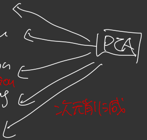
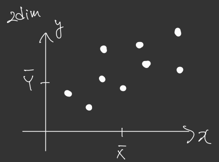
PCA (principal compouend anaysis)

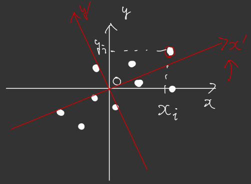
- Axis rotation { How ○
  Criterion ○



- ⓧ clustering
- ⊛ classification
- ⊛ Data compression
  dimensionality reduction
- ＊ Image processing
- ＊ regression

- ○ Axis rotation



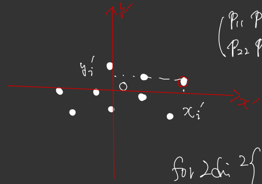


⇓ $(x_i, y_i) \mapsto (x_i', y_i')$

function to rotate axis

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

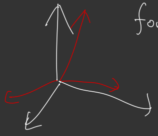
$$\begin{pmatrix} p_{11} & p_{12} \\ p_{22} & p_{22} \end{pmatrix}\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} x_i' \\ y_i' \end{pmatrix}$$



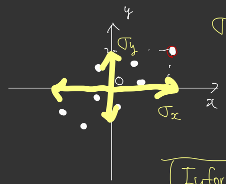
$$P\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} x_i' \\ y_i' \end{pmatrix}$$

for 2dim $2\left\{\underbrace{\begin{pmatrix} p_{11} & \\ & p_{22} \end{pmatrix}}_{2}\right.$ $\boxed{2\times 2}$

for 3dim $3\left\{\underbrace{\begin{pmatrix} p_{11} & & \\ & p_{22} & \\ & & p_{33} \end{pmatrix}}_{3}\right.$



Information of the data

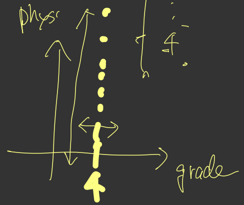


$\sigma_x : \sigma_y = 6:4$

	Math	Physic	. .	grade
①				4
②				4
⋮				⋮
(100)				4

Information

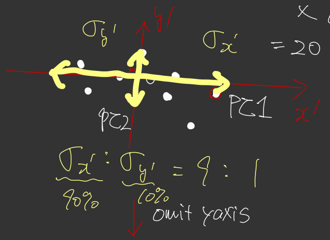
Entropy



2values/each data point

× 10

= 20

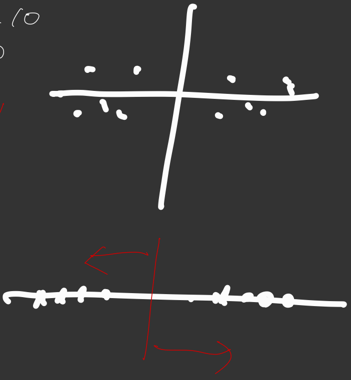


$\sigma_{x'} : \sigma_{y'} = 9:1$

90% 10%

omit y axis

10 values

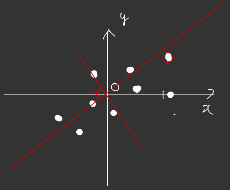


mathematically

goal: to select the axis which has the highest amount of Information

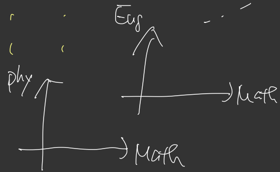
PCA

find axis such that the variance is maximum



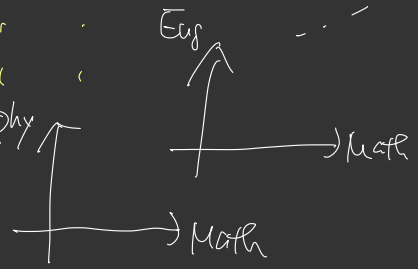
Examples (clustering)

	Math	Phy.	Eng.	Jap.
①	60	20		
②	15	40		
⋮	⋮	⋮		
ⓝ				



Example (clustering)

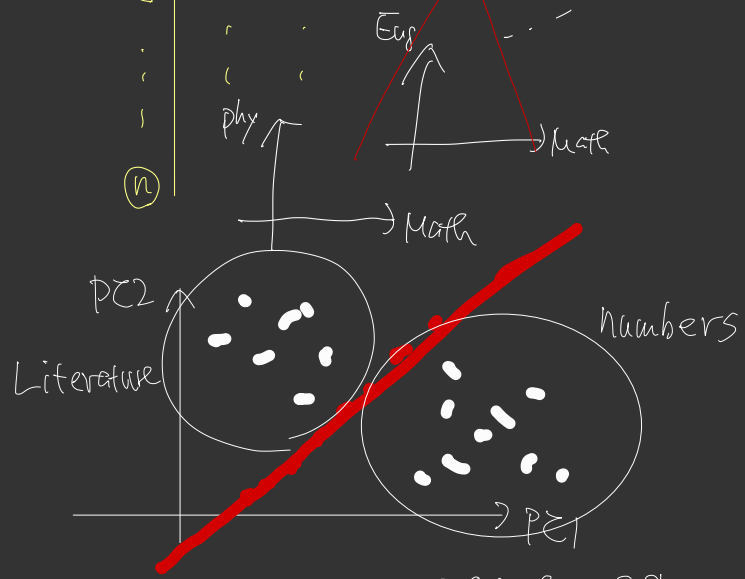
	Math	Phy.	Eng.	Jap.
1	60	20		
2	15	40		
⋮	⋮	⋮		
n				



PCA →

Example (clustering)

	PC1 (55%)	PC2 (30%)	~~PC3 (10%)~~	~~PC4 (5%)~~
1				
2				
⋮	⋮	⋮		
n				



0.4 Math + 0.3 Phy

+ 0.15 Jap + 0.15 Eng

	var1	var2	var20	var10,000
①	PC1	· ·	—	PC10,000
②	15%	10%	· ·	
⋮				
⑩			95%	

PC1 ~ PC20

Data → regression → result

dim = 10,000 10hours

dim = 20 3 minutes

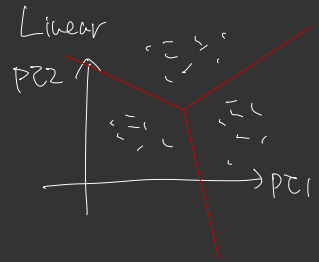
peri . . — — — .

1 ⋮ 569

30

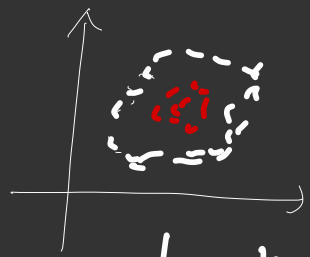
30 dim

Linear



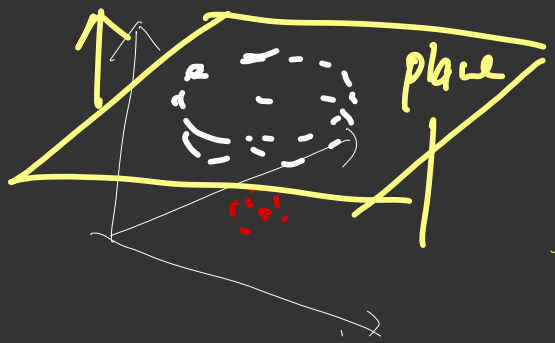
superplane

超平面



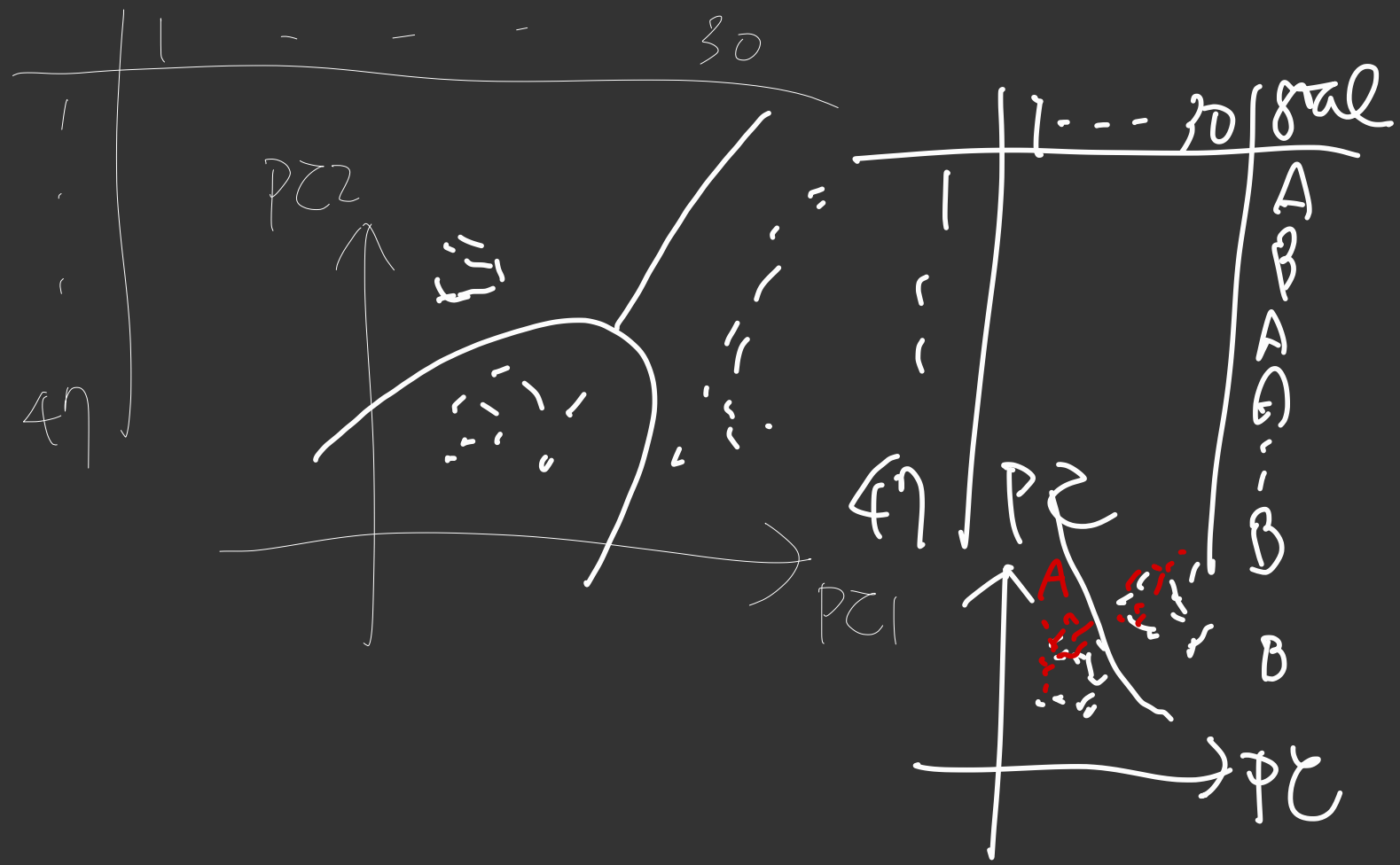
↓ non linear PCA

(e.g. kernel PCA)



$$\begin{pmatrix} 0.3 & 0.2 & 0.4 \\ 0.5 & 0.6 & 0.1 \\ 0.1 & 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$= \begin{pmatrix} 0.3x + 0.2y + 0.4z \\ 0.5x + 0.6y + 0.1z \\ \vdots \end{pmatrix}$$


1 - - - - 30
1
47
PC2
PC1
1 - - - 30
gal
1
47
A
B
A
A
B
B
PC
A
PC