If $z = 0$

$\Rightarrow X = A$, $Y = B$

As A & B are independent

$\Rightarrow$ X & Y are also independent.

$\Rightarrow \rho = 0$ ~~pq~~

-1	0	$+1$
Perfect -ve linear association.	No linear Association.	Perfect +ve linear association

$$\Rightarrow \cancel{\sigma_X^2 = p} \quad \rho = \frac{pq \; var(Z)}{\sqrt{(p^2 \cdot var(Z) + var(A))} \cdot \sigma_Y}$$

Also for σ_Y

$$\rho = \frac{pq \cdot var(Z)}{\sqrt{(p^2 \cdot var(Z) + var(A))} \cdot \sqrt{q^2 var(Z) + var(B)}}$$

$\rightarrow$ If (A, B are constants and not random variables

then $var(A) = var(B) = 0$

$$\Rightarrow \rho = \frac{pq \cdot var(Z)}{\left(\sqrt{p^2 var\, Z + \cancelto{0}{var(A)}}\right) \sqrt{q^2 var(Z) + \cancelto{0}{var(B)}}}$$

$\Rightarrow$ As std deviations are always +ve

$$\frac{pq \cdot var(Z)}{|pq \cdot var(Z)|} = \pm 1$$

$\Rightarrow \rho = 1$

If p	q	ρ
+	+	+1
+	−	−1
−	+	−1
−	−	+1

Covariance ☆

① measures of joint variability of 2 random variables
② measures association 2 variables.

[Variance : R : var( ) ☆

Correlation:

Eg: $X = pZ + A$
$Y = qZ + B$

Z, A, B : independant random variables

A, B: Noises

$$Y = (q/p)\,X + \left(B - \frac{q}{p}A\right)$$

$$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

☆
$$cov(X,Y) = cov(pZ + A, qZ + B)$$
$$= cov(pZ, qZ) + \cancel{cov(pZ, B)}^{0}$$
$$+ \cancel{cov(qZ, A)} + \cancel{cov(A, B)}^{0}$$
$$= cov(pZ, qZ)$$
$$= pq \cdot var(Z)$$

They are independent ($cov(a,b) = 0$)

$$\Rightarrow \rho = \frac{pq \cdot var(Z)}{\sigma_X \sigma_Y}$$

$\sigma_X^2 = var(pZ + A)$
$= p^2 \cdot var(Z) + var(A)$

On prev page :

Association b/w X+Y ☆

$$cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

$$s^2_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n-1} = cov(X,Y)$$

n−1 for unbiased estimate (in place of n)

$-\infty \le cov(X,Y) \le +\infty$ [R: cov()

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

* Pearson's Correlation

[Correlation does not imply causation]

☆ $$\rho_{xy} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

$\rho \in [-1, 1]$

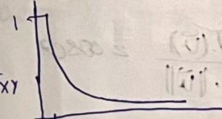
$$r_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[x_i - \bar{x}]^2 [y_i - \bar{y}]^2}}$$

[Read from prev. pages.

Good measure of linear association and only measure of linear association.

[R: cor() specific pearson's method)

If $X \perp Y$



n → sample size

do a t-test for r_{xy}

$H_0: \rho_{xy} = 0$

If Pvalue > cutoff (for eg: 0.01)
⇒ H_0 is not rejected.

* Spearman's correlation

→ Non linear & Monotonic → If data is
→ when the data is rank data.

$$\rho_{xy} = \frac{cov(R(X), R(Y))}{\sigma_{R(X)} \cdot \sigma_{R(Y)}}$$

```
d <- read.csv('data.csv')
rs <- cor(d$v1, d$v2,
      method = "spearman")
rs <- cor(d, method = "s-m")   # whole data
```

Covariance ☆

① measures of joint variability of 2 random variables
② measures association 2 variables.

[Variance : R: var( ) ☆

Correlation:

Eg: $X = pZ + A$; $Y = qZ + B$ — Z, A, B : independant random variables

A, B: Noises

$Y = (q/p)X + (B - \frac{q}{p}A)$

$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

☆ $cov(X,Y) = cov(pZ + A, qZ + B)$
$= cov(pZ, qZ) + \cancel{cov(pZ, B)}^{0}$
$+ \cancel{cov(qZ, A)} + \cancel{cov(A, B)}^{0}$

They are independant ($cov(a,b) = 0$)

$= cov(pZ, qZ)$
$= pq \cdot var(Z)$

$\Rightarrow \rho = \frac{pq \cdot var(Z)}{\sigma_X \sigma_Y}$

$\sigma_X^2 = var(pZ + A)$
$= p^2 \cdot var(Z) + var(A)$

On prev page;

☆ Association b/w X + Y ☆

$cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$

$s_{xy}^2 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n-1} = cov(X,Y)$

$n-1$ ← for unbiased estimate (in place of n)

$-\infty \leq cov(x,y) \leq +\infty$ [R: cov()

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

* Pearson's Correlation

[Correlation does not imply causation]

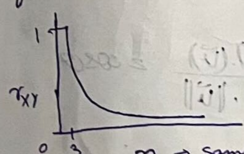
☆ $\rho_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$ [$\rho \in [-1, 1]$]

$r_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[x_i - \bar{x}]^2 [y_i - \bar{y}]^2}}$ [Read from prev. pages.

Good measure of linear association
and only measure of linear association.

[R: cor() specific pearson's method)

If $X \perp Y$



$n \rightarrow$ sample size

do a t-test for r_{xy}

$H_0: \rho_{xy} = 0$

If pvalue > cutoff (for eg: 0.01)
$\Rightarrow H_0$ is not rejected.

* Spearman's correlation

→ Non linear & Monotonic → If data is
→ when the data is rank data.

$$\boxed{\rho_{xy} = \frac{Cov(R(x), R(y))}{\sigma_{R(x)} \cdot \sigma_{R(y)}}}$$

```
d <- read.csv("data.csv")
rs <- cor(d$v1, d$v2,
    method = "spearman")
rs <- cor(d, method = "s m")   # whole data
```

X	Y
x_1	y_1
x_2	y_2
⋮	⋮
x_n	y_n
$\bar{x}$	$\bar{y}$

$$\vec{u} = \begin{vmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{vmatrix} \qquad \vec{v} = \begin{vmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{vmatrix}$$

$$\vec{u} \cdot \vec{u} = \|\vec{u}\|^2$$

$$S_x^2 = \frac{\|\vec{u}\|^2}{n} \qquad S_y^2 = \frac{\|\vec{v}\|^2}{n}$$

$$S_{xy}^2 = \frac{(\vec{u}) \cdot (\vec{v})}{n}$$

$$r_{xy} = \frac{S_{xy}^2}{S_x S_y}$$

$$\Rightarrow r_{xy} = \frac{(\vec{u}) \cdot (\vec{v})}{\|\vec{u}\| \cdot \|\vec{v}\|} = \cos\theta$$

$$|\vec{u} \cdot \vec{v}| \leq \|\vec{u}\| \|\vec{v}\|$$

$$\vec{X} = c\vec{Y} \; [=]$$

$$\vec{X} = c\vec{Y} + b\vec{Z}$$

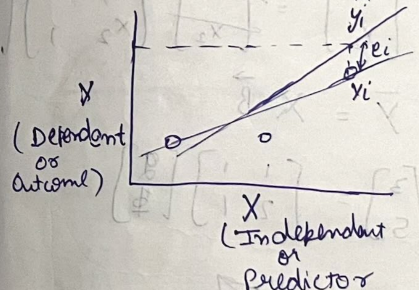
θ

COV MATRIX

$$= \begin{vmatrix} S_x^2 & S_{xy}^2 \\ S_{xy}^2 & S_y^2 \end{vmatrix}$$

Association: Does <u>not</u> imply causation

$$y = f(x)$$
$$= ax + b$$



X	Y
x_1	y_1
x_2	y_2
x_3	y_3

* For given data pts 4
* for an x_i we use the eq. $mx_i + c = y_i$
so for $x_i \Rightarrow mx_i + c = \hat{y}_i$
$\Rightarrow \hat{y}_i$ may not be equal to y_i (from the data.)

$\hat{a}$

$\hat{b}$

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

→ Diff b/w the values

$$e_i = (y_i - \hat{y}_i) \quad \text{For some } \hat{a} \text{ \& } \hat{b}$$

As $\vec{e}_i$ can be +ve or -ve we square it.

$$\boxed{RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Residual sum of squares. from data (Real value) estimated find $\hat{a}$ & $\hat{b}$ so RSS

Our purpose is to minimise the RSS

$$\boxed{\hat{a} = \frac{\sum^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}}$$

Identity

$\sum_{i=1}^{n} (y_i - \bar{y}) = 0$

$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$

$$= \frac{\mathrm{cov}(X,Y)}{S_x^2}$$ [To remember]

$$\boxed{\hat{b} = \bar{y} - \hat{a}\bar{x}}$$ ["]

Eg: $y = 2x+1$

X	Y
1	3
2	5

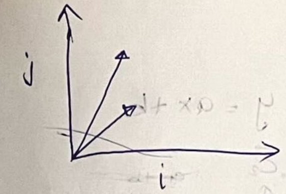
$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$

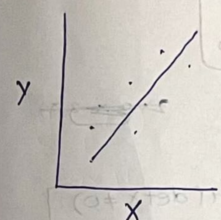
$\bar{y} = X\vec{B}$

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

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

$\vec{X}_{c1} + \vec{X}_{c2} \quad \vec{v}_N$





In tab1 we can make set of eqn: $y_i = mx_i + c$

$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix}$

$\hat{Y} = \hat{X} M$

$y = ax + b$

Derivation at back

$\boxed{\hat{y}_i = ax_i + b}$

Tab 1

X	Y
x_1	y_1
⋮	
x_i	y_i
⋮	⋮
x_N	y_N

$N > 2 \quad \min_{a,b} [RSS] = \left[\sum_{\text{all } i} (y_i - \hat{y}_i)^2\right]$

$\cancel{\vec{q} = x c_1}$

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

$c_1 \quad c_2$

$\vec{q} = x\vec{c}_1 + y\vec{c}_2$

q in in Col_A (column space of A)

Eg:

X	Y
1	4
2	6

$$y = ax + b$$

$$\begin{array}{ccc} & \vec{C_1} & \vec{C_2} \\ \begin{bmatrix} 4 \\ 6 \end{bmatrix} = & \begin{bmatrix} 1 \\ 2 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{array} \begin{bmatrix} a \\ b \end{bmatrix}$$

~~a+b~~ ~~2a+b~~

$Y \qquad X \qquad B$

$\boxed{B = X^{-1} Y}$ [X must be singular ($\det X \neq 0$)]

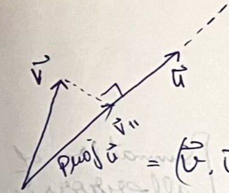
$Y = a\vec{C_1} + b\vec{C_2}$ [$\vec{Y}$ is in col X]

$\vec{Y}$ col X

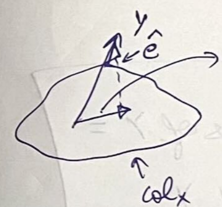
N>2

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

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



$\mathrm{Proj}_{\vec{u}} \vec{v} = (\vec{v} \cdot \vec{u})\, \hat{u}$



$\mathrm{Proj}_{\mathrm{col}\,x} \vec{Y}$

$\hat{Y} = X \cdot \vec{Y} \qquad \hat{e} = \vec{Y} - \hat{Y}$

$= X^T \vec{Y}$

$= XB$ ← For some value of a & b

$\hat{e} \perp X$

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

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

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

$\Rightarrow X^T (Y - XB) = 0$

$\Rightarrow X^T X B = X^T Y$

Hat Matrix

Pseudoinverse

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

$\boxed{\vec{B} = (X^T X)^{-1} (X^T)(Y)}$ ← Sol^n

Linear Reg. Using Linear Algebra

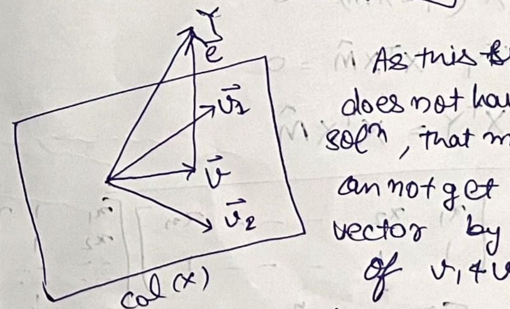
x_i	y_i
x_1	y_1
x_2	y_2
⋮	⋮
x_n	y_n

$$y_i = m x_i + c$$

$$\downarrow$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{array}{cc} v_1 & v_2 \\ \end{array} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix}$$

$$\boxed{Y = XM}$$



As this system does not have a unique solⁿ, that means we can not get 'Y' vector by linear comb. of v_1 + v_2

$\Rightarrow$ Y is is not in Col(X)

Task

Find a vector which is in col(X) and closest to 'Y'

Solⁿ: That vector is: Proj of Y on col(X)

Req. vector: $\underline{\vec{u}} = \text{Proj}_{\text{col}(X)} \vec{Y}$

$\Rightarrow u = X\hat{m}$ $\quad \hat{m} = \begin{bmatrix} \hat{m} \\ \hat{c} \end{bmatrix}$

$\vec{e} = \vec{Y} - \vec{u}$

Replace $u \to X\hat{m}$

$\Rightarrow \vec{e} = Y - u = Y - X\hat{m}$

$\Rightarrow$ e is orthogonal to $\underline{\text{col}(X)}$

$\Rightarrow \underline{X^T e = 0}$

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

$\Rightarrow X^TY - X^TX\hat{m} = 0$

$\Rightarrow X^TY = X^TX\hat{m}$

$$X^TX = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{bmatrix} \to 2\times2 \text{ Matrix}$$

Also

$$X^TY = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$X^TY = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

$$\Rightarrow \hat{m}\sum_{i=1}^{n} x_i^2 + \hat{c}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \quad \text{——— (1)}$$

&

$$\hat{m}\sum_{i=1}^{n} x_i + \hat{c}n = \sum_{i=1}^{n} y_i \quad \text{——— (2)}$$

From (1) & (2) we get

$$\hat{c} = \frac{\sum_{i=1}^{n} y_i - \hat{m}\sum_{i=1}^{n} x_i}{n}$$

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

Also

2) $$\hat{m} = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i}$$

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

⇒ $\hat{e} \cdot \vec{1} = 0$ error

⇒ $(y - \hat{y})^T \cdot \vec{1} = 0$

$\vec{y}^T \vec{1} = \frac{1}{N} \hat{y}^T \vec{1}$ [Summation of all errors is 0]

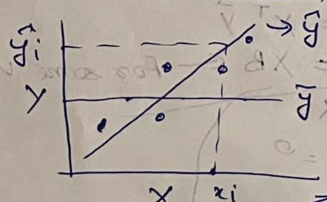
↓

$\bar{y} = \bar{\hat{y}}$

$\boxed{\text{Mean of estimated values of } y = \text{mean of original value of } y}$

$y = Ax + B$ (2)

X	Y



→ $\hat{y} = mx + c$

[Error → $\Sigma(y_i - \hat{y}_i)$

$\boxed{y = c = \bar{y}}$ → average. (1) worst model

Alternate model

→ y does not depend on X

Error for $y = ax + b$

$RSS = \Sigma(y_i - \hat{y}_i)^2$ →

for

$y = c = \bar{y}$ → ★

$\boxed{TSS = \Sigma(y_i - \bar{y})^2}$

Total sum of squares errors.

★ $\boxed{RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$

Reduced sum of squares.

error in linear model.

$\frac{TSS - RSS}{TSS} = R^2$

error in worst model

coefficient of determination.

~~1-RSS~~ $\boxed{1 - \frac{RSS}{TSS} = R^2}$ ★

$TSS = RSS + ESS$ → is true only when there is intercept in the model ($y = ax + b$)

$ESS = \Sigma(\hat{y}_i - \bar{y})^2$ → $R^2 = \frac{ESS}{TSS}$

Explained sum of squares.

$R^2 = \frac{ESS}{TSS}$

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

→ $R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

$= \frac{Var(\hat{y})}{Var(y)}$ ← fitted model

When $R^2 = 0 \Rightarrow$ Fitted linear model is bad model
($RSS = TSS$) $\quad 0 \le R^2 \le 1$

If R^2 close to 1 $\Rightarrow$ our model ($Y = AX+B$) is better
than ($Y = C$) [$C = \bar{Y}$]

$y = ax + b$

$y = \alpha x + \beta$

$\boxed{\bar{X}}$, μ_X

Sample Statistics — Population Parameter

$\boxed{y_i = \alpha x_i + \beta} + \varepsilon_i$ ← The error

Normally distributed.

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

t-test for regression co-efficient

$H_0: \alpha = 0 \Leftarrow$ y does not depend on x

$H_1: \alpha \neq 0$

$$t = \frac{\text{Sample Stat} - \text{Population Parameter}}{\text{SE of Sample Statistics}}$$

SE = Sample error

If H_0 is true

then, $t = \dfrac{a - 0}{SE\ of\ (a)}$

$(n-2) \longrightarrow$ df of t [If $n=2$ then there is a perfect fit [only a st. line]]

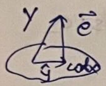
$P = (t \mid H_0 = T) >$ cut off

2-sided t-test

1) Do linear regression & Find a & b
2) Find R^2.
3) Do a t-test for a & b

If $b = 0$ & $a \neq 0$, then do the steps again with $b = 0$

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



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

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

$\Rightarrow \boxed{Y^T \hat{Y} = \hat{Y}^T \hat{Y}}$

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

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

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

$\hat{e} \cdot \bar{y}\vec{1}$

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

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

$\Rightarrow Y^T \bar{Y} = \hat{Y}^T \bar{Y}$

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

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

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

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

$Y^T \hat{Y} = \hat{Y}^T \hat{Y}$

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

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

Gets cancelled

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

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

X	Y
x_1	y_1
$\vdots$	$\vdots$
x_N	y_n

x_1	x_2	Y
x_{11}	x_{21}	y_1
$\vdots$	$\vdots$	$\vdots$
x_{1n}	x_{2n}	y_n

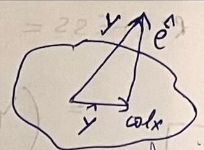
Multiple Linear Regression

$y = a_1 x_1 + a_2 x_2 + b$

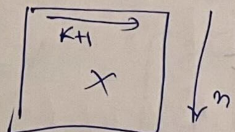
$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ $X = \begin{bmatrix} x_{11} & x_{21} & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & & 1 \end{bmatrix}$

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

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



$\hat{Y} = XB$



$n >> K+1$

If any column is linearly dependent on any of the other columns then it is called

Multilinearity Problem

$t = \frac{x - \bar{x}}{SE}$

$SE \rightarrow (X^T X)^{-1}$

Multilinearity

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

* $^{(K+1)}C_2$

$(x_1, x_2) \rightarrow e_1$

$(x_1, 1) \rightarrow e_2$

$(x_2, 1) \rightarrow e_3$

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

→ Auxiliary regression

$X_j = P_1 X_1 + P_2 X_2 + P_{j-1} X_{j-1} + P_{j+2} X_{j+2} + P_K X_K + q$

↑

R_j^2

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

↑ Trouble

$y = ax + b$

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

$R^2 = 1 - \frac{RSS}{TSS}$

$TSS - \Sigma (y_i - \bar{y})^2$

↑

Coefficient of Determination

$0 \leq R^2 \leq 1$

$y = \bar{y}$

$y = a_1 x_1 + b \longrightarrow RSS_1 = 26$

$y = a_2 x_2 + b \longrightarrow RSS_2 = 13$

$y = a_1 x_1 + a_2 x_2 + b \longrightarrow RSS_3 =$

$a_1 \neq 0$

Every time is any independent element is added the RSS will reduce

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

F-test

Small model $\Rightarrow RSS_S$ $\quad RSS_{\text{small model}} > RSS_{\text{Large model}}$.

Large model $\Rightarrow RSS_L$

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

H_0 : all $\alpha = 0$

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

H_1 : Atleast one $\alpha_1 \neq 0$

$H_0 \Rightarrow y = \bar{y}$

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

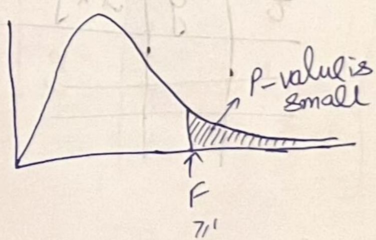
$RSS_L = \Sigma (y_i - \hat{y}_i)^2$

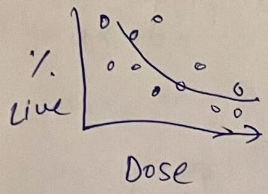
$$\frac{ESS_L}{RSS_L} \Rightarrow \frac{ESS_L / (K)}{RSS_L / (n-K-1)}$$

= F-value

Fdist

$(K, n-K-1)$





$$\frac{dx}{dt} = -Kx$$

$$x = x_0 e^{-Kt}$$

$$\frac{x}{x_0} = e^{-Kt}$$

t	y	ln y

Linearizing

Non linear

↓

Linear

Non-linear

Least square

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

$y_i = e^{K t_i}$ → calc. K in this model.

& minimize → $[\Sigma (y_i - \hat{y}_i)^2]$

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

↳ Polynomial regression.

x	y	$\overset{y}{x^2}$	$\overset{z}{x^3}$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_1 & x_1^2 & x_3^2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & x_n^2 & x_n^3 & 1 \end{bmatrix}$$

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

For supervised learning → we req. labelled data

XRAY → lung cancer (D = Y)

XRAY → normal (D = N)

Classifier ← Trained ← Labblled data

Previous Class

* **Non-linear regression**

① Linearization

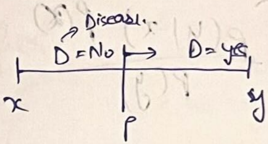
② Polynomial Regression

$$y = a_1 x + a_2 \underbrace{x^2}_{X} + b$$

③ NL Least sequence

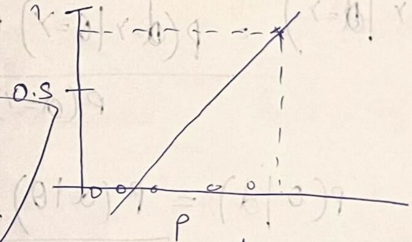
④ Supervised or Unsupervised learning.

Eg:- what amount of disease causing protein 'P' above which will cause disease

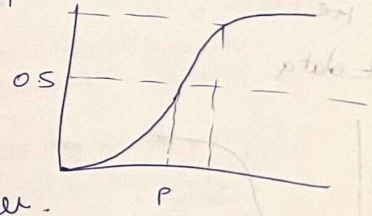


Logistic Regression

$$y = \frac{1}{1-e^{-(ax+b)}}$$



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



Prob of having disease.

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

odd

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

Baye's Theorem

$$P(x/y) = \frac{P(y/x)\,P(x)}{P(y)}$$

theta, data

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

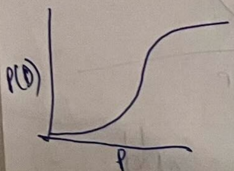
$\theta = \{a, b\}$

↓

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

$$P(\theta \mid d) = \frac{P(d \mid \theta) . P(\theta)}{P(d)}$$

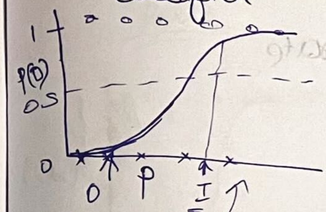
d=data



Logistic Regression

↓

Classifier



Different values of P in blood.

- Every classifier starts with labelled data.
- we try to fit a sigmoidal curve

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

logistic eqn

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

Maximum likelihood estimate:

↓

A method by which we fit a sigmoid in the data.

Test data

Contigency table

Confusion Matrix

		Actual 1	Actual 0
Prob	1	TP	FP
	0	FN	TN

P → Positive

N → Negative

T/F → True/False

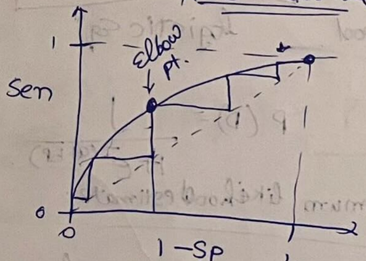
Sensitivity: True Positivity rate.

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

Specificity: True Neg. rate

$$= \frac{TN}{TN+FP}$$

ROC Curve

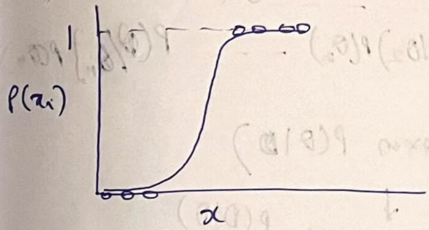


$$y = \frac{1}{1+e^{-(a+b_1x_1+\dots+b_nx_n)}}$$

~~D+~~ND	D	V ← Vulnerable
~~N~~D+V	ND+V	D+ND
P(ND)	P(D)	P(V)

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

Rev



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

Maximum Likelihood Method

↓ ~~Base~~ Baysian

$$P(y|x) = \frac{P(x|y) \cdot P(y)}{P(x)}$$

Eg: Prob of head

$pH = P \Rightarrow \boxed{Q}$

$\boxed{n = 19, \quad h = 4}$

head Data

$Q = \{pH\}$

$Q = \theta, = 0.2$

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

$P(\theta|D)$ → Posterior

$P(\theta)$ ← Prior

$P(D)$ ← marginal

$\theta = \theta_1, \theta_2, \ldots \theta_n$

$$P(D) = P\left(\frac{D}{\theta_1}\right)P(\theta_1) + P(D|\theta_2)P(\theta_2) \ldots P(D/\theta_n)\,P(\theta_n)$$

If we want to maxm $P(\theta|D)$

↓

maximize $P(D|\theta)$