Today's outline

① Multivariate Normal

② Independence

③ Covariance

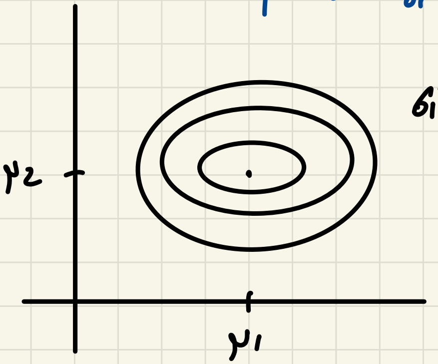
④ PCA

⑤ Data miniworkshop

$(p=2)$

$$\underset{2\times 2}{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, \quad f(x_1, x_2) = \underbrace{f_1(x_1)}_{N(\mu,\sigma_1^2)} \; \underbrace{f_2(x_2)}_{N(\mu,\sigma_2^2)}$$

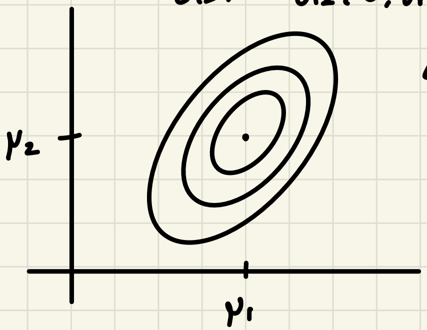
$$f(x_1, x_2) = \underbrace{\frac{1}{2\pi} \cdot \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2}}}_{\text{constant}} \cdot \exp\left[\frac{-1}{2} \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right) \right]$$



$\sigma_1^2 > \sigma_2^2$ in this case

When Σ's not diagonal

$\sigma_{12} \neq 0$ $\sigma_{12} > 0$, or $\sigma_{12} < 0$



← in this case, $\sigma_{12} > 0$ or $\sigma_{12} < 0$

Ans: $\sigma_{12} > 0$

Properties of $N_p(\mu, \Sigma)$

① $X_{p\times 1} \sim N(\mu, \Sigma)$

$A \in \mathbb{R}^{m\times p}$, $b \in \mathbb{R}^m$

$\Rightarrow AX + b \sim N(A\mu + b, A\Sigma A^T)$

$Y = AX + b$

← This property has 3 parts

① $EY = A\mu + b$
② $Cov(Y) = A\Sigma A^T$
] holds for all distribution

③ $Y \sim N(\cdot, \cdot)$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{pmatrix} \sim N(\mu, \Sigma), \text{ if } Cov(X_i, X_j) = 0, \; X_i \underset{\text{independent}}{\perp\!\!\!\perp} X_j$$

Proof: Σ diagonal $\Rightarrow f(x_1, \dots x_p) = f(1) \cdot f(2) \cdots f(p)$

$Cov(X_i, X_j) = 0 \Rightarrow \Sigma$ diagonal

$\Rightarrow f(x_1, x_2, x_3, \dots, x_p) = f(x_1) \cdot f(x_2) \cdots f(x_p)$

$\Rightarrow x_1, x_2, \dots, x_p$ are indep

True or False

① $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \sim N_p(\mu, \Sigma)$

. if X_i, X_j indep, $\forall i \neq j$ (mutually indep)

$\Rightarrow X_1, X_2, \ldots, X_p$ independent.

pf $X_i \perp\!\!\!\perp X_j$

$\Rightarrow Cov(X_i, X_j) = 0$

$\Rightarrow \Sigma$ diagonal

$\Rightarrow X_1, X_2, \ldots, X_p$ indep

Only applies for normal

② X, Y are two normal R.V.

(T or F) $Cov(X,Y) = 0 \Rightarrow X \perp\!\!\!\perp Y$ F

(T or F) $X+Y$ is Normal F

X, Y are normal

$\not\Rightarrow \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2(\mu, \Sigma)$

Covariance is a simple measure for dependency two R.V. X, Y

$$Cov(X,Y) = E[(X-EX)(Y-EY)]$$

Correlation Coefficient

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

p.f.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$$\begin{aligned} Cov(X,Y) &= Cov(X, aX+\cancel{b}) \\ &= Cov(X, aX) \\ &= a\,Cov(X,X) \\ &\Rightarrow a\,Var(X) \end{aligned}$$

$$\begin{aligned} Var(Y) &= Var(aX+b) \\ &= Var(aX) \\ &= a^2 Var(X) \end{aligned}$$

$$Corr(X,Y) = \frac{a\,Var(X)}{\sqrt{Var(X)a^2 Var(X)}} = 1$$

remark:

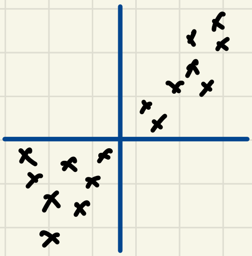
if $Corr(X,Y)=1$

$Y = ax+b$

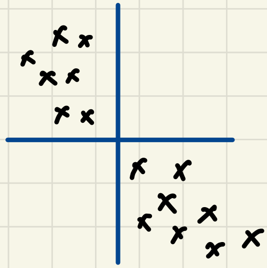
for some $a, b \in \mathbb{R}$

p.f.

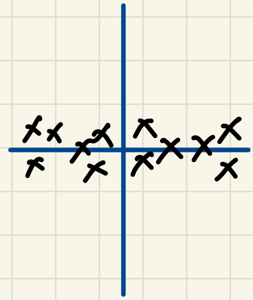
Quiz:



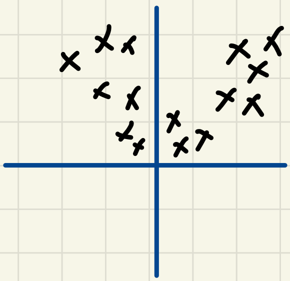
Cov > 0



Cov < 0



Cov = 0



in this case Cov = 0

Implication of last quiz:

Corr( ) only capture "linear independance"

In the last example $Y = X^2$ (highly dependent)
but $Corr(X, Y) = 0$

Correlation is not causility But correlation is useful

Usage of correlation

① Generate potential causal hypothesis
"Pair Plot"

② Risk hedging

two stocks with return

R_1, R_2

$ER_1 = ER_2 > 0$

$Var(R_1) = Var(R_2) > 0$

$Cov(R_1, R_2) < 0$

Consider the portofolio

$\tilde{R} = \frac{1}{2}R_1 + \frac{1}{2}R_2$

$Var(\tilde{R}) = Var(\quad)$

$= Var(\frac{1}{2}R_1) + Cov(\frac{1}{2}R_1, \frac{1}{2}R_2) + Var(\frac{1}{2}R_2)$

$= 2 \cdot \frac{1}{4} Var(R_1) + \underbrace{\frac{1}{4} Cov(R_1, R_2)}_{<0}$

⇒ The variance must lower than individual stock.

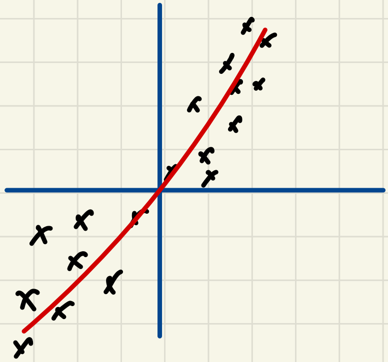
③ Prediction:

See X guess Y

Prediction's about exploiting correlation

(Does not have to be causal)

④ Dimension reduction:



$y = 2x$ Why bother to store (x, y)?

$y \approx 2x$ $(x, y) \approx (x, 2x)$

We can only store one variable without losing information.

For $p=2$, we can eye ball for p large, use algorithm

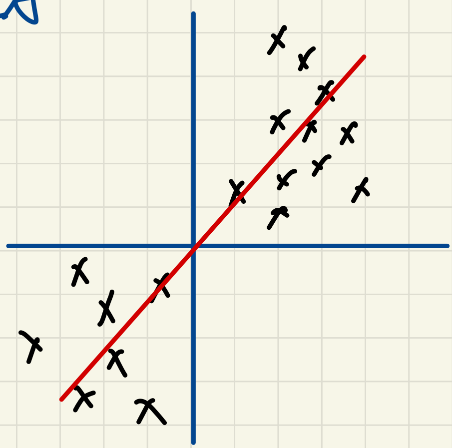
One such algorithm is PCA 主成份分析

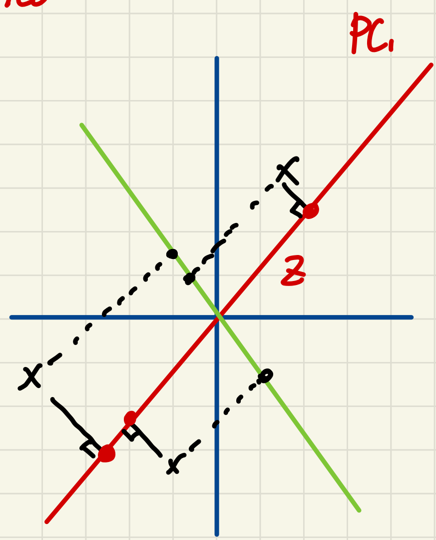
Remind the students who skipped class.

PCA stands for "Principal Component Analysis"

NOT "Porsche Club of America"

PCA

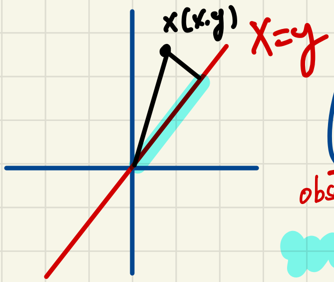




If has 2 dimensions x & y, but after PCA, projected to the point on line z (reduces to 1 dimension)

/ has smaller reconstruction error than \

$\Leftrightarrow$ The projected variable / has larger variance (less information loss)



inner product

$$\underbrace{\begin{pmatrix} x \\ y \end{pmatrix}}_{\text{observation}} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (x, y)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = x+y$$

$$= \frac{(x+y)}{\sqrt{2}}$$

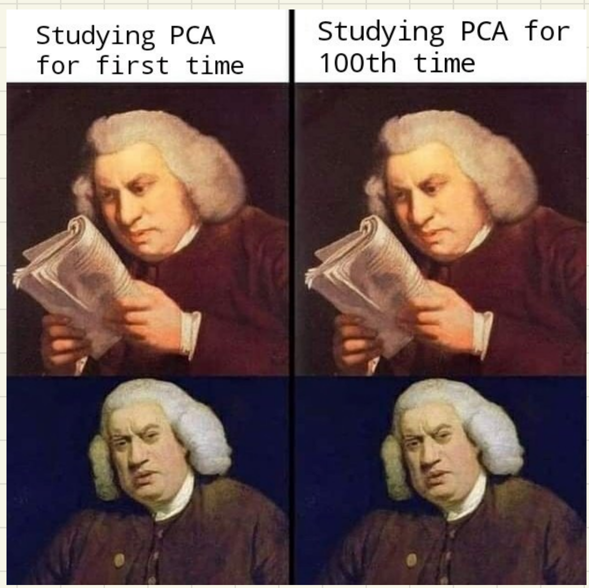
$$\vec{a} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad \vec{b} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

$$\vec{a} \cdot \vec{b} = \vec{a}^T \vec{b}$$

$$= x_1 x_2 + y_1 y_2 \text{ (inner product)}$$

project $\vec{a}$ to $\vec{b}$

$$(\vec{a} \cdot \vec{b}) \frac{1}{\|\vec{b}\|} \cdot \vec{b}$$



$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \text{ random vector}$$

Want to find:

a direction a

so that $Var(a^T X)$ is largest

Def:

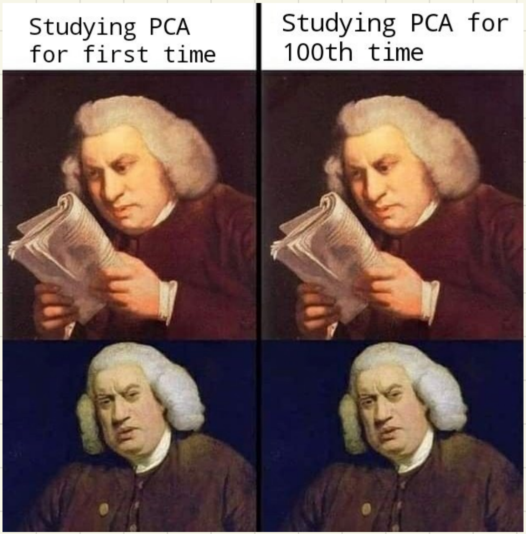
$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}, \quad \underset{p\times 1}{EX} = \underset{p\times 1}{0}, \quad Cov(X) = \Sigma_{p\times p}$$

First Principal Component (P.C.)

a_1. s.t.

$$\max Var(a_1^T X)$$

$$\text{s.t. } \|a_1\| = \sqrt{a_1^T a_1} = 1 \quad (a_1 \text{ is unit vector. } \|a_1\| = 1)$$



The second P.C. is given by $\max_{a_2} Var(a_2^T X)$

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

$a_1^T a_2 = 0$

↳ innerproduct $=0$

$\Leftrightarrow$ orthogonal

j-th P.C ($j \le p$) is given by

$\max Var(a_j^T . X)$

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

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