classmate
Date
Page

19/8/24 → Testing of Hypothesis ↗ Null & alternative
↘ Can be simple or composite

	T	F
T	✓	FP
F	FN	✓

Type I error – FP (actually true, predict false)

Type II error – FN

Neyman – Pearson Lemma (NP Lemma)

$H_0 : \theta = \theta_0 \in \Theta_0$ & $H_1 : \theta = \theta_1 \in \Theta_1$

Existence :- $\exists\, \phi(x)$ & k for $H_0 : \theta = \theta_0$ & $H_1 : \theta = \theta_1$

a) $\mathbb{E}_{\theta_0}(\phi(x)) = \alpha$

b) $\phi(x) = \begin{cases} 1 & p_{\theta_1} > k p_{\theta_0} \\ 0 & p_{\theta_1} \le k p_{\theta_0} \end{cases}$ MP ≡ Most Powerful

Sufficiency :- If $\phi(x)$ satisfies a) & b) ϕ is MP test

Necessary condⁿ :- ϕ is MP ⇒ ϕ satisfies a) & b) unless there exist test ϕ whose size is strictly less than α and power 1.

$\boxed{\mathbb{E}_{\theta_0}(\phi(x)) < \alpha \text{ \& } \mathbb{E}_{\theta_1}(\phi(x)) < 1}$

Pf :-

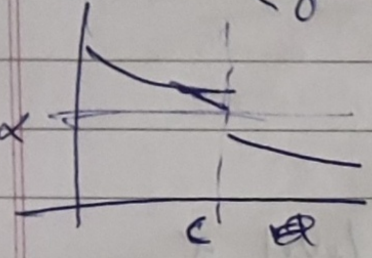
$r(x) = \dfrac{p_{\theta_1}(\underline{x})}{p_{\theta_0}(\underline{x})} \sim F_\theta(x)$

$\alpha(c) = P[r(X) > c] = 1 - F_\theta(c)$

$\lim_{\epsilon \to 0^+} \alpha(c+\epsilon) = \alpha(c) \neq \alpha(c-\epsilon)$

may or may not be equal

$$\phi(x) = \begin{cases} 1 & r(x) = \dfrac{p_{\theta_1}}{p_{\theta_0}} > k \\ \gamma & r(x) = k \\ 0 & r(x) < k \end{cases} \qquad c = k$$



$$\alpha = E_{\theta_0}(\phi(x)) = P[r(X) > c] + \gamma P[r(X) = c]$$
$$= \alpha(c) + \gamma[\alpha(c) - \alpha(c^-)]$$

$\alpha(c) \le \alpha \le \alpha(c^-)$

$$\gamma = \frac{\alpha - \alpha(c)}{\alpha(c) - \alpha(c^-)}$$

$0 < \gamma < 1$

Sufficiency cond^n :-

$E_{\theta_1}(\phi) \ge E_{\theta_1}(\phi^*)$ $\forall \phi^*$ you need to prove

ϕ^* is level-α test

$$\int \overbrace{(\phi - \phi^*)(p_{\theta_1} - k p_{\theta_0})}^{\text{int}} \, d\mu \ge 0$$

$$\Leftrightarrow \left[\int \phi p_{\theta_1} d\mu - \int \phi^* p_{\theta_1} d\mu\right] \ge k\left[\int \phi p_{\theta_0} d\mu - \int \phi^* p_{\theta_0} d\mu\right]$$

$$= k\left(\underbrace{E_{\theta_0}(\phi) - E_{\theta_0}(\phi^*)}_{\ge 0}\right)$$

$$\Leftrightarrow E_{\theta_1}(\phi) - E_{\theta_1}(\phi^*) \ge 0$$

$\phi^* < 1$

$\phi = 1$

$\Leftrightarrow p_{\theta_1} > k p_{\theta_0}$

int ≥ 0

$\phi = 0, \; p_{\theta_1} < k p_{\theta_0}$

then int ≥ 0

<u>Necessary cond^n :-</u>

ϕ^* is another test of level α

$E_{\theta_0}(\phi^*) \leq \alpha$

If $E_{\theta_0}(\phi^*) \leq \alpha$ it is level α &

$E_{\theta_0}(\phi^*) = \alpha$ it is size α

ϕ is an MP test

ϕ^* is another test that satisfies a) & b).

T.S.T $\phi = \phi^*$ except for all pts for which $p_{\theta_1} = k p_{\theta_0}$ a.e μ

$S^+ = \{\underline{x} : \phi(x) > \phi^*(x)\}$

$S^- = \{\underline{x} : \phi(x) < \phi^*(x)\}$

$S^0 = \{\underline{x} : \phi(x) = \phi^*(x)\}$

i.e

$\mu\{\underbrace{x : \phi \neq \phi^* \;\&\; p_{\theta_1} \neq k p_{\theta_0}}_{S}\} = 0$ — Proved later.

$\int(\phi - \phi^*)(p_{\theta_1} - k p_{\theta_0})\,d\mu = 0$

$E_{\theta_1}(\phi) - E_{\theta_1}(\phi^*) = k\left(E_{\theta_0}(\phi) - E_{\theta_0}(\phi^*)\right) \leq 0$ → ①

$E_{\theta_1}(\phi) \geq E_{\theta_1}(\phi^*)$ — ② ~~because~~ ϕ is MP & ϕ^* is $\leq \alpha$

As ϕ is a MP Test

9/1/2x Q) Let $X_i \; i=1:n \sim N(\theta, 1)$. $H_0: \theta = \theta_0 \quad H_1: \theta \neq \theta_0$

$$\lambda(\underline{x}) = \frac{\sup_{\theta \in H_0} L(\theta)}{\sup_{\theta \in H_0 \cup H_1} L(\theta)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_0)^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

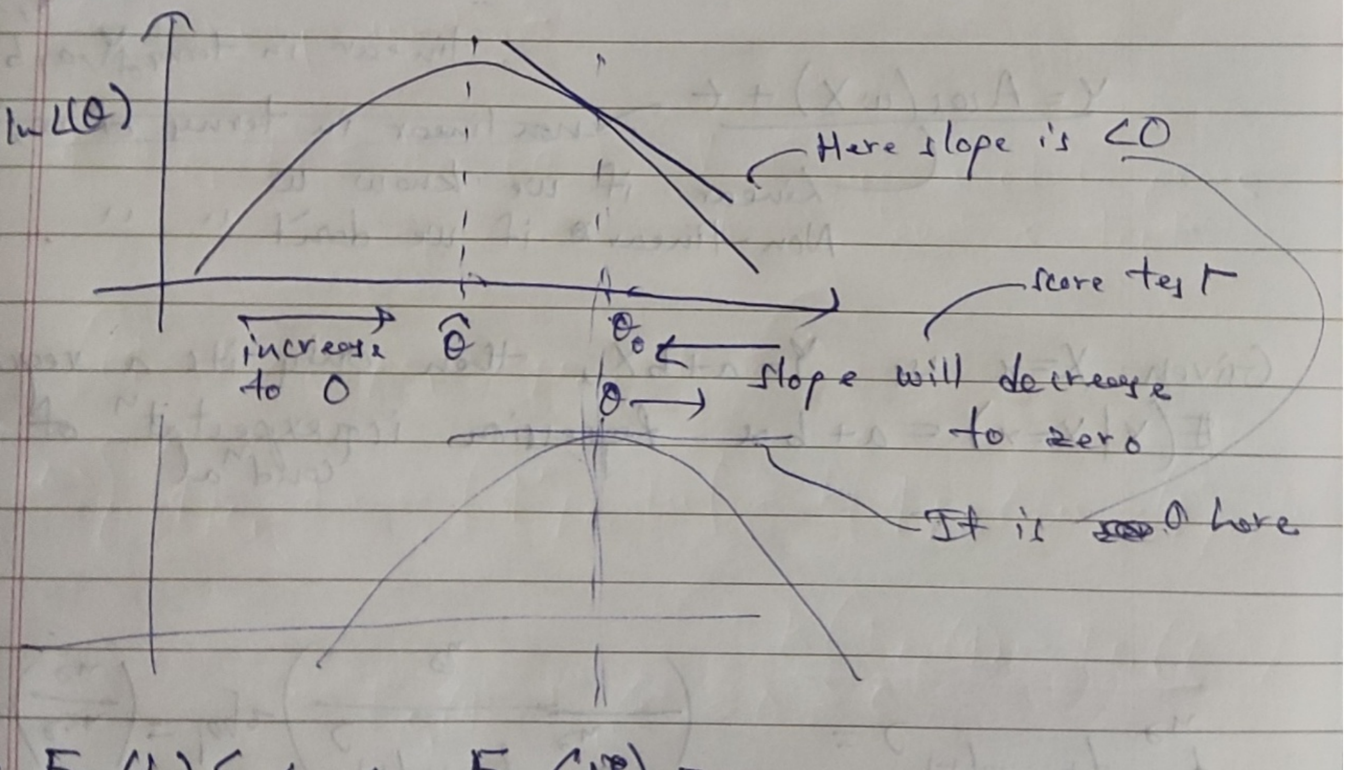
$$= e^{-\frac{n}{2}(\bar{x}-\theta_0)^2}$$

$$R = \{x : \lambda \leq c\} = \{x : e^{-\frac{n}{2}(\bar{x}-\theta_0)^2} \leq c\}$$

$$R = \left\{x : |\bar{x}-\theta_0| \geq \sqrt{\frac{-2}{n}\ln c}\right\}$$

Equivalent pf

Wald test $\equiv$ Likelihood test $\equiv$ Score test



$E_{\theta_0}(\phi) \leq \alpha$ or $E_{\theta_0}(\phi^*) = \alpha$

Likelihood ratio test:-

$$\lambda(x) = \frac{L(\theta_0)}{L(\hat{\theta})} < c$$

$$\ln \lambda = \ln L(\theta_0) - \ln L(\hat{\theta}) < \ln c$$

as $\ln c < 0$, $|\ln c| = -\ln c$ (as $0 < c < 1$)

$$|\ln L(\theta_0) - \ln L(\hat{\theta})| > |\ln c|$$

2/9/24. Regression (Multiple Linear) - MLR

X	Y
factor	response
Covariates	Dependent var
Independent variable	

$Y = a + bX + \epsilon$ — Noise

— linear in terms of Y, X

— " " " " Y, a, b

$Y = a + be^{X} + \epsilon$ — Non linear in terms of Y, X

linear in terms of Y, a, b

$Y = A\cos(\omega X) + \epsilon$ — ~~Non linear in terms of ω, X~~

— Linear if we know ω

Non-linear if we don't " " .

Given $X = x$ is $Y = a + bX$, then only its a regression.

$E(Y|X=x) = a + bx$ Regression is conditional expectation of response.

$$\log_q\left(\frac{e_m}{e_{n+1}}\right) = \log_q\left(\frac{e_n}{\frac{c}{1+p+\dots+p^n}\, e_n}\right)$$

$e_{n+1} = C e_n p$ $\quad$ $e_{n+m} = C e_{n+m-1} p = C\left(C e_{n+m-2}\, p\right) p = \frac{c}{1+p+\dots+p^n}\, e_n\, p^m$

$Y_1 = a_1 + a_2 X_{21} + a_3 X_{31} + \epsilon_1$
$Y_2 = a_1 + a_2 X_{22} + a_3 X_{32} + \epsilon_2$

$Y_n =$

$$Y = \begin{pmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ \vdots & \vdots & \vdots \\ 1 & X_{2n} & X_{3n} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\boxed{Y^{n\times 1} = X^{n\times 3} \beta^{3\times 1} + \epsilon^{n\times 1}}$$

Assumpt^ns :-

1) $\mathbb{E}(\epsilon_i) = 0$ 2) $Var(\epsilon_i) = 1$ 3) $Cov(\epsilon_i, \epsilon_j) = 0 \iff i \neq j$

$\mathbb{E}(\underline{\epsilon}) = \underline{0}$
$Var(\underline{\epsilon}) = I$

8/9/24 Regression

$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ $\quad \mathbb{E}(\underline{\epsilon}) = 0 \quad V(\underline{\epsilon}^{n\times 1}) = \sigma^2 I$ ✗

$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots \\ \cdots & & \cdots \end{pmatrix}$

Auto-correlat^n

$$\rho^* = \frac{\sum_{t=1}^{n}(\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^{n} \epsilon_t^2} = 2(1-\rho)$$

$$\rho = \frac{\sum \epsilon_i \epsilon_{i-1}}{\sum \epsilon_{i-1}^2}$$

$\hat{\epsilon} = f(x_i)$

Choose P s.t
$\mathbb{E}(P\underline{\epsilon}) = 0$
$V(P\underline{\epsilon}) = P\Sigma P'$ (after spectral decomp)
$= I$

$P\underline{Y} = PX\underline{\beta} + P\underline{\epsilon}$

$\mathbb{E}(P\underline{\epsilon}) = 0$

$V(P\underline{\epsilon}) = P\Sigma P' \quad P = \Sigma^{-1/2}$
$= I$

$Y^* = X^*\beta + \underline{\epsilon}^*$ $\quad \mathbb{E}(P\underline{\epsilon}) = P\mathbb{E}(\underline{\epsilon})$
$= 0$

1) Validat^n of different assumpt^ns in regression

$$\tilde{\epsilon}_j^* = \frac{\hat{\epsilon}_j}{\sqrt{Var}} \qquad \underset{\sim}{Y} \sim N(X\underline{\beta}, I\sigma^2)$$

$\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta}$ $\qquad H = X(X'X)^{-1}X'$
$= Y - X(X'X)^{-1}X'Y$
$= (I-H)Y \Rightarrow \mathbb{E}(\hat{\epsilon}) = 0$

$Var(\hat{\epsilon}) = (I-H)V(Y)(I-H) = (I-H)\sigma^2$

is normal

Y is normal $\Rightarrow (I-H)Y \Rightarrow \hat{\epsilon} \sim N(0, (I-H)\sigma^2)$

$\hat{\epsilon}_j = e_j^T \hat{\epsilon} \sim N(0, (1-h_{jj})\sigma^2)$

$(\hat{\epsilon} \sim N(0, (I-H)\sigma^2))$

$\tilde{\hat{\epsilon}}_j = \dfrac{\hat{\epsilon}_j}{\sqrt{1-h_{jj}}\,\hat{\sigma}_{(j)}}$ (and remember $\hat{\sigma}^2 = \dfrac{\hat{\epsilon}^T \hat{\epsilon}}{n-r-1}$)

not the usual notat^n

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{1\,\overline{r+1}} \\ \vdots & & \\ X_{n1} & \cdots & X_{n\,\overline{r+1}} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{\overline{r+1}} \end{pmatrix}$$

$\hat{\sigma}_j^2 = \dfrac{\hat{\epsilon}^{(j)T}\hat{\epsilon}^{(j)}}{\overline{n-1} - \overline{(r+1)}}$

$\sim t_{n-j-1}$

t-distribut^n $\underline{\underline{\text{H.W}}}$

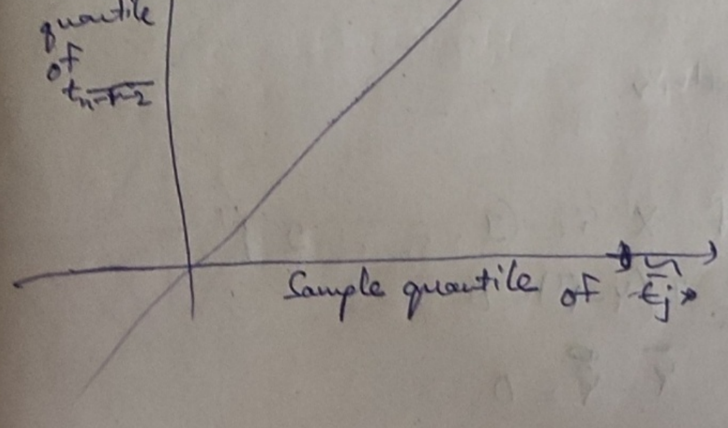
$\tilde{\epsilon}_{j*} = \dfrac{\dfrac{\hat{\epsilon}_j}{\sqrt{1-h_{jj}}\,\sigma}}{\sqrt{\dfrac{(n-r-1)\hat{\sigma}_j^2}{\sigma^2 (n-r-1)}}}$ — $\dfrac{\hat{\epsilon}_j}{\sqrt{1-h_{jj}}\,\sigma} \sim N(0,1)$, $(n-r-1)\hat{\sigma}_j^2 / \sigma^2 \sim \chi^2_{n-r-1}$

both are independent
$\hat{\sigma}_j$ is obtained by deleting j^{th} row in y and X

$$\boxed{\text{IF } X \sim N(0,1) \text{ and } Y \sim \chi_r^2 \text{ \& } X \text{ and } Y \text{ are independent, } W = \frac{X}{\sqrt{\frac{Y}{r}}} \sim t_r}$$

$\tilde{\hat{\epsilon}}_{j*} = \dfrac{\hat{\epsilon}_j}{\sqrt{(1-h_{jj})\,\hat{\sigma}_j^2}} = \dfrac{\dfrac{\hat{\epsilon}_j}{\sqrt{(1-h_{jj})\,\sigma^2}}}{\sqrt{\dfrac{(\overline{n-1} - \overline{r+1})}{(\overline{n-1} - \overline{r+1})} \dfrac{\hat{\sigma}_j^2}{\sigma^2}}} \sim t_{\overline{n-1} - (\overline{r+1})}$

QQ plot



1/4/24

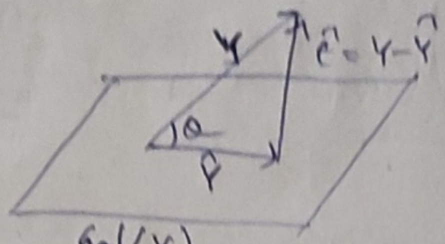
9th problem – Statistical Inference parameters

$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ $\quad X = x \quad \mathbb{E}(\epsilon) = 0 \quad V(\epsilon) = \sigma^2 I$

$\hat{\beta}_{LSE} = (X'X)^{-1}X'Y$

$\underline{\hat{Y}} = X\hat{\beta}$ = predict^n of Y

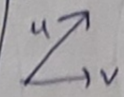
$\underline{\epsilon} = (\underline{Y} - \underline{\hat{Y}})$



Col(X)

$\hat{Y} \perp \hat{\epsilon}$

$\hat{\epsilon}^T\hat{Y} = 0$



Proj of u on v

$= \|u\| \cos\theta \dfrac{v}{\|v\|}$

cosine similarity

$= \left\langle \dfrac{u}{\|u\|}, \dfrac{v}{\|v\|} \right\rangle$

$= \underline{u}^T\underline{v}$

$= \mathbb{E}\left(\dfrac{u}{\|u\|} \dfrac{v^T}{\|v\|}\right)$

= Pearsonian correlat^n (X, Y)

$= \mathbb{E}\left(X - \mathbb{E}(\underline{X})\right)\left(Y - \mathbb{E}(\underline{Y})\right)^T$ / $\|X - \mathbb{E}(X)\| \cdot \|Y - \mathbb{E}(Y)\|$

same as normal correlat^n

$= \dfrac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}}$

$\hat{Y} = X\hat{\beta}$ = projexn of vector y on Col(X)

$\|y\|\cos\theta$

Pf:- $\underline{\hat{Y}} \perp \hat{\epsilon}$

$\hat{\epsilon} = (I - H)\hat{Y}$ $\quad H = X(X'X)^{-1}X'$ $\quad$ H is symmetric

$\cancel{\hat{\epsilon}^T\hat{Y} = Y^T(I-H)X}$ $\quad HX = X \quad H^T = H$

$\hat{\epsilon}^T\hat{Y} = Y^T(I-H)X(X'X)^{-1}X'Y$

$= Y^T[X - HX]XX'XY \rightarrow ?$

$= 0$

$Y_1 = \beta_0 + \beta_1 X_{11} + \ldots + \beta_r X_{r1} + \epsilon_1$

$\vdots$

$Y_n = \beta_0 + \beta_1 X_{1n} + \ldots + \beta_r X_{rn} + \epsilon_n$

$$y = \underbrace{\begin{pmatrix} 1 & X_{11} & \ldots & X_{r1} \\ \vdots & & & \\ 1 & X_{1n} & \ldots & X_{rn} \end{pmatrix}}_{\text{design matrix}} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_r \end{pmatrix}$$

Proj of u on v

$= \langle \underline{u}, \underline{v} \rangle \dfrac{v}{\|v\|^2}$

$= \langle \underline{u}, \underline{v} \rangle \dfrac{v}{v^T v}$

$X^T\underline{\epsilon} = 0$

As 1st column of X is $(1 \ldots 1)^T$

$1'\underline{\epsilon} = 0 \Rightarrow \frac{1}{n} 1'(Y - \hat{Y}) = 0$

$\bar{Y} - \bar{\hat{Y}} = 0$

$$\sum_{i=1}^{n} Y_i^2 = Y'Y = (Y-\hat{Y}+\bar{Y})^T(Y-\hat{Y}+\bar{Y})$$
$$= \underbrace{(Y-\hat{Y})}_{\hat{\epsilon}}{}^T(Y-\hat{Y}) + \hat{Y}^T\hat{Y} - n\bar{Y}^2$$
$$\sum Y_i^2 - n\bar{Y}^2$$
$$= \sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_{i=1}^{n}(Y-\hat{Y}_i)^2 = \hat{\epsilon}^T\hat{\epsilon} = \sum_{i=1}^{n}\hat{\epsilon}_i^2 + \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2$$

17/8/24

Measure of Goodness of fit in Regression Setup

R^2 = inter variation of predictors $\hat{Y}_i$ w.r.t total variances in Y_i

$$= \frac{\sum(\hat{Y}_i-\bar{\hat{Y}}_i)^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = 1 - c\,\frac{\sum_{i=1}^{n}\hat{\epsilon}_i^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

Measures linear relat^n ship in a regression

(from some specific c that makes it χ^2 distribut^n)

$0 \le R^2 \le 1$ $\quad R^2 \to 1 \Rightarrow \sum\hat{\epsilon}_i$ is very very small

$R^2 \to 0 \Rightarrow$ Regression is good.

very large

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \epsilon_i$$

Variance of each β can be very high

$$\boxed{\hat{y} = X\hat{\beta}}$$

Variations of Estimate of $\hat{\beta}$

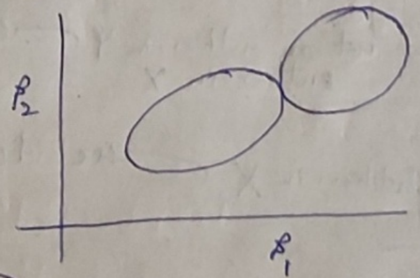
so we $\min \sum \epsilon_i^2 \ni \sum_{i=1}^{n} \beta_i^2 \le c \Leftrightarrow V(\hat{\beta}) \le c$

This regression is called ridge regression

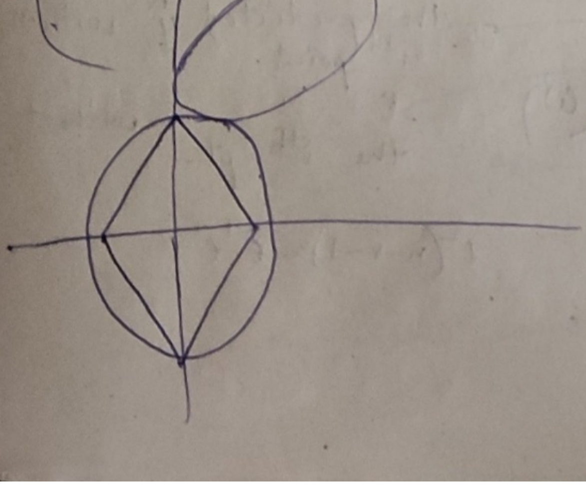
$$L(\beta) = (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) + \lambda \underline{\beta}'\underline{\beta}$$

LASSO

this is ellipsoid



instead of this, $\sum_{i=1}^{n}|\beta_i| \le c$



Elastic Net

Combination of Lasso and Ridge

$$\lambda\left(\sum_{i=1}^{n}|\beta_i|\right) + (1-\lambda)\sum_{i=1}^{n}\beta_i^2 \le c$$

Statistical Interference in a Regression set up

$$y = X\beta + \epsilon \qquad E(\epsilon)=0$$

$$V(\epsilon) = \sigma^2 I$$

Aim:- Estimate of β & σ^2 using MLE

Data:- $(x^{(i)}, y^{(i)})$
$i = 1(1)n$

Given $X = x$,

$$L(\beta, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{(y - x^T\beta)^T(y - x^T\beta)}{2\sigma^2}}$$

Given σ^2

$$\beta_{MLE} = \arg\max_{\beta} L = \arg\min_{\beta} (y - x\beta)'(y - x\beta) = \hat{\beta}_{LSE}$$

$$\frac{\partial(\ln L)}{\partial\sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}(y - x\hat{\beta})'(y - x\hat{\beta}) = \frac{\sum_{i=1}^{n}\epsilon_i^2}{n} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n}$$

10/9/24

Statistical Inference:-

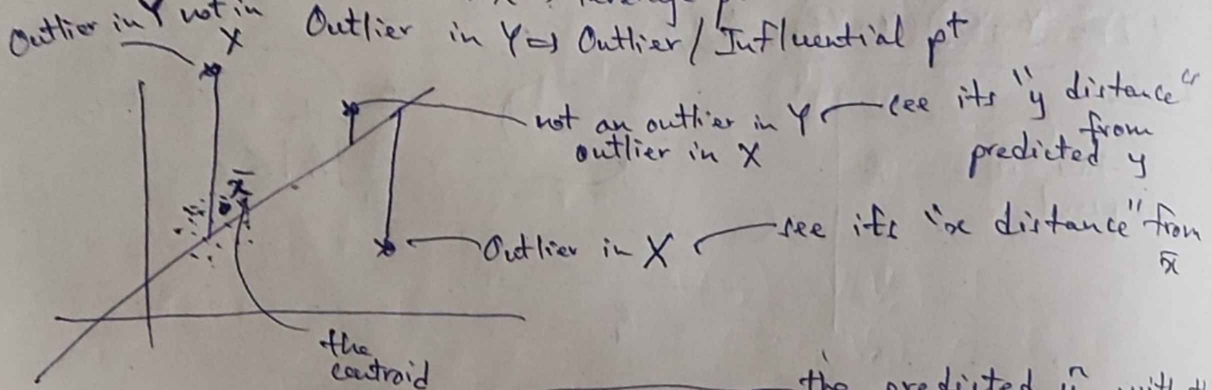
* "Interval Guess" * Model selection problem * Outlier detection

$$\{(x^{(i)}, y^{(i)}) : i = 1(1)n\}$$

$$Y = X\beta + \epsilon$$

Outlier in $X \Rightarrow$ leverage pt

Outlier in $Y \Rightarrow$ Outlier / Influential pt



Cook's distance for i^{th} observation (to detect outlier)

$$= \frac{(\hat{y} - \hat{y}^{(i)})^T(\hat{y} - \hat{y}^{(i)})}{s(r+1)}$$

$\hat{y}$: the predicted $\hat{y}$ with the i^{th} point

$\hat{y}^{(i)}$: " " " without the i^{th} pt.

$$s^2(n-r-1) = \hat{\epsilon}^T\hat{\epsilon}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \hline x_{n1} & \cdots & x_{np} \end{pmatrix}^{n-1\times 1} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}^{p\times 1} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}^{n-1\times 1} \Rightarrow y^{(i)} = X^{(i)}\beta + \epsilon^{(i)} \Rightarrow$$

$$\hat{\beta}^{(i)} = X$$

$$\hat{\beta}^{p\times 1} = (X'X)^{-1}X'Y$$

$$\hat{\beta}^{(i)p\times 1} = (X^{(i)'}X^{(i)})^{-1} X^{(i)T} Y^{(i)}$$

$$\text{Cook dist} = \frac{(X\hat{\beta} - X\hat{\beta}^{(i)})^T (X\hat{\beta} - X\hat{\beta}^{(i)})}{(r+1)s}$$

$$= \frac{(\hat{\beta} - \hat{\beta}^{(i)})(X'X)(\hat{\beta} - \hat{\beta}^{(i)})}{(r+1)s}$$

<u>Outlier in X</u>

$$\hat{y} = HY$$

$$\hat{y}_j = \sum_{i=1}^{n} h_{ji} y_i = h_{jj} y_j + \sum_{i\neq j} h_{ji} y_i$$

h_{jj} = measures the distance b^n $\underline{x}_j$ & centroid of $\underline{x}$

$\equiv$ very high

$\Rightarrow j^{th}$ obs is an outlier in X

<u>22/4/24</u> How eigen vectors are best for extracting var (PCA):-

For λ_1
$$\frac{x^T B^{p\times 1} x}{x'x} = \frac{x^T P \Lambda P' x}{x^T P P} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_{i=1}^{p} \lambda_i y_i^2}{\sum y_i^2} = \lambda_1 \frac{\cancel{\sum y_i^2}}{\cancel{\sum y_i^2}} = \lambda$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

For some λ_{k+1},

$$\frac{x^T B \underline{x}}{x^T x} = \frac{\underline{x}^T P \Lambda P' \underline{x}}{\underline{x}' x} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_{i=1}^{p} \lambda_i y_i^2}{\sum y_i^2} \to (1)$$

$$y = P'x$$

$$x = Py = [e_1 \cdots e_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = y_1 e_1 + \cdots + y_n e_n$$

$$e_i^T \underline{x} = y_1 e_i^T e_1 + y_2 e_i^T e_2 + \cdots + y_n e_i^T e_n$$

$$= y_i$$

$x \perp e_i \quad i = 1(1)k \qquad e_i^T x = 0 \quad i = 1(1)k$

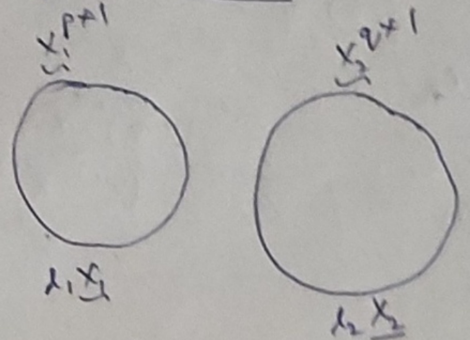
$$(1) = \frac{\sum_{k+1}^{p} \lambda_i y_i^2}{\sum_{k+1}^{p} y_i^2} \le \lambda_{k+1}$$

Two ways of saying what component is dominant:
- e_{ij} is dominant
- $Cov(Y_i, X_k)$ is max

23/4/24

Canonical Correlatⁿ :-

PCA



$\underset{l_1, l_2}{\operatorname{argmax}} \ \text{Corr}(l_1 \underset{\sim}{X_1}, l_2 \underset{\sim}{X_2}) = (l_1^*, l_2^*)$

~~argmax~~ U can create min{p,q} direxns. $p \le q$

$$\underset{\sim}{X}^{\overline{p+q}\times 1} = \begin{bmatrix} X_1^{p\times 1} \\ X_2^{q\times 1} \end{bmatrix} \quad \mathbb{E}(\underset{\sim}{X}) = \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad V(\underset{\sim}{X}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

1st canonical correlatⁿ (U_1, V_1) which are linear combinatⁿ $\underline{X_1}$ & $\underline{X_2}$ that maximizes the $\text{corr}(U_1, V_1)$.

ith canonical correlatⁿ (U_i, V_i) " " " " " " " $\ni$ it maximizes the $\text{corr}(U_i, V_i)$ & the pair is uncorrelated with prev canonical components.

$U_1 = e_1^T \Sigma_{11}^{-1/2} X^{(1)} \qquad V_1 = f_1^T \Sigma_{22}^{-1/2} X^{(2)} \qquad \left(U_k = e_k^T \Sigma_{11}^{-1/2} X^{(1)} \quad V_k = f_k^T \Sigma_{22}^{-1/2} X^{(2)} \right)$

$\text{Corr}(U_1, V_1) = \rho_1^*$ where $\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \geq \rho_p^{*2}$, and $(e_1, e_2, \dots, e_p)$ are eigen values & eigen vectors of $\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} = M_1$

and $\rho_1^{*2} \geq \dots \geq \rho_p^{2}$ and $(f_1, f_2, \dots, f_p)$ are " " " of $\Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2} = M_2$

Pf:-

$U = a' X^{(1)} \qquad V = b' X^{(2)}$

$$\underset{a, b}{\operatorname{argmax}} \ \text{Corr}(U, V) = \frac{a' \Sigma_{12} b}{\sqrt{a' \Sigma_{11} a} \sqrt{b' \Sigma_{22} b}} \to (1)$$

$= (a_1, b_1)$

$V(U_k) = V(V_k) = 1$

$\text{Cov}(U_k, V_l) = 0 \quad k \neq l$

$\text{Cov}(U_k, U_l) = 0 \quad k \neq l$

$\text{Cov}(V_k, V_l) = 0 \quad k \neq l$

$\underbrace{a' \Sigma_{11}^{1/2}}_{a^{*\prime}} \underbrace{\Sigma_{11}^{1/2} a}_{a^*}$

$a^* = \Sigma_{11}^{1/2} a \Rightarrow a = \Sigma_{11}^{-1/2} a^*$

$$(1) = \frac{a^{*\prime} \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} b^*}{\sqrt{a^{*\prime} a^*} \sqrt{b^{*\prime} b^*}}$$

$$(\underbrace{a^{*'} \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}}_{c'} b^*)^2 \le \sqrt{(c'c)(b'b)} \quad \text{(Cauchy Schwartz)}$$

$$\le \sqrt{c'c}$$

$$\frac{(a^{*'} \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} b^*)}{\sqrt{b^{*'} b^*}} \le \sqrt{a^{*'} \underbrace{\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2}}_{M_1} a^*}$$

$$\le \sqrt{a^{*'} M_1 a^*} = \sum \sqrt{\rho_i^{*2} y_i^2}$$

$$\sqrt{\underbrace{a^{*'} P}_{y'} \Lambda P' a^*} \qquad \le \rho_1^* \underbrace{\sqrt{\sum_{i=1}^{n} y_i^2}}_{\sqrt{a^{*'} a^*}}$$

$a^* = e_1$

$\underline{a} = \Sigma_{11}^{-1/2} a^* = \Sigma_{11}^{-1/2} e_1$