has p dimensions. $x_i \in \mathbb{R}^p$

Assume: $E\, x_i = 0$

$\frac{1}{n} \sum x_i = 0$

(You can always centered data)

We want to find a function f:

$$f: \mathbb{R}^p \to \mathbb{R}^q, \quad q < p$$

that "encodes" x_i to

$$z_i = f(x_i) \in \mathbb{R}^q$$

($f(\cdot)$ is "encoder")

Generally we also want a decoder function $g(\cdot)$

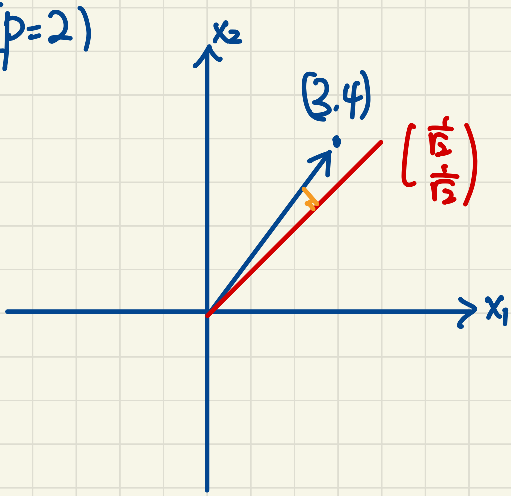
$$g: \mathbb{R}^q \to \mathbb{R}^p$$

We would lose information. But we want the reconstruction minimized.

$$\frac{1}{n} \sum_{i=1}^{n} \| x_i - g(f(x_i)) \|^2$$

$\|\cdot\|$ = norm / length 長度

(p=2)



$a_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

$A = (a_1)$

$f(x) = A^T x,\ x \in \mathbb{R}^2$

$g(z) = Az,\ z \in \mathbb{R}^1$

quiz:

$x = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$. What is

$f(x) = A^T x$

$z = f(x) = ?$

$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{3}{\sqrt{2}} + \frac{4}{\sqrt{2}} = \frac{7}{\sqrt{2}} = z$

$\hat{x} = g(z) = ?$

$g(z) = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \cdot \frac{7}{\sqrt{2}} = \begin{pmatrix} \frac{7}{2} \\ \frac{7}{2} \end{pmatrix} = \hat{x}$

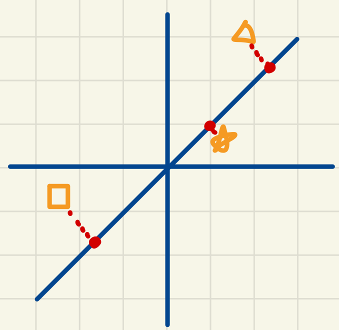
$\|x - \hat{x}\| = ? \Rightarrow \sqrt{\left(3 - \frac{7}{2}\right)^2 + \left(4 - \frac{7}{2}\right)^2}$

$= \sqrt{\frac{1}{4} + \frac{1}{4}}$

$= \frac{1}{\sqrt{2}}$ #

$g(f(x))$ = projection of x on a_1

$a_1^T x$ = inner product of a_1 & x



What if $p>3$ ($q>3$) dimensions >3

We can't eyeball anymore

For $q=1$

PCA find a line to projection onto so that reconstruction error is minimized.

q	Find	Need Vector
$q=1$	a line	$a_1 \in \mathbb{R}^p$
$q=2$	a plane	$a_1, a_2 \in \mathbb{R}^p$
$q=3$	3D space	$a_1, a_2, a_3 \in \mathbb{R}^p$

The vectors $a_1, a_2, \ldots a_q$ these are the "principal Components" (PC)

Rmk: if $x \in \mathbb{R}^p$, we at most have p PCs

We require:

$$\|a_j\| = \sqrt{a_j^T a_j} = 1$$

$$a_j^T a_{j'} = 0 \quad j \neq j' \iff \text{orthogonal}$$

Quiz: Suppose $p=3.\ q=2$

which of the following can be the first PC?

1. $a_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $a_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
2. $a_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$ $a_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$
3. $a_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $a_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$
4. $a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $a_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

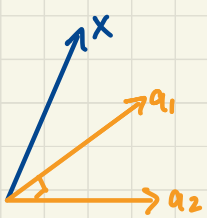
Ans: 4. since $p=3$, a has to be 3×1
and has to be unit vector

How to define projection.

for $q>1$

Suppose $q=2$

$p=3.\ q=2$



$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} a_1^T x \\ a_2^T x \end{pmatrix} = \begin{pmatrix} a_1^T \\ a_2^T \end{pmatrix} x = (a_1\, a_2)^T x$$

$$\hat{x} = z_1 a_1 + z_2 a_2$$

$$= \underbrace{a_1^T x a_1}_{=\text{projection on } a_1} + \underbrace{a_2^T x a_2}_{\text{projection on } a_2} = (a_1 a_2)\,(a_1 a_2)^T x$$

$a_1, a_2, \dots, a_q$

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

$f(x) = A^T x$

$g(z) = Az$

Optimal $a_1, a_2, \dots, a_q$

$$\min \frac{1}{n}\sum_{i=1}^{n} \|x_i - AA^T x_i\|^2$$

s.t. $\|a_j\| = 1 \quad j = 1, 2, \dots, q$

$a_j^T a_{j'} = 0 \quad j \neq j'$

① Find $a_1 \in \mathbb{R}^p$

$$\min_{a_1 \in \mathbb{R}^p} \frac{1}{n}\sum \|x_i - a_1^T x a_1\|^2$$

s.t. $\|a_1\| = 1$

② Given a_1, solve

$\tilde{x} = x_i - a_1^T x a_1$

$$\min_{a_2 \in \mathbb{R}^p} \frac{1}{n}\sum \|\tilde{x}_i - a_2^T \tilde{x} a_2\|^2$$

s.t. $\|a_2\| = 1$

$a_1^T a_2 = 0$

Claim:

minimize reconstruction error

$=$ max variation of $a_1^T x$

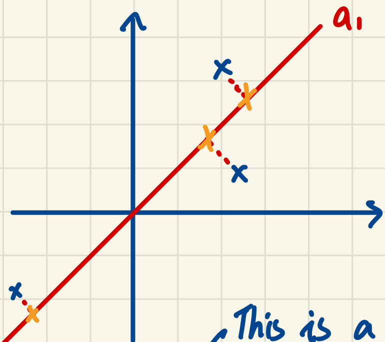
Originally, a_1 solves

$$\min_{a_1} \frac{1}{n}\sum \|x_i - a_1 a_1^T x_i\|^2$$

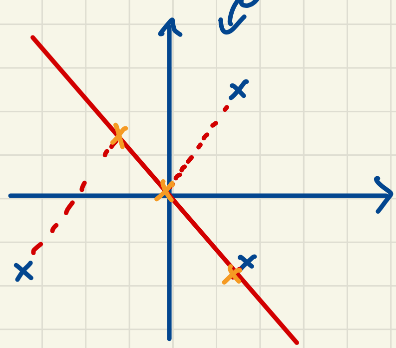
s.t. $\|a_1\| = 1$

$\max_{a_1} \mathrm{Var}(a_1^T x_i)$ when n large.

Intuition:



This is a shitty projection.



$$\frac{1}{n}\sum \|x_i - a_1 a_1^T x_i\|^2$$

$$\Rightarrow \|X\|^2 = \underbrace{\|x_i - a_1 a_1^T x_i\|^2}_{\text{want small}} + \underbrace{\|a_1 a_1^T x\|^2}_{\text{large}}$$

$$\min \frac{1}{n}\sum \|x_i - a_1 a_1^T x\|^2$$

$$\Updownarrow$$

$$\max \frac{1}{n}\sum \|a_1 a_1^T x\|$$

$$\|x_i - a_1 a_1^T x_i\|^2 = \|x\|^2 - \|a_1 a_1^T x\|^2$$

$$\frac{1}{n}\sum \|a_1 a_1^T x\|^2$$

$$= \frac{1}{n}\sum (a_1^T x)^2 \cancel{\|a_1\|}^{=1}$$

$$= \frac{1}{n}\sum_{1}^{n} (a_1^T x)^2$$

$$\xrightarrow{P} E[(a_1^T x)^2] \text{ when } n \to \infty$$

n large:

$$\max E[(a_1^T x)^2]$$

$$E[(a_1^T x)(a_1^T x)]$$

$$= E[(a_1^T x)(x^T a_1)]$$

$$= E[a_1^T x x^T a_1]$$

$$a_1^T (E x x^T) a_1$$

$$= a_1^T \mathrm{Cov}(x)\, a_1$$

$$= \mathrm{Var}(a^T x)$$

Recall: $E x_i = 0$

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ \vdots \\ x_{ip} \end{pmatrix}$$

$$x_i^T = (x_{i1}, x_{i2} \dots x_{ip})$$

$$x_i x_i^T = \begin{pmatrix} x_{i1}^2 & x_{i1}x_{i2} & \cdots \\ x_{i2}x_{i1} & x_{ij}^2 & \vdots \\ \vdots & \cdots & \ddots \end{pmatrix}$$

diagonal $E[x_{ij}^2]$

$= \mathrm{Var}(x_{ij})$

off diagonal:

$x_{ij} \cdots x_{ij'}$

$= \mathrm{Cov}(x_{ij}, x_{ij'})$

$E[x_i x_i^T] = \mathrm{Cov}(x_i)$

$= \Sigma$

Summary

We can find $a_1, a_2 \dots a_q \in \mathbb{R}$ by

$$\max_{a_1} \operatorname{Var}(a_1^T \cdot x)$$
$$\text{s.t. } \|a_1\| = 1$$

$$\max_{a_2} \operatorname{Var}(a_2^T \cdot x)$$
$$\text{s.t. } \|a_2\| = 1$$
$$a_1^T a_2 = 0$$

Quiz:

$p=3 \quad q=2$

Which A corresponds then plane $x=y$

$$\hat{x} = A A^T x$$

Hint:

Let $A = (a_1, a_2)$

What a_1, a_2 span the plane $x=y$

A simpler problem:

Which A project the plane $z=0$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

$z=0$

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad a_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$A = (a_1, a_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

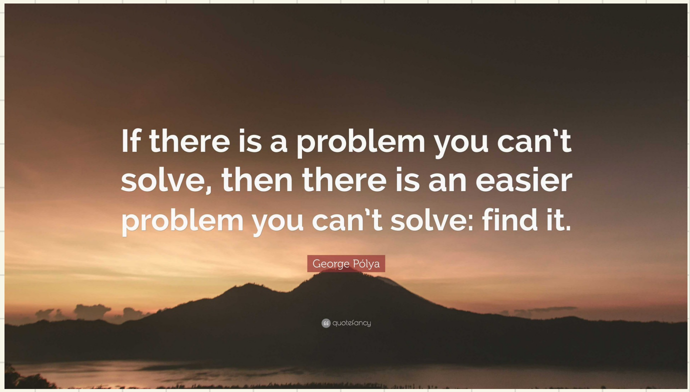
inner product

$$AA^T x = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \underline{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

$x=y$

$a_1=\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{pmatrix}$ $A=(a_1, a_2)$ $a_2=\begin{pmatrix}0\\0\\1\end{pmatrix}$



Original problem:

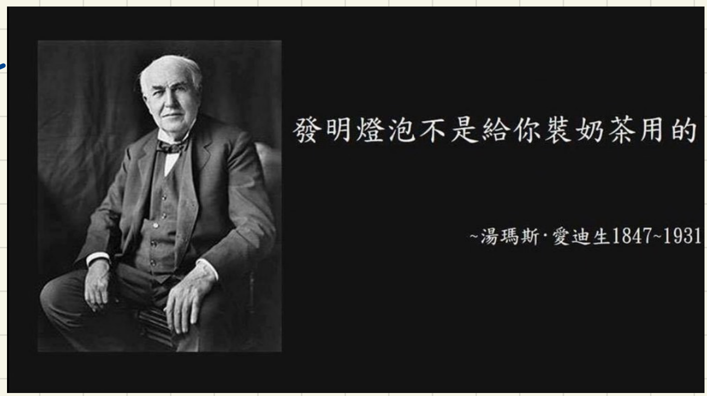
$$\min_{a_1, a_2 \dots a_q \in \mathbb{R}^p} \frac{1}{n}\sum \|x_i - AA^T x_i\|^2$$

s.t. $\|a_j\|=1$

$a_j^T a_{j'}=0 \quad j\neq j'$

$A=(a_1 a_2 \dots a_q)$

Last class, we can solve $a_1, a_2, \dots, a_q$ by (when n large)



$$\max_{a_1} \ Var(a_1^T x)$$

s.t. $\|a_1\|=1$

$$\max_{a_2} \ Var(a_2^T x)$$

$\|a_2\|=1 \quad a_1^T a_2=0$

$$\max_{\substack{a_1 \\ \|a_1\|=1}} \mathrm{Var}(a_1^T x)$$

$$\mathrm{Var}(a_1^T x) = \underset{1\times p}{a_1^T} \underset{p\times p}{\underset{\sim}{\Sigma}} \underset{p\times 1}{\underset{\sim}{a_1}} \in \mathbb{R}$$

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

Without loss of generality.

$$\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_p$$

Objective

$$\mathrm{Var}(a_1^T x) = a_1^T \Sigma a_1$$

$$= (a_{11}\ a_{12} \cdots a_{1p}) \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_p \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1p} \end{pmatrix}$$

$$= \lambda_1 a_{11}^2 + \lambda_2 a_{12}^2 + \cdots + \lambda_p a_{1p}^2$$

$$\|a_1\| = 1$$

$$a_{11}^2 + a_{12}^2 + \cdots + a_{1p}^2 = 1$$

$$\max \ \lambda_1 a_{11}^2 + \lambda_2 a_{12}^2 + \cdots + \lambda_p a_{1p}^2$$

$$\text{s.t. } a_{11}^2 + a_{12}^2 + \cdots + a_{1p}^2 = 1$$

Solution:

$a_{11}^*=1 \quad a_{12}^*=a_{13}^*=\cdots=a_{1p}^*=0$

$a_1=\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} \quad A=(a_1)$

$$AA^T\begin{pmatrix}x_{11}\\x_{12}\\\vdots\\x_{1p}\end{pmatrix}=\begin{pmatrix}x_{11}\\0\\\vdots\\0\end{pmatrix}$$

$\max_{a_2} \quad Var(a_2^T x)$

$\to a_{21}^2\lambda_1+a_{22}^2\lambda_2+\cdots+a_{2p}^2\lambda_p$

s.t. $\|a_2\|=1 \to a_{21}^2+a_{22}^2+\cdots+a_{2p}^2=1$

$a_1^Ta_2=0$

$$(1.0.0\ldots0)\begin{pmatrix}a_{21}\\a_{22}\\\vdots\\0\end{pmatrix}=0 \quad \Rightarrow a_{21}=0$$

$\max_{a_2} \quad a_{22}^2\lambda_2+a_{23}^2\lambda_3+\cdots+a_{2p}^2\lambda_p$

s.t. $a_{22}^2+\cdots+a_{2p}^2=1$

$\Rightarrow a_{22}=1$

$a_{23}=a_{24}=\cdots=a_{2p}=0$

$a_{21}=0$ (by orthogonality)

$$a_2^* = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$a_3^* = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 5 \\ 3 \\ 2 \\ 1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} 5 \\ 3 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Thats the end of the lecture today

中秋節快樂