WED NOV 8 OFFICE HRS MW 12:30 - 1:30

TODAY LINEAR MODELS

$$y_i = a + b x_i + \varepsilon_i$$

Least squares, Maximum Likelihood, Multivariate linear model

Robust models

AT RANDOM VARIABLE LEVEL: (x, y) jointly distributed random variables

Best linear predictor of Y, $aX + b$

$a =$

$b =$

Q1 HW 6

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

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_5 \end{bmatrix} \qquad Z_1, \ldots, Z_5 \text{ IID } N(0,1)$$

$$\mathrm{cov}(X) = A\,\mathrm{cov}(Z)\,A^T = AA^T$$

$$\mathrm{cov}(Z) = I_{5\times5}$$

$$\underset{5\times1}{X} = \underset{5\times5}{A}\,\underset{5\times1}{Z}$$

$$\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

$$\rightarrow X_{11} \quad X_{12} \quad X_{13} \quad X_{14} \quad X_{15}$$

$$\mathrm{cov}(X.T)$$

$$\underline{X}$$

$$\vdots$$

$$\mathrm{cov}_S(\underline{X})_{i,j} = \mathrm{cov}_S(X_{\cdot i}, X_{\cdot j})$$

5×5

$$= \frac{1}{N} \sum_{\ell=1}^{N} (X_{\ell i} - \bar{X}_{\cdot i}) \times (X_{\ell j} - \bar{X}_{\cdot j})$$

$$\rightarrow X_{N1} \quad X_{N2} \quad X_{N3} \quad X_{N4} \quad X_{N5}$$

THEORY: RANDOM VECTOR $(X_1, \ldots, X_5) = X$

$$\mathrm{cov}(X)_{i,j} = \mathrm{cov}(X_i, X_j)$$

$$\mathrm{cov}(\underset{r\times k}{A} \underset{k\times 1}{Z}) = \underset{r\times k}{A}\, \underset{k\times k}{\mathrm{cov}(Z)}\, \underset{k\times r}{A^T}$$

$r \times r$

$$Z = \begin{matrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \end{matrix}$$

Z

$$\bar{X} \rightarrow \mu \qquad SD(\bar{X}) = \frac{SD(X)}{\sqrt{n}}$$

LINEAR MODELS

(X, Y)

Best linear predictor of Y using X

$aX + b$

$$a = \frac{\mathrm{COR}(X,Y)}{SD(X)} SD(Y) \qquad b = EY - a\,EX$$

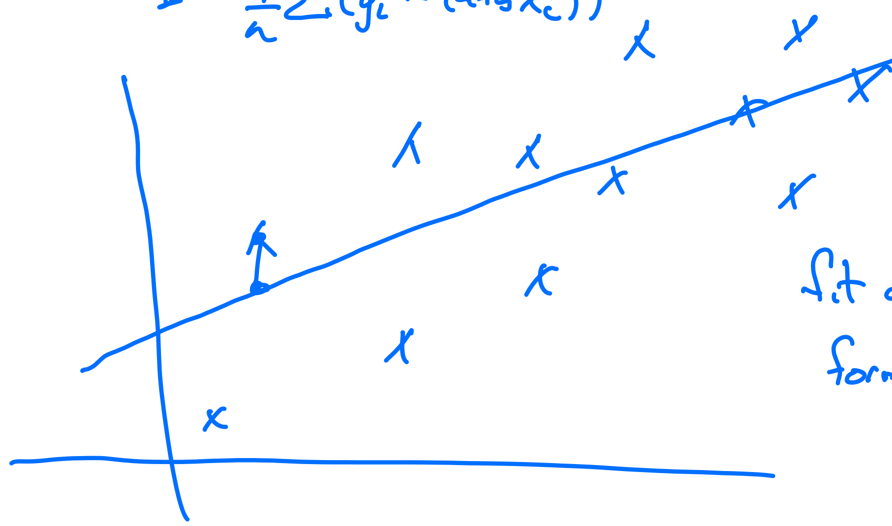
$\uparrow$

Choice of a,b minimizes $E(Y-(aX+b))^2$
MSE

DATA $(x_1,y_1), \ldots, (x_n,y_n)$

minimize$_{a,b}$ $\frac{1}{n}\sum_i (y_i-(a+bx_i))^2$ ← sum of squared errors

$y=a+bx$



fit a line of form $y=a+bx$

$$\hat{a} = \frac{\rho_{xy}}{s_x} s_y \qquad \hat{b} = \bar{y} - \hat{a}\,\bar{x}$$

$$\rho_{xy} = \text{sample correlation} = \frac{\frac{1}{n}\sum_i (x_i-\bar{x})(y_i-\bar{y})}{s_x\, s_y}$$

s_x = sample SD of x
s_y = sample SD of y

Maximum Likelihood

Specific Model assumption

$$Y_i = a + bX_i + \varepsilon_i \qquad \varepsilon_i \sim N(0,\sigma^2)$$

assume $x_1, \ldots, x_n$ are fixed non-random

Unknown parameters: a, b, σ^2 unobserved parameters

Goal: Estimate a,b from data

Maximum Likelihood Estimation

STEP 1: Write down likelihood function (or log-likelihood)

$\to Y_i \sim N(a+bx_i, \sigma^2)$

likelihood of $Y_i \sim \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu_i)^2\right)$

$$\mu_i = a+bx_i$$

log likelihood of $Y_i = -\frac{1}{2\sigma^2}(y_i - (a+bx_i))^2 - \frac{1}{2}\log(2\pi\sigma^2)$

log-likelihood of $(Y_1, \dots, Y_n) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (a+bx_i))^2$

$- n/2 \log(2\pi\sigma^2)$

STEP 2

Find parameter values a, b, σ^2 maximizing this function (σ fixed)

maximizing likelihood $\longleftrightarrow$ minimizing

$$\sum_{i=1}^{n}(y_i - (a+bx_i))^2$$

a, b minimizing are least squares $\hat{a}, \hat{b}$

MORAL If we assume model is Normal, MLE estimates of a, b are exactly the least-squares estimates

---

Univariate model

$$y_i = a + bx_i + \varepsilon_i$$

DATA

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \begin{bmatrix} x_1 & z_1 \\ x_2 & z_2 \\ \vdots & \\ x_n & z_n \end{bmatrix}$$

Multivariate MODEL

$$y_i = a + bx_i + cz_i + \varepsilon_i \qquad i = 1, 2, \dots, n$$

$$\underset{n\times 1}{Y} = \underset{n\times k}{X}\underset{k\times 1}{\beta} + \underset{n\times 1}{\varepsilon} \qquad \mathrm{cov}(\varepsilon) = \sigma^2 I$$

$$X = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ 1 & \vdots & \vdots \\ 1 & \vdots & \vdots \\ 1 & x_n & z_n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Goal: Estimate $\underset{k \times 1}{\beta}$ ← parameters

$\|Y - X\beta\|^2$ ← pick β to minimize

$$= \sum_{i=1}^{n} (y_i - (X\beta)_i)^2$$

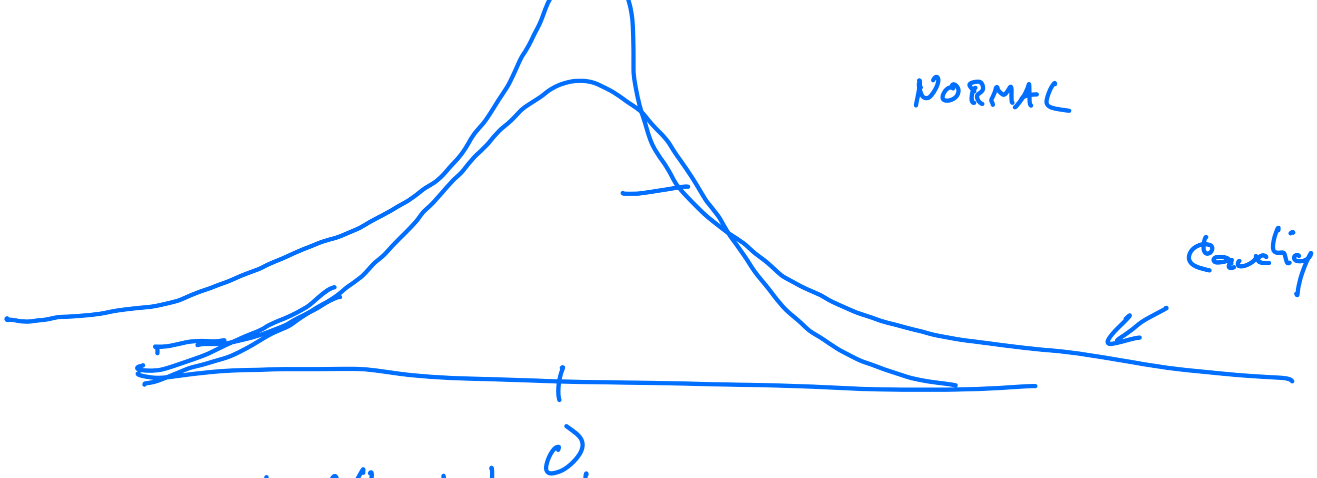
DIFFERENTIATE WRT β, SET $= 0$

MATRIX EQN

$$\underbrace{\underset{k \times n}{(X^T} \underset{n \times k}{X)}} \underset{k \times 1}{\beta} = \underset{k \times n}{X^T} \underset{n \times 1}{y}$$ NORMAL EQUATIONS

$$\hat{\beta} = (X^T X)^{-1} X^T y$$ IS SOL'N

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



"Fat tailed" distribution

If errors are Not Normal, L.S. estimates may perform badly

estimated values far from truth!

SOL'N: USE MLE with Cauchy density