4/3/2024 BT307 - post midsem

- End-sem = 30 marks
  Quiz = 10 marks
  Lab = 5 marks
  Attendance = 5 marks

  Total = 50 marks

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

  $-\infty \le cov(X,Y) \le +\infty$ $= E(XY) - E(X)E(Y)$

  $cov(X,Y) = 0 \Rightarrow$ 

  $Cov(X,Y) > 0 \Rightarrow$ 

  $Cov(X,Y) < 0 \Rightarrow$ 

  $X \perp Y \Rightarrow Cov(X,Y) = 0$

  $\boxed{\text{(proof)}}$

- $$S^2_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$$

* Pearson's correlation: (good measure of linear association) (can be used for non-linear as well)

- $$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}}$$ $(-1 \le \rho_{XY} \le +1)$

- $$r_{xy} = \frac{S^2_{xy}}{S_x S_y}$$ (If $n = 2$, $r_{xy} = 1$, ~~...~~ prove)

- Spearman's correlation (non-linear, ~~...~~ monotonic)

5/3/2024

- t-test for r_{xy},

  $H_0 : \rho_{xy} = 0$

  p-value > cut-off (0.01)

  $\Rightarrow$ accept H_0.

- $\vec{u} = \begin{vmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{vmatrix}$

$$\vec{u}^T \vec{u} = \vec{u} \cdot \vec{u} = \|\vec{u}\|^2$$ (square of norm)

$$s_x^2 = \frac{\|\vec{u}\|^2}{n}$$

$\vec{v} = \begin{vmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{vmatrix}$ $\quad s_y^2 = \frac{\|\vec{v}\|^2}{n}$

$$s_{xy}^2 = \frac{\vec{u} \cdot \vec{v}}{n}$$

$$r_{xy} = \frac{s_{xy}^2}{s_x s_y} = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|}$$

$$= \cos\theta$$

$(-1 \leq \cos\theta \leq 1)$

for absolute values,

$\hookrightarrow |\vec{u} \cdot \vec{v}| \leq \|\vec{u}\| \|\vec{v}\|$ $\quad (0 \leq |\cos\theta| \leq 1)$

they are equal $\Rightarrow$ linearly dependent

$$\vec{x} = c\vec{y}$$

c: non-zero

$$\vec{x} = c\vec{y} + b\vec{z}$$

c, b: non-zero

$$\text{cov matrix} = \begin{vmatrix} s_x^2 & s_{xy}^2 \\ s_{xy}^2 & s_y^2 \end{vmatrix}$$

conformed link

Association does not imply causation

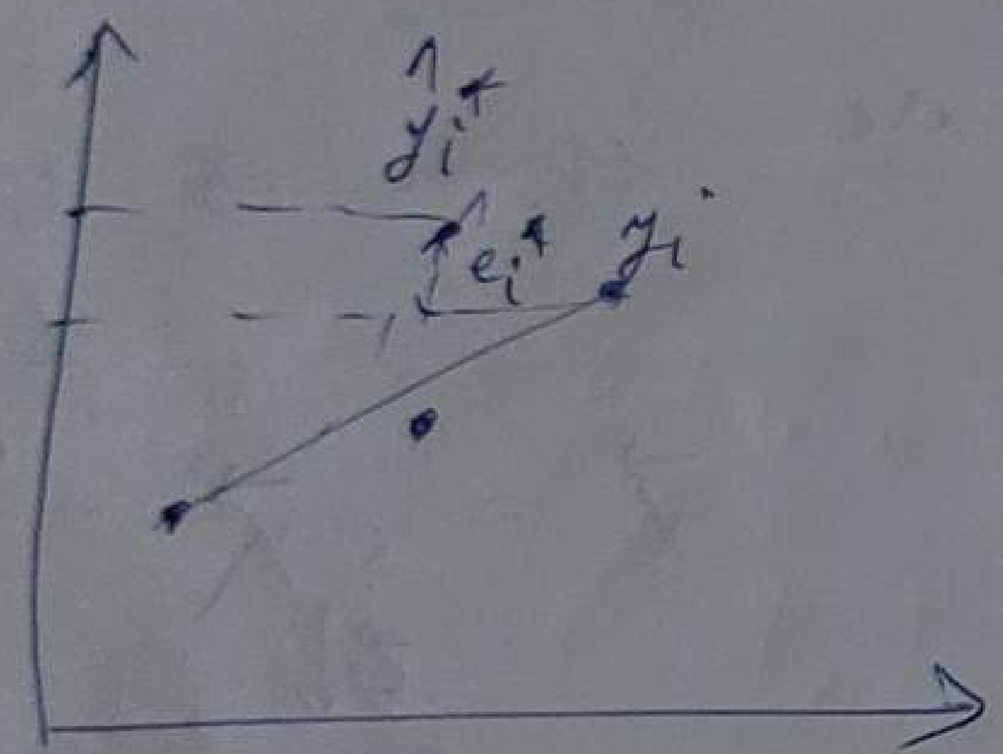
- $y = f(x)$

  $= ax + b$

  (Linear model)

x	y
x_1	y_1
x_2	y_2
x_3	y_3

y (Dependent/outcome)

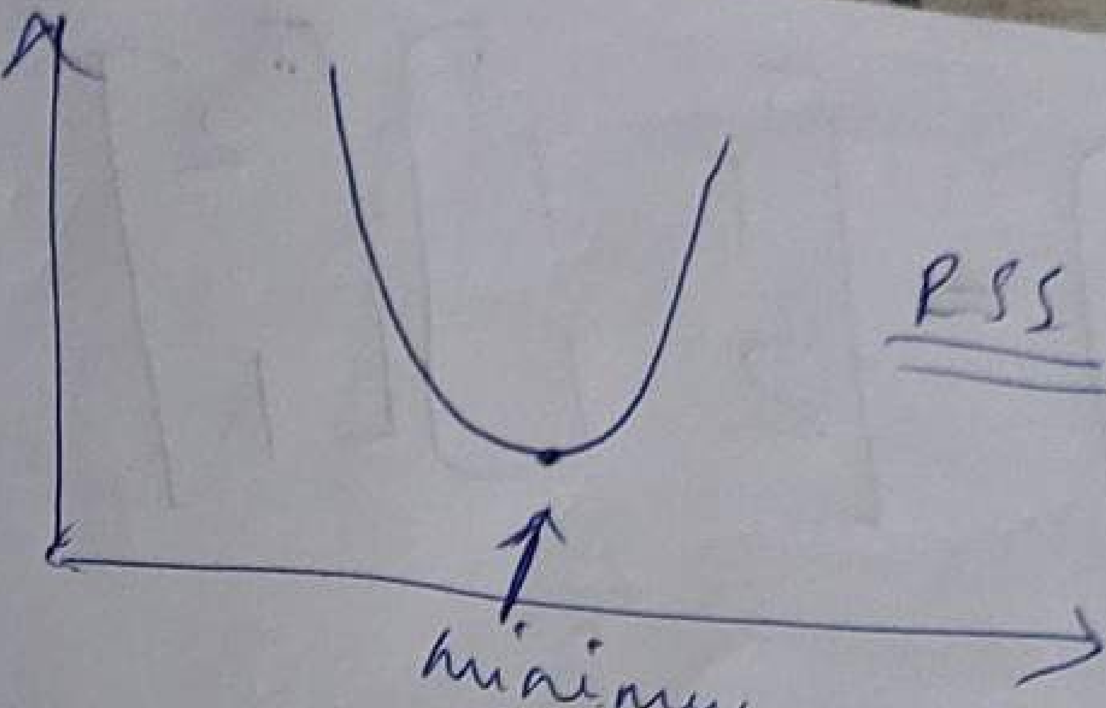


x (independent/predictor)

$$\hat{y}_i = \hat{a}x_i + \hat{b}$$

$$e_i = (y_i - \hat{y}_i)^2$$

Residual sum of squares (RSS) $= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$



$$\hat{a} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{cov(X,Y)}{S_X^2} = \frac{cov(X,Y)}{var(X)}$$

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

- $y = \underset{\hat{a}}{2}x + \underset{\hat{b}}{1}$

X	Y
1	3
2	5

$$\vec{Y} = X\vec{B}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 1 \\ X_2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix}3\\5\end{bmatrix} = \begin{bmatrix}1&1\\2&1\end{bmatrix}\begin{bmatrix}2\\1\end{bmatrix}$$

$$a\begin{vmatrix}x_1\\x_2\end{vmatrix} + b\begin{vmatrix}1\\1\end{vmatrix} = \begin{bmatrix} \\ \end{bmatrix}$$

$\vec{x}_{c_1}$ $\quad$ $\vec{x}_{c_2}$ $\quad$ $\vec{v}_N$

$\downarrow$

same column space

$\vec{y}$ is in the column space of X.

(col_X)

$\downarrow$

(coefficient matrix)

11/3/2024

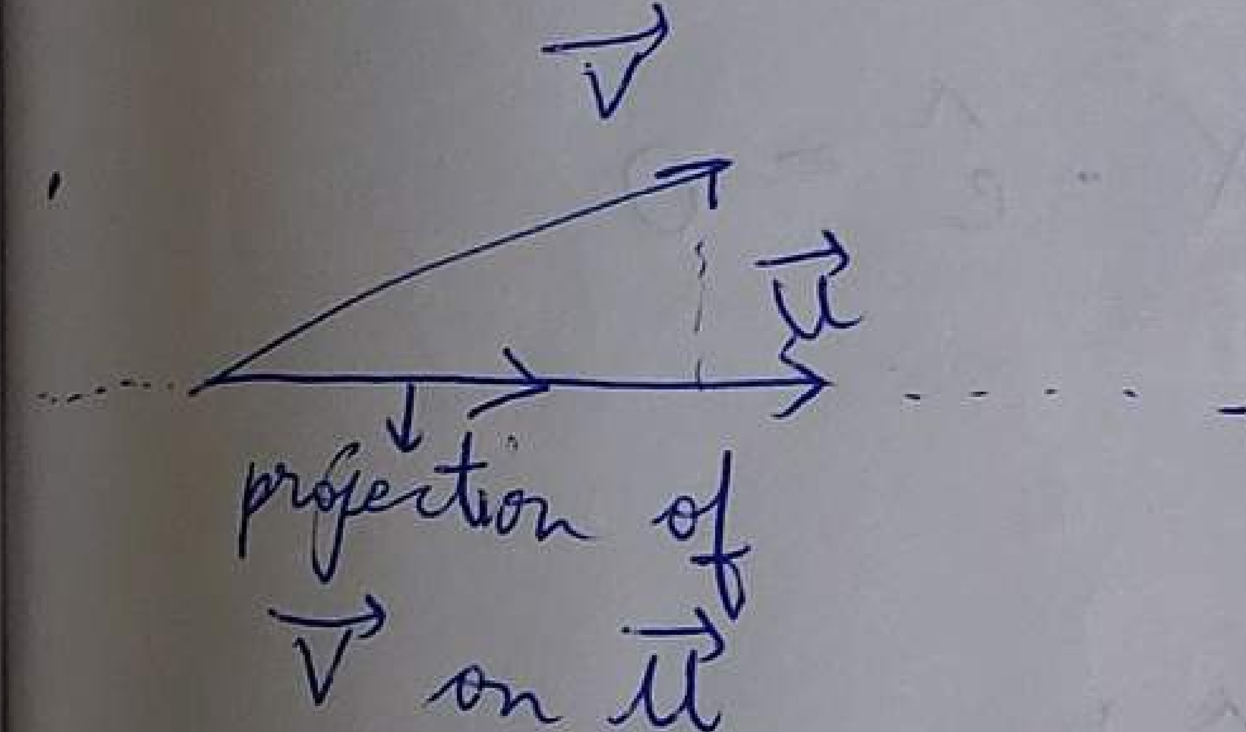
- $\begin{bmatrix}1&8\\2&4\end{bmatrix} = A$

$\vec{c}_1$ $\quad$ $\vec{c}_2$

$\vec{q} = x\vec{c_1} + y\vec{c_2} \Rightarrow q$ is in col_A.

$$\begin{bmatrix}y_1\\ \vdots\\ y_N\end{bmatrix} = \begin{bmatrix}x_1&1\\ \vdots&\vdots\\ x_n&1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}$$

$\vec{Y}$ $\quad$ X $(N>2)$ $\quad$ $\vec{B}$



$$\hat{Y} = X\cdot\vec{Y}$$

$$\Rightarrow \hat{Y} = X^T\vec{Y}$$

$\hat{e}$ - error $= \vec{Y} - \hat{Y}$

$\hat{e} \perp \vec{u} \Rightarrow$ minimum error.

$\hat{e} \perp X \Rightarrow X \cdot \hat{e} = 0$

$X^T \hat{e} = 0$

$X^T(\vec{Y} - \hat{Y}) = 0$

$X^T(\vec{Y} - X\vec{B}) = 0$

$\Rightarrow X^T X \vec{B} = X^T \vec{Y}$

$\Rightarrow \vec{B} = (X^T X)^{-1} X^T \vec{Y}$

$(X^T X)^{-1} X^T$ is hat matrix (pseudo inverse of X)

$$X = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$$

$\hat{e} \cdot \vec{1} = 0 \Rightarrow (Y - \hat{Y})^T \cdot 1 = 0$

$\Rightarrow \frac{\vec{Y}^T \vec{1}}{N} = \frac{\hat{Y}^T \vec{1}}{N}$

$\bar{Y} = \bar{\hat{Y}}$

$\Sigma(y_i - \hat{y}_i) = 0$

$\Rightarrow$ sum of all errors $= 0$ (that's why we take square in RSS).

15/3/2024

- Alternate model: y does not depend on x. (To compare model with a model we know is bad).

$$Y = c = \bar{y}$$

Error:

$Y = ax + b$

$RSS = \sum (y_i - \hat{y}_i)^2$

$Y = c = \bar{y}$

TSS (total) $= \sum (y_i - \bar{y})^2$

R^2 (Coefficient of determination) $= \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$

$TSS = RSS + ESS$ (only true when there is an intercept)

$$ESS = \sum (\hat{y}_i - \bar{y})^2$$

$\Rightarrow R^2 = \frac{ESS}{TSS}$ → Explained S.S.

$$= \frac{\sum (\hat{y}_i - \bar{y})^2 / n}{\sum (y_i - \bar{y})^2 / n}$$

$$= \frac{Var(\text{predicted } y)}{Var(y)}$$

$$0 \le R^2 \le 1$$

(but it can be -ive (if $RSS > TSS$))

$\bar{x}$ sample statistic

μ_x population parameter

$y = \boxed{a} x + \boxed{b}$ (statistic)

$\downarrow \quad \downarrow$

$y = \alpha x + \beta$ (parameter)

$y_i = \alpha x_i + \beta + \varepsilon_i$

$\varepsilon_i \sim N(0, \sigma^2)$

t-test for regression coefficient :

$\left.\begin{array}{l} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{array}\right\}$ (at population level)

$$t = \frac{\text{sample stat.} - \text{pop. parameter}}{\text{S.E. of sample stat.}}$$

H_0 is true,

$$t = \frac{a - 0}{\text{S.E. of } (a)}$$

division by $(n-2) \longrightarrow$ df of t

$P(t \mid H_0 = T) >$ cut-off

flippdle ⇒ 2-sided t-test

18/3/2024

- $\hat{e} \cdot \hat{Y} = 0$

$\Rightarrow (Y-\hat{Y})^T \hat{Y} = 0$

$\Rightarrow Y^T \hat{Y} - \hat{Y}^T \hat{Y}$

$\hat{e} \cdot \vec{1} = 0$

$\hat{e} \cdot (\alpha \vec{1}) = 0$

$(\hat{e} \cdot \bar{y} 1) = 0$

$\hat{e} \cdot \bar{Y} = 0$

$$\downarrow$$

$$\begin{bmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix}$$

$\Rightarrow (Y-\hat{Y})^T \bar{Y} = 0$

$\Rightarrow Y^T \bar{Y} - \hat{Y}^T \bar{Y}$ (only when there are intercept in the model)

$TSS = \sum (y_i - \bar{y})^2 = (Y-\bar{Y})^T (Y-\bar{Y})$

$= Y^T Y - 2\bar{Y}^T Y + \bar{Y}^T \bar{Y}$

$RSS + ESS = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$

$= (Y-\hat{Y})^T (Y-\hat{Y}) + (\hat{Y}-\bar{Y})^T (\hat{Y}-\bar{Y})$ so

$= Y^T Y \underbrace{- 2\hat{Y}^T Y + \hat{Y}^T \hat{Y}}_{=0} + \hat{Y}^T \hat{Y} - 2\hat{Y}^T \bar{Y} + \bar{Y}^T \bar{Y}$

$= y^T y - 2\hat{y}^T \bar{y} + \bar{y}^T \bar{y}$

$= TSS$

(multiple linear)

* multilinear regression

X_1	X_2	Y
x_{11}	x_{21}	y_1
⋮	⋮	⋮
x_{1n}	x_{2n}	y_n

(multi-variate $\Rightarrow \hat{y}$ more than one)

$y = a_1 x_1 + a_2 x_2 + b$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & x_{21} & 1 \\ \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} a_1 \\ a_2 \\ b \end{bmatrix}$$

$B = (X^T X)^{-1} X^T Y$

$\hat{y} = XB$

$$t = \frac{\dots}{SE}$$

$\uparrow$

$(X^T X)^{-1}$

$\longrightarrow (X^T X)^{-1} \uparrow\uparrow \rightarrow \text{var} \uparrow\uparrow$

${}^{k+1}C_2$

$(x_1, x_2) \longrightarrow e_1$

$(x_1, 1) \longrightarrow e_2$

$(x_2, 1) \longrightarrow e_3$

(Pearson coefficients)

~~Auxiliary~~

Auxiliary regression :

- R_j^2 (close to 1 $\Rightarrow$ X_j has linear dependency) (> 0.9)

Variance inflation factor

$$= \frac{1}{1 - R_j^2} > 10$$

- multilinearity problem / multicolinearity problem

2/8/2024

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad B = \begin{bmatrix} a_1 \\ a_2 \\ b \end{bmatrix}$$

$$X = \begin{bmatrix} X_{11} & X_{21} & 1 \\ \vdots & \vdots & \vdots \\ X_{1n} & X_{2n} & 1 \end{bmatrix}$$

$$B = (X^T X)^{-1} X^T Y$$

Adjusted $R^2 = 1 - \dfrac{\frac{RSS}{(n-k-1)}}{\frac{TSS}{(n-1)}}$

d.o.f. of RSS = $(n - k - 1)$

* F-test

don't do repeated t-test on the same data

- $\dfrac{RSS_S - RSS_L}{RSS_L}$

small model $\Rightarrow RSS_S$
large model $\Rightarrow RSS_L$

$y = \underbrace{a_1}_{\alpha_1} x_1 + \underbrace{a_2}_{\alpha_2} x_2 + \underbrace{b}_{\beta}$

H_0 : all $\alpha = 0$, $\Rightarrow y = \bar{y}$

H_1 : at least one $\alpha \neq 0$.

$RSS_{H_0} = \Sigma (y_i - \bar{y})^2$

$RSS_L = \Sigma (y_i - \hat{y}_i)^2$

$$\Rightarrow \frac{\Sigma(y - \bar{y})^2 - \Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \hat{y}_i)^2}$$

$$= \frac{TSS_L - RSS_L}{RSS_L}$$

$$= \frac{ESS_L}{RSS_L} \Rightarrow \frac{\frac{ESS_L}{k}}{\frac{RSS_L}{n-k-1}}$$

(like variances with Bessel correction)

$=$ F-value $\rightarrow$ F-distribution $(k, n-k-1)$

p-value small $\Rightarrow$ reject H_0
($\Rightarrow$ F-value is large)
(small model is bad model)

26/3/2024

* Non-linear ~~re~~ regression
(least square)

- for eg: double exponential
(cannot be made linear)
(linearization)

- $\min \left[\Sigma (y_i - \hat{y}_i)^2 \right]$

$y_i = e^{-kt_i}$

* Polynomial regression

- $y = a_1 x + a_2 x^2 + a_3 x^3 + b$

$y = a_1 x + a_2 w + a_3 z + b$

x	y	w x^2	z x^3

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_1 & x_1^2 & x_1^3 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & x_n^2 & x_n^3 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ b \end{bmatrix} = (X^T X)^{-1} X^T Y$$

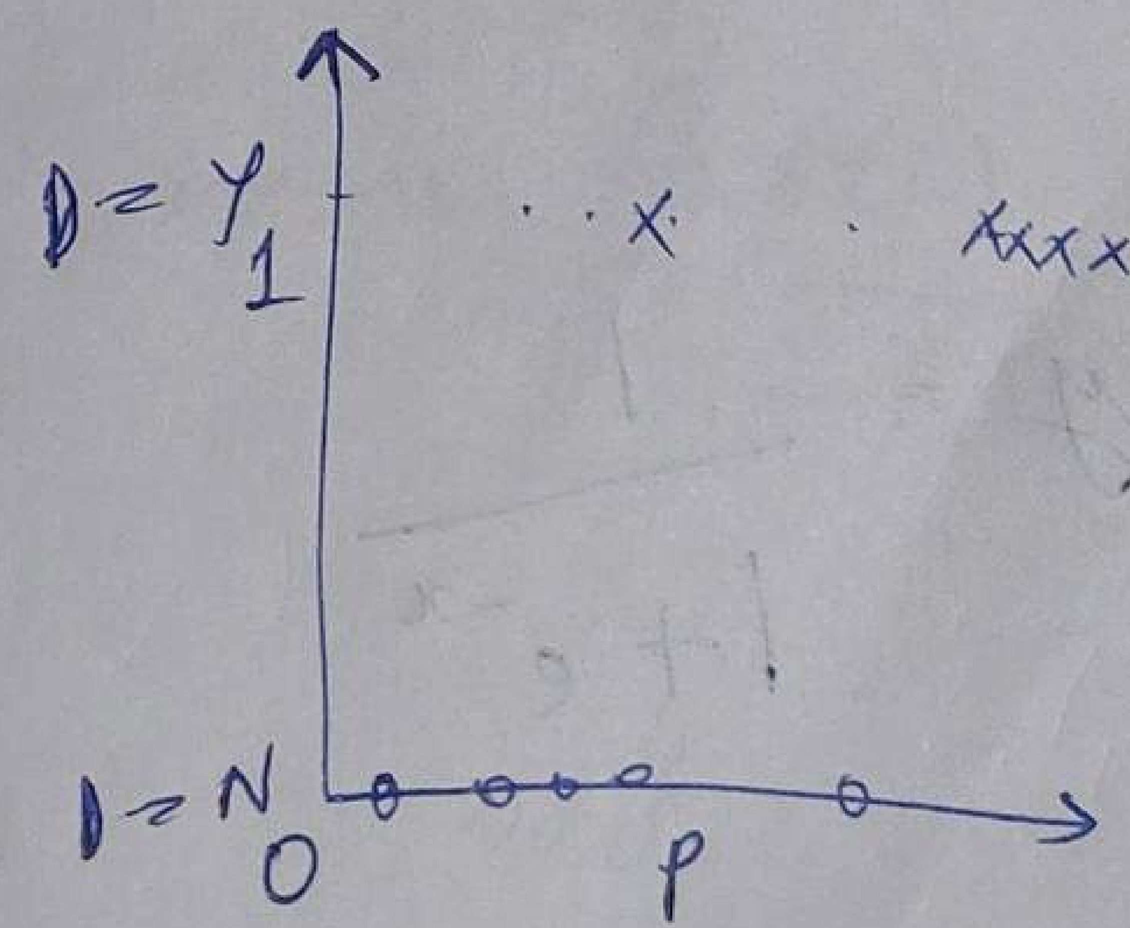
Convert data to matrices (question in exam)
(Y, X & B)

- supervised and unsupervised learning

27/3/2024

-

P	D
	Y
	Y
	⋮
	N



supervised learning with discrete and continuous variables (in linear regression, both are continuous).

$$D = f(P)$$

$$D = a.p + b$$

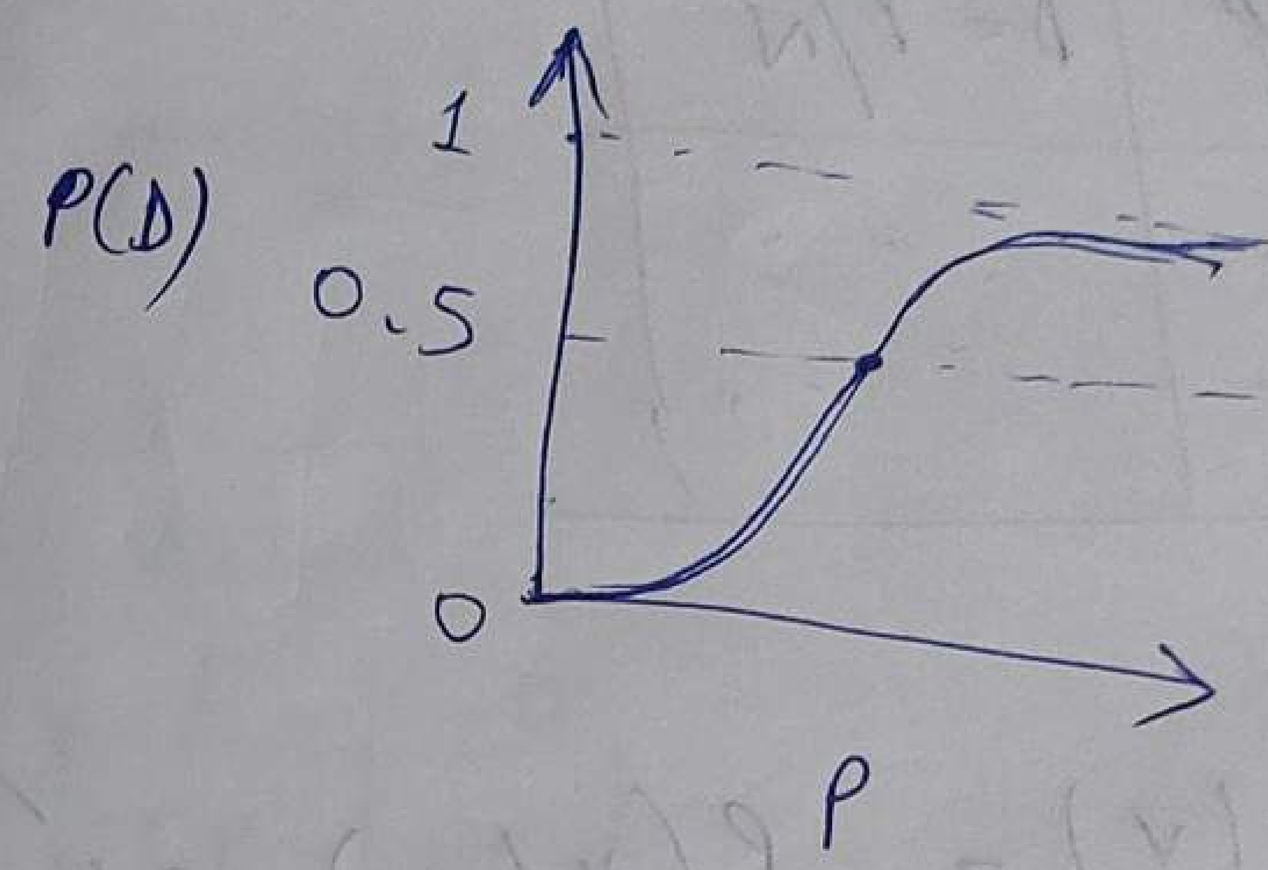
$$p(D) = ap + b$$

step function $\xrightarrow{\text{smoother (differentiable)}}$ sigmoid function

$$y = \frac{1}{1+e^{-x}}$$

$$y = \frac{1}{1+e^{-(a+bx)}}$$

$a<0$, $b>0$ (for this model)



Logistic classifier

Logistic regression

$$P(D) = \frac{1}{1+e^{-(a+bx)}}$$

$$a+bx = \ln\left(\frac{P(D)}{1-P(D)}\right)$$

"log of odd"

odd

$\theta = \{a, b\}$ (set of parameters)

d = data

P ~~θ~~	$D = Y/N$
...	...

$$P(X|Y) = \frac{P(Y/X)\,P(X)}{P(Y)}$$

Prior prob. of X

Posterior prob. of X

$$P(\theta = Y / d = Y) = \frac{P(d = Y / \theta = Y)\,P(\theta = Y)}{P(d = Y)}$$

$$P(\theta | d) = \frac{P(d|\theta) \cdot P(\theta)}{P(d)}$$

1/4/2024

1. Test data :

Contingency table

Confusion matrix

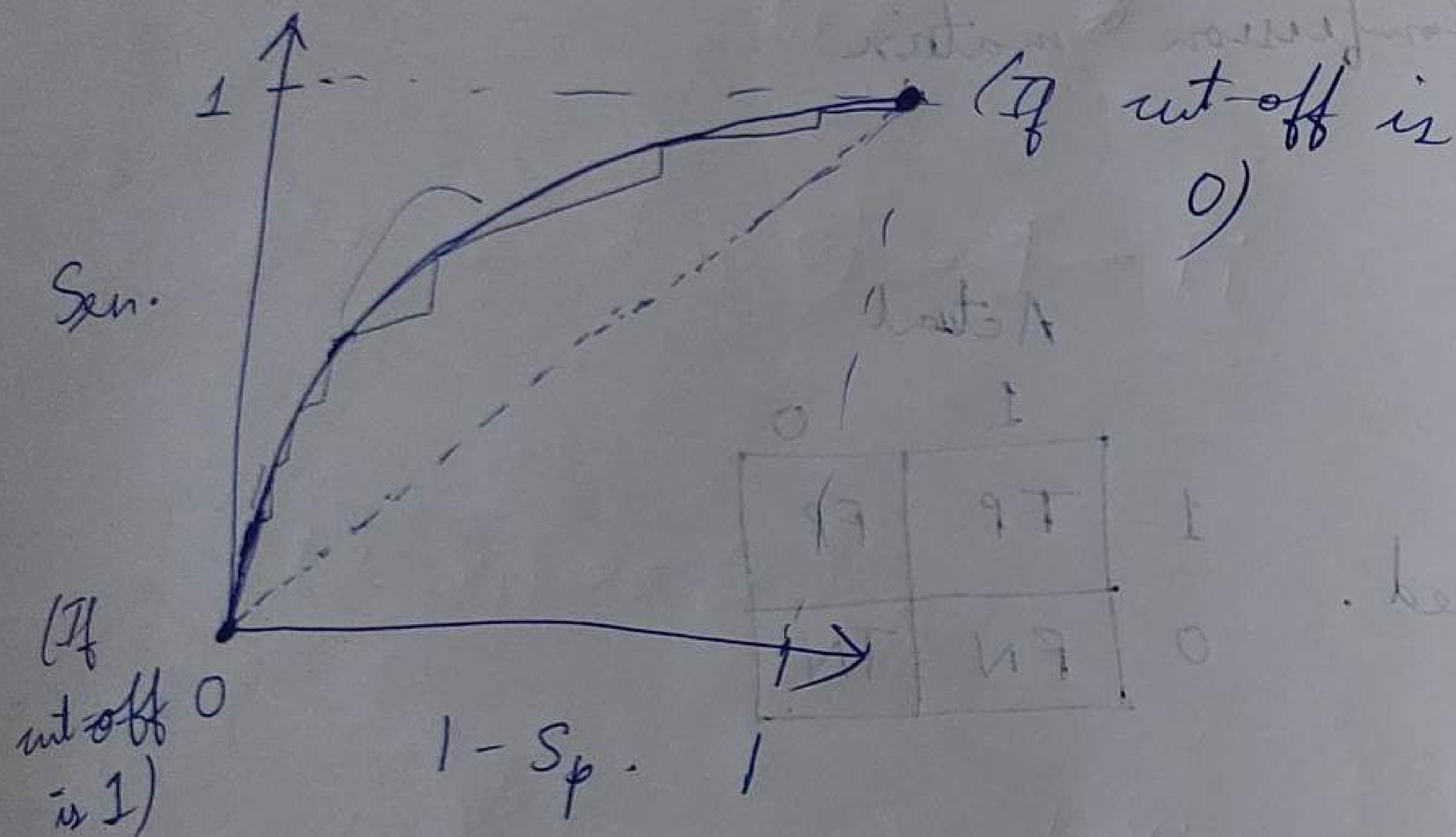
		Actual 1	Actual 0
Pred.	1	TP	FP
	0	FN	TN

- Sensitivity = $\dfrac{\text{true positive}}{\text{total positive}}$

$$= \frac{TP}{TP + FN}$$

- Specificity = $\dfrac{TN}{TN + FP}$

• ROC Curve (Receiver Operator Characteristic Curve)



• $y = \dfrac{1}{1+e^{-(a+b_1x_1+\dots+b_nx_n)}}$

5/4/2024

Maximum Likelihood

Likelihood of θ ($L(\theta)$) — Prior

• $P(\theta|D) = \dfrac{P(D|\theta)\,P(\theta)}{P(D)}$

Posterior prob. of θ

$P(D) \longrightarrow$ Marginal of D

$P(D) = P(D|\theta_1)\cdot P(\theta_1) + P(D|\theta_2)\cdot P(\theta_2) + \dots$

maximize $P(D|\theta)$ to maximize $P(\theta|D)$

Likelihood because it doesn't sum to 1.

eg: $D \Rightarrow n = 10,\ h = 4$

$\theta = \{p_H\}$

$P(D|\theta) = {}^{n}C_{h}\, p_H^{\,h}(1-p_H)^{n-h}$

$= {}^{10}C_4\,(x)^4(1-x)^6$

• $P(Y_i \text{ in class } 1) = P(X_i) = P_i$

$$= \frac{1}{1+e^{-(a+bx)}}$$

$P(Y_i \text{ in class } 0) = 1 - P_i$

$P(X = x_1) = \sigma(x_1)$

$$\prod_{i=1}^{n} \sigma(X_i)^{Y_i} \left[1-\sigma(X_i)\right]^{1-Y_i}$$

$= P(D|\theta)$

$= P(D|a,b)$

$Ln[L(\theta)]$

max of $Ln[L(\theta)]$ will give max $L(\theta)$

Algorithm : Gradient Descent or Ascent

12/4/2024

• Classification (supervised learning)

* Clustering (unsupervised learning)

• Feature space

distance :

$\vec{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$\vec{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

1. Euclidean Distance

$$d_{xy} = \left[\sum_{i=1}^{n} (X_i - Y_i)^2\right]^{1/2}$$

When n is very large, Euclidean distance may be lesser than what is true.

2. Manhattan distance

$$d_{xy} = \sum_{i=1}^{n} |X_i - Y_i|$$

3. Minkowski distance:

$$d_{xy} = \left|(x_i - x_i)^p\right|^{1/p}$$

$$p \geq 1$$

- Euclidean dist.

$$d_{xy}^2 = \sum_{i=1}^{n} (x_i - y_i)^2$$

$$= (\vec{x}-\vec{y})^T(\vec{x}-\vec{y})$$

$$= (\vec{x}-\vec{y})\, S^{-1}\, (\vec{x}-\vec{y})$$

$$S = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

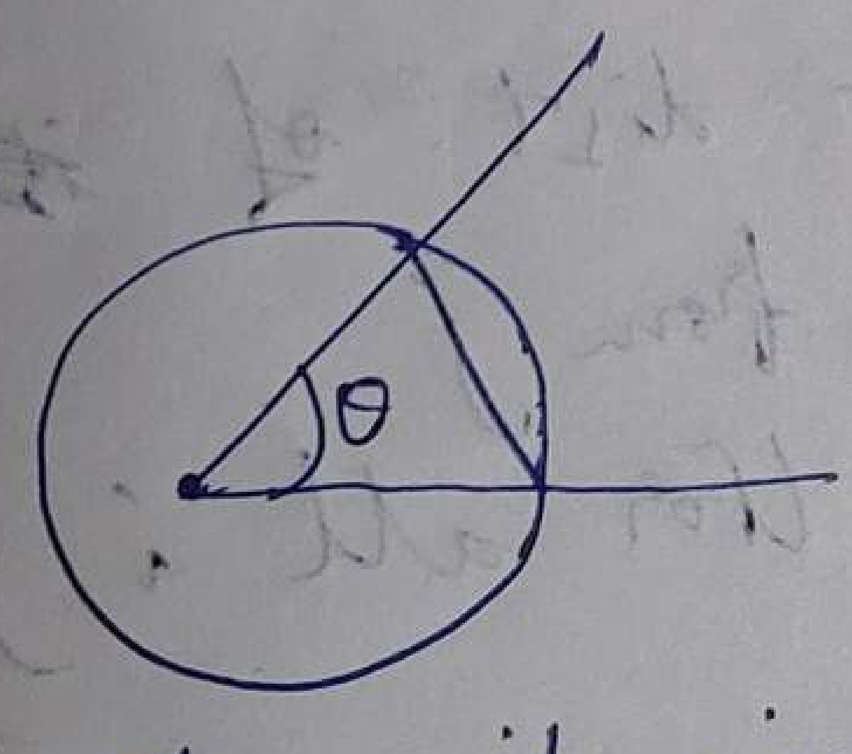
$$S = \begin{bmatrix} v_1 & c \\ c & v_2\end{bmatrix}$$

$$S^{-1} = \frac{1}{\Delta}\begin{vmatrix} v_2 & -c\\ -c & v_1\end{vmatrix}$$

4. Mahalohobis distance:

$$d_{xy} = \left[(\vec{x}-\vec{y})^T S^{-1}(\vec{x}-\vec{y})\right]^{1/2}$$

5. Chord distance



unit circle

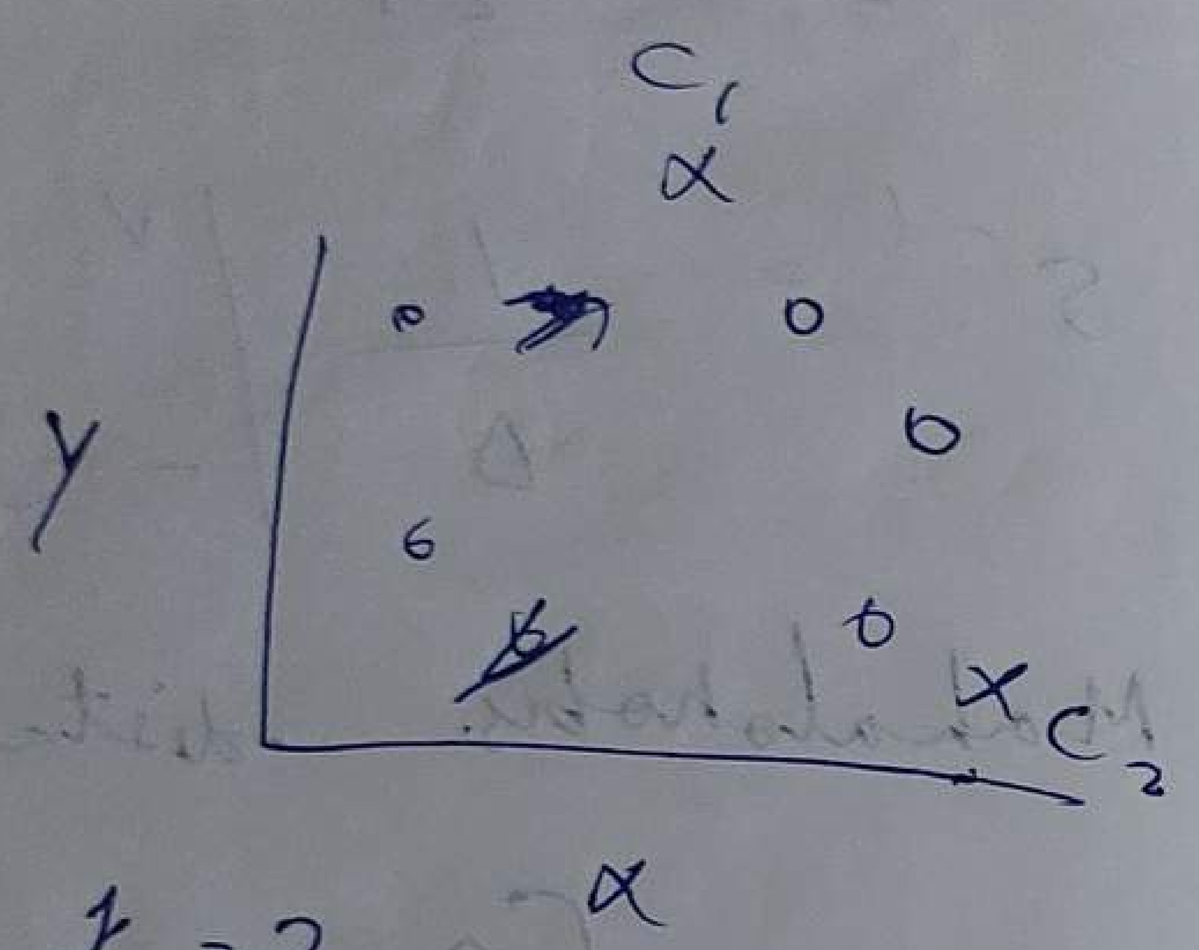
$$\cos\theta = \frac{\vec{x}\cdot\vec{y}}{|x||y|}$$

$$d_{xy} = \sqrt{2(1-\cos\theta)}$$

19/4/2024

k-means

x	y
x_1	y_1
⋮	⋮
⋮	⋮
x_n	y_n



$k = 2$

1. define k
2. seed k centroids randomly $(c_1, \dots, c_k)$
3. calculate dist.

   d_{ij} = dist. of i^{th} point from j (for all i, j)

4. Get minimum (d_{ij}) for i.
5. based on min., assign i data to the cluster.

6. Calculate the real centroids of clusters

~~step~~

~~step~~

Loop until membership of no point changes.

(check if membership of each point has changed (not effective)

OR

check if centroid has changed)

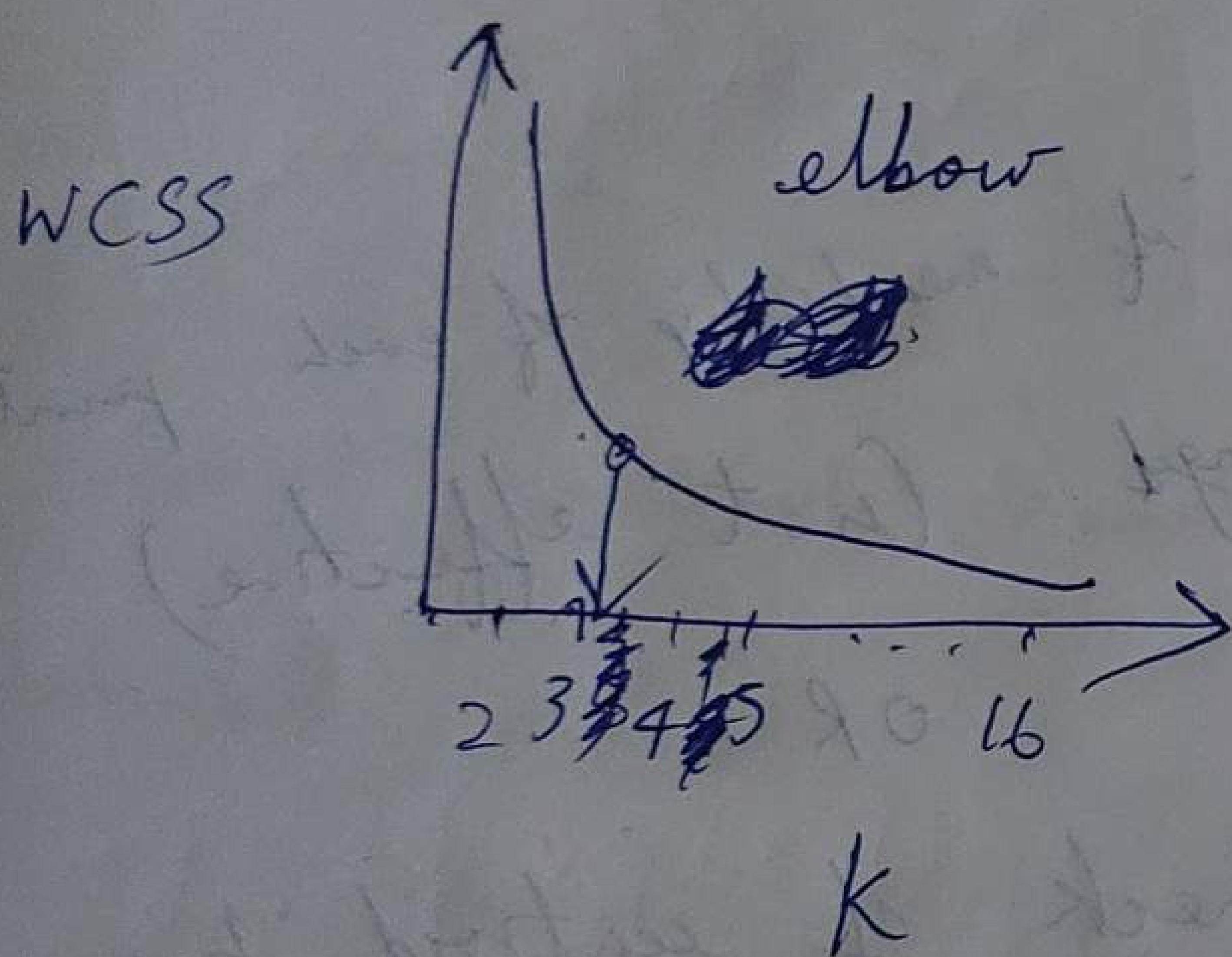
• Within cluster sum of squares (WCSS):

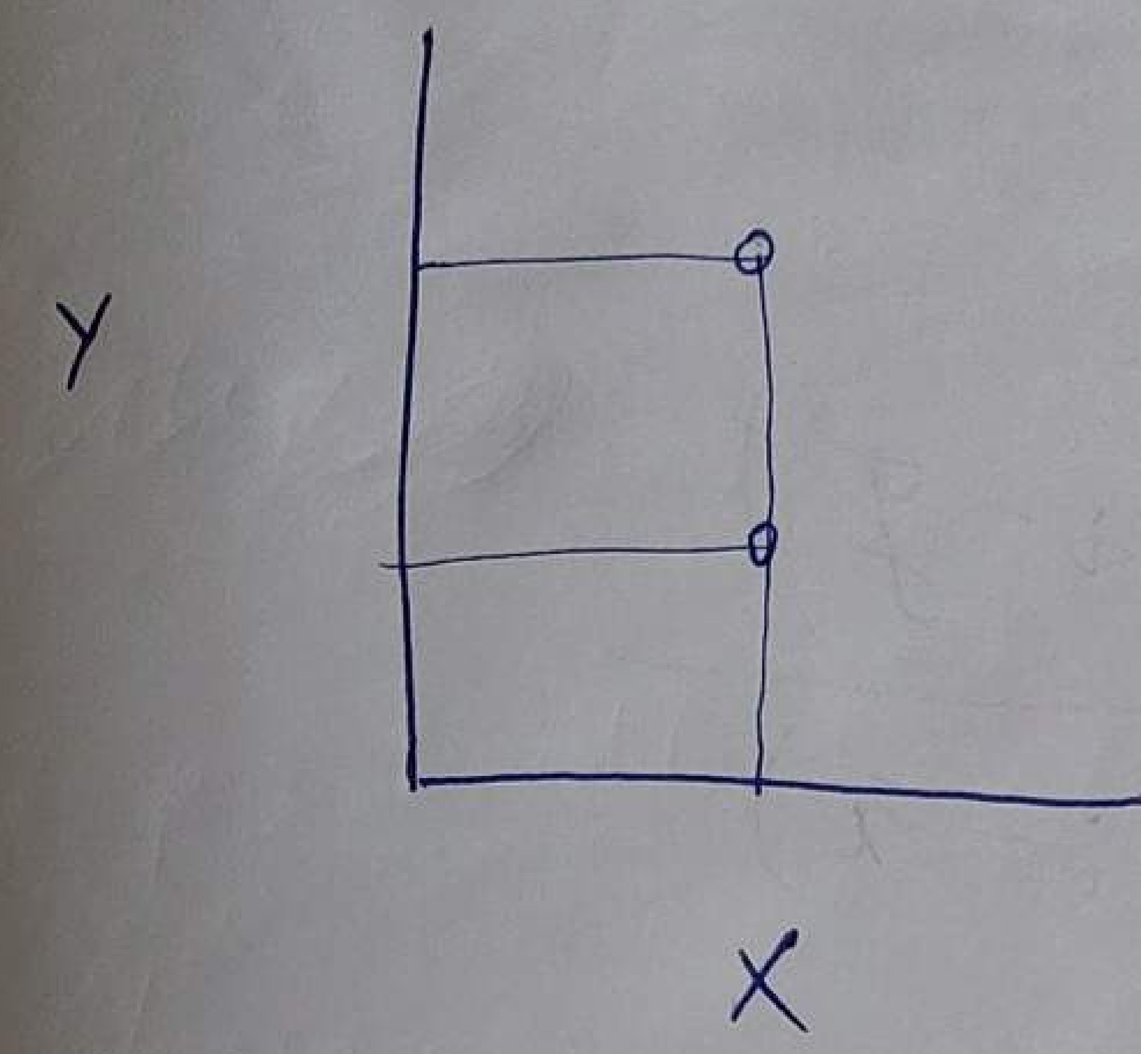
$$WCSS = \sum_{k} \sum_{ij} \left(d_{ij}^{k} - \bar{d}^{k}\right)^2$$

- Between cluster sum of squares (BCSS):

$$BCSS = \sum_{m,n} \left(d_{nm} - \bar{d}_c \right)^2 \; !$$

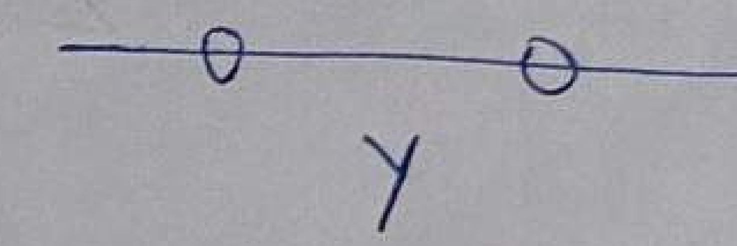
m, n are clusters, $m \neq n$

- 

- 

2D → 1D

high variance



$$X = \begin{vmatrix} x_{11} & \cdots & x_{d1} \\ \vdots & & \vdots \\ x_{1n} & \cdots & x_{dn} \end{vmatrix}_{n \times d}$$

variables (Dimensions)

$\vec{\mu}_{d\times 1}$

$$P_1 = \frac{\begin{bmatrix} x_{11} \\ \vdots \\ x_{d1} \end{bmatrix} \cdot \vec{\mu}}{|\vec{\mu}|} = \frac{1}{|\vec{\mu}|} \underset{(1\times d)}{\begin{bmatrix} x_{11} & \cdots & x_{d1} \end{bmatrix}} \underset{(d\times 1)}{\vec{\mu}}$$

(projection)

$$\underset{(n\times 1)}{P} = \frac{X \vec{\mu}}{|\vec{\mu}|}, \text{ if } |\vec{\mu}| = 1, \; P = X\vec{\mu} = \begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix}$$

Linear regression:

$c = \bar{y} - m\bar{x}$

$$m = \frac{\sum (y_i - \bar{y})}{\sum (x_i - \bar{x})}$$

$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}) \rightarrow cov(x,y)}{\sum (x_i - \bar{x})^2 \rightarrow var(x)}$$

$$\underset{\vec{u}}{\arg\max} \left[var(X\vec{u}) \right]$$

constraint : $|u|^2 = 1$

main objective of PCA (Principle Component Analysis)

$$V(X) = E(X^2) - [E(x)]^2$$

$$V(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{(x_1 + x_2 + x_3 + \dots)}{n^2}$$

$$E(x^2) = \overline{(x^2)} = \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{x_1^2 + x_2^2 + x_3^2 + \dots}{n}$$

$$[E(x)]^2 = \left(\frac{\sum_{i=1}^{n} (x_i)}{n} \right)^2 = \frac{(x_1 + x_2 + x_3 + \dots)^2}{n^2}$$

$$= \frac{n x_1^2 + n x_2^2 + \dots - (x_1 + x_2 + x_3 + \dots)^2}{n^2}$$

22/4/2024 BT307

- Dimension reduction

$$x_{n\times d} \longrightarrow x_{n\times d'}$$

$$d' << d$$

PCA

$x_{n\times d} \rightarrow \vec{v}$

$$\underset{n\times 1}{p} = \underset{(n\times d)(d\times 1)}{X\vec{v}}$$

$$X = \begin{bmatrix} x_{11} & \cdots & x_{d1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x_{1N} & & x_{dN} \end{bmatrix}$$

$$\arg_{\vec{u}} \max [\mathrm{var}(p)]$$

$$\arg_{\vec{u}} \max [\mathrm{var}(X\vec{u})]$$

constraint : $|\vec{u}|^2 = 1$

$$\vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_2 \end{bmatrix}, \quad \vec{v}_c = \begin{bmatrix} v_1 - \bar{v} \\ v_2 - \bar{v} \end{bmatrix}$$

$$\frac{1}{n} \vec{v}_c^T v_c = \mathrm{var}(\vec{v})$$

$$\mathrm{var}(X\vec{u}) = \frac{1}{n}(X\vec{u})_C^T (X\vec{u})$$

centred

$$= \frac{1}{n}(X_C\vec{u})^T (X\vec{u})$$

$x_c = x$ (centered)

$$= \frac{1}{n} \vec{u}^T X^T X u^T$$

$$= \vec{u}^T \frac{X^T X}{n} \vec{u}$$

$$= \vec{u}^T S \vec{u}$$

$$\arg\max (\vec{u}^T S u)$$

Lagrange multiplier :

L = objective function $-$ λ (constraint)

$$L(u, \lambda) = \vec{u}^T S u - \lambda (|\vec{u}|^2 - 1)$$

$u^T u$

$(\because |\vec{u}|^2 - 1 = 0)$

$$\frac{\partial L}{\partial \vec{u}} = \frac{\partial}{\partial \vec{u}}(\vec{u}^T S u) - \frac{\partial}{\partial \vec{u}} \lambda (u^T u - 1) = 0$$

$$[x \ y]^T \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= ax^2 + 2bxy + cy^2$$

$$\frac{dh}{d\vec{x}} = \nabla h = \begin{bmatrix} \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial y} \end{bmatrix} = \begin{bmatrix} 2ax + 2by \\ 2bx + 2cy \end{bmatrix}$$

$$= 2 \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Rightarrow \frac{\partial L}{\partial \vec{u}} = 2S\vec{u} - \lambda(2\vec{u}) = 0$$

$$\Rightarrow S\vec{u} = \lambda\vec{u}$$

$\vec{u}$ is an eigenvector of S,

λ is the eigenvalue.

$X_{n \times d}$ - If variables of X are independent,

$S_{d \times d}$

and $n > d$

then, $\vec{u}_1, \ldots \vec{u}_d$

$\lambda_1, \ldots \lambda_d \geq 0$

When $n < d$,

At least, one $\lambda = 0$.

$S\vec{u}_i = \lambda_i \vec{u}_i$

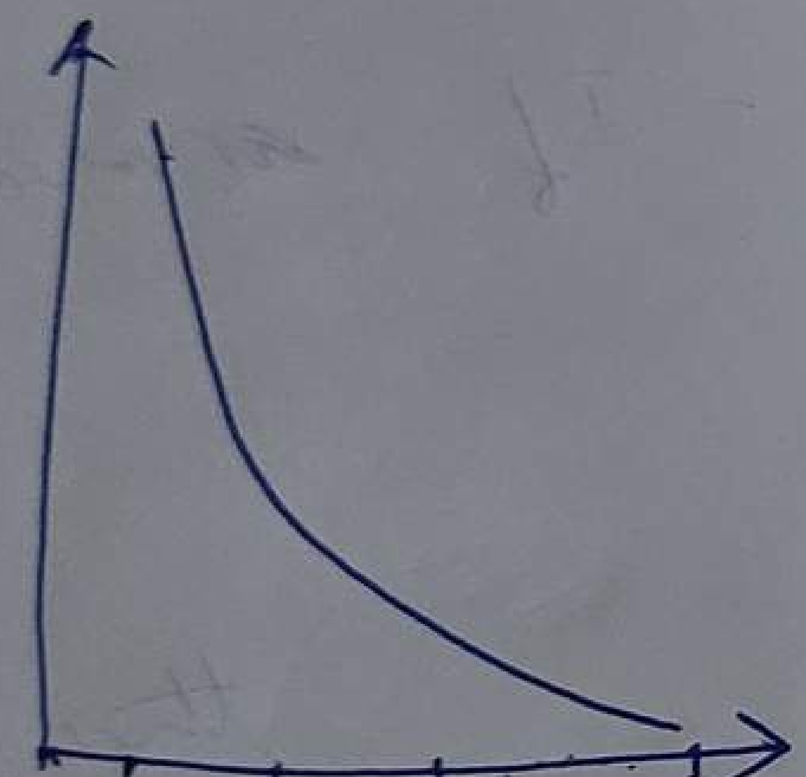
sum of eigenvalues = trace

product of eigenvalue = determinant

$var(X\vec{u}_i) = \lambda_i$

$\sum \lambda_i = var(X)$

$\frac{\lambda_i}{\sum \lambda_i}$



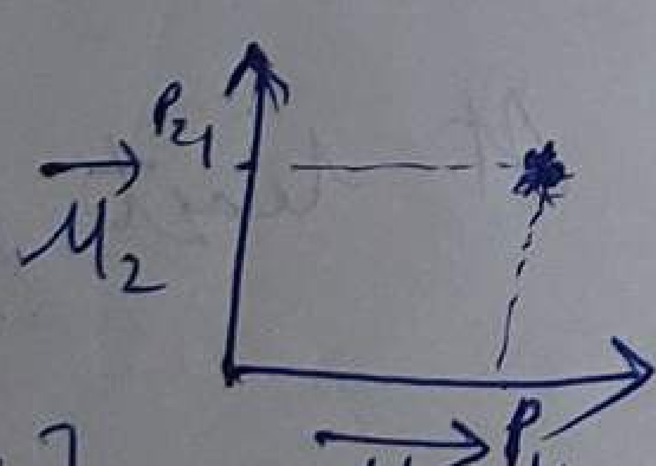
$u_1 \quad u_2 \quad u_3 \quad u_d$

$\downarrow \quad \downarrow \quad \downarrow$

$PC_1 \quad PC_2 \quad PC_3 \ldots$

$X\vec{u}_1 = P_1$

$X\vec{u}_2 = P_2$



$P_1 = \begin{bmatrix} p_{11} \\ \vdots \\ p_{1N} \end{bmatrix}$

$P_2 = \begin{bmatrix} p_{21} \\ \vdots \\ p_{2N} \end{bmatrix}$

$\underset{n\times d}{X} \begin{bmatrix} | & | \\ \vec{u}_1 & \vec{u}_2 \\ | & | \end{bmatrix}_{d\times 2} = \tilde{P} = \begin{bmatrix} | & | \\ \vec{P}_1 & \vec{P}_2 \\ | & | \end{bmatrix}_{n\times 2}$

loading matrix

projection matrix

Linear method of dim. red.

(global dim. red.)

no subjectivity (unique analytical sol.)

(only no. of dim. is chosen)

26/4/2024

$U = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_d \\ | & | & & | \end{bmatrix} \rightarrow$ loading matrix

$X u \in \mathbb{R}_{n \times d}$

$$\begin{bmatrix} \vec{p}_1 & \vec{p}_2 & \cdots & \vec{p}_d \end{bmatrix}$$

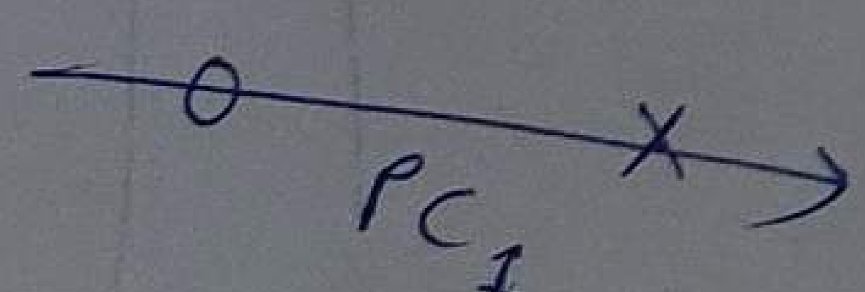
$$X = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

~~[illegible]~~

P.C. 1

$$\vec{u}_1 = \begin{bmatrix} \vec{v}_{11} \\ \vec{v}_{12} \end{bmatrix}$$

$$X\vec{u} = \begin{vmatrix} x_1 v_{11} + y_1 v_{12} \to 0 \\ x_2 v_{11} + y_2 v_{12} \to X \end{vmatrix}$$



Linear combination ⇒ PCA is linear method of dim. red.

global method : because it only consider var. and cov. (global data)

tSNE (t-distributed stochastic neighbour embed)

UMAP

non-linear and local