LECTURE 37 : MINIMUM MEAN SQUARE ERROR ESTIMATION PART 2

- TOPICS TO COVER (BASED ON CH 4.9)

→ LINEAR ESTIMATORS

→ LINEAR ESTIMATORS

RESPONSE → Y : A RANDOM VARIABLE WITH KNOWN PROBABILITY DISTRIBUTION F_Y

FEATURE → X : A RANDOM VARIABLE WITH KNOWN PROBABILITY DISTRIBUTION F_X

QUESTION : SUPPOSE Y IS NOT OBSERVED BUT X IS AND WE WISH TO ESTIMATE Y BASED ON AN OBSERVATION FROM X SUCH THAT THE MEAN SQUARE(D) ERROR IN ESTIMATION IS MIMIMIZED.

ANSWER :

LET $g(X)$ BE A FUNCTION OF X THAT WE WISH TO USE TO ESTIMATE THE RV Y. THEN

ERROR IN ESTIMATION $= Y - g(X)$

SQUARE(D) ERROR IN ESTIMATION $= (Y - g(X))^2$

MEAN SQUARE(D) ERROR IN ESTIMATION $= E(Y - g(X))^2$

MEAN ← EXPECTED / AVERAGE

WE WANT TO CHOOSE g SUCH THAT THE MSE IS MIMIMIZED, I.E., SOLVE

$$g^*(X) = \arg\min_{g(X)} E(Y - g(X))^2$$

WE HAVE ALREADY SHOWN THAT (LECTURE 36) $g^*(X) = E(Y \mid X)$

FOR SIMPLICITY AND REASONS BEYOND ARE SCOPE, WE FOCUS ON FUNCTIONS g

THAT ARE LINEAR, I.E.,

$$\underbrace{g(X)}_{\underbrace{L(X)}_{\text{ANOTHER NOTATION}}} = aX + b$$

LINEAR $\rightarrow L(X)$

a, b: UNKNOWN

FINDING $g(X)$ AMONG LINEAR FUNCTIONS SIMPLIFIES THE PROBLEM AS WE ONLY NEED TO FIND THE UNKNOWN a AND b.

MEAN SQUARE(D) ERROR IN LINEAR ESTIMATION $= E(Y - L(X))^2$

$= E(Y - (aX + b))^2$

WE WANT TO CHOOSE a AND b SUCH THAT THE MSE IS MIMIMIZED,

I.E., SOLVE

$$a^*, b^* = \arg\min_{a,b} E(Y - (aX + b))^2$$

HOW?

→ DIFFERENTIATE $E(Y - (aX + b))^2$ W.R.T. a AND b AND

→ SET EACH OF THE ABOVE TWO EQUAL TO ZERO

: $\underbrace{\text{TWO EQUATIONS}}_{}$ AND TWO UNKNOWNS

$\leftarrow$ NORMAL EQUATIONS

BUT, WE WILL TAKE A DIFFERENT APPROACH.

CONSTANT ESTIMATOR

CONSIDER $$E(Y-(aX+b))^2 = E(\underbrace{(Y-aX)}_{\text{NEW RV}}-b)^2 \quad \cdots (1)$$

MINIMIZING (1) W.R.T. b FOR A FIXED a IS SAME AS FINDING THE MINIMUM MSE CONSTANT ESTIMATOR OF $Y-aX$, WHICH IS GIVEN BY

$$b^* = E(Y-aX) = EY - aEX = \mu_Y - a\mu_X$$

HENCE FOR A FIXED a, THE OPTIMAL LINEAR ESTIMATOR IS

$$L_{b^*}(X) = aX + b^*$$
$$= aX + \mu_Y - a\mu_X$$
$$= \mu_Y + a(X-\mu_X)$$

HENCE FOR A FIXED a, MSE AT $L_{b^*}(X)$ IS

$$E(Y-(aX+b^*))^2 = E(Y-\mu_Y-a(X-\mu_X))^2$$
$$= E(\underbrace{Y-aX}_{\text{NEW RV}} - \underbrace{(\mu_Y - a\mu_X)}_{E(\text{NEW RV})})^2$$
$$= Var(Y-aX)$$

LECTURE 34

$$= Cov(Y-aX, Y-aX)$$

LECTURE 34

$$= \text{Cov}(Y, Y) - 2a\,\text{Cov}(Y, X) + a^2\,\text{Cov}(X, X)$$

$$= \text{Var}(Y) - 2a\,\text{Cov}(Y, X) + a^2\,\text{Var}(X) \quad \dots (2)$$

WE NEED TO FIND a SUCH THAT (2) : MSE AT $L_{b^*}(X)$ IS MINIMIZED, I.E.

SOLVE:

$$a^* = \underset{a}{\arg\min}\; E(Y - aX + b^*)^2$$

HOW?

→ DIFFERENTIATE $E(Y - aX + b^*)^2$ W.R.T. a AND

→ SET IT EQUAL TO ZERO : ONE EQUATION AND ONE UNKNOWN

CONSIDER

$$\frac{d}{da} E(Y - aX + b^*)^2 = \frac{d}{da}\left(\text{Var}(Y) - 2a\,\text{Cov}(Y, X) + a^2\,\text{Var}(X)\right) = 0$$

$$= -2\,\text{Cov}(Y, X) + 2a\,\text{Var}(X) = 0$$

UNOBSERVED RESPONSE

REGRESSION COEFFICIENT OF Y ON X ← OBSERVED FEATURE

$$\Rightarrow \quad a^* = \frac{\text{Cov}(Y, X)}{\underbrace{\text{Var}(X)}}$$

σ_X^2 : NOTATION FOR $\text{Var}(X)$

CORRELATION COEFFICIENT BETWEEN X AND Y

RECALL THAT $\rho_{X,Y} = \dfrac{\text{Cov}(Y, X)}{\sigma_X \sigma_Y}$

$$\Rightarrow \quad a^* = \frac{\rho_{X,Y}\,\sigma_X\,\sigma_Y}{\sigma_X^2} = \frac{\rho_{X,Y}\,\sigma_Y}{\sigma_X}$$

$\therefore$ MINIMUM MSE LINEAR ESTIMATOR :

$$L^*(X) = a^* X + b^*$$

$$a^* = \frac{\rho_{X,Y}\,\sigma_Y}{\sigma_X} \qquad b^* = \mu_Y - a^* \mu_X$$

$$L^*(X) = \mu_Y + \rho_{X,Y}\,\sigma_Y \left(\frac{X - \mu_X}{\sigma_X} \right)$$

$\therefore$ MINIMUM MSE AT $L^*(X)$ USING (2) IS

$$E\left(Y - a^* X + b^*\right)^2 = Var(Y) - 2a^* Cov(Y, X) + a^{*2} Var(X) \quad \cdots (2)$$

$$= \sigma_Y^2 - 2\frac{\rho_{X,Y}\,\sigma_Y}{\sigma_X} Cov(Y, X) + \frac{\rho_{X,Y}^2\,\sigma_Y^2}{\sigma_X^2}\sigma_X^2$$

$$= \sigma_Y^2 - 2\frac{\rho_{X,Y}\,\sigma_Y}{\sigma_X} \rho_{X,Y}\,\sigma_X\,\sigma_Y + \rho_{X,Y}^2\,\sigma_Y^2$$

$$= \sigma_Y^2 - 2\rho_{X,Y}^2\,\sigma_Y^2 + \rho_{X,Y}^2\,\sigma_Y^2$$

$$= \underbrace{\sigma_Y^2}_{\text{MINIMUM MSE OF CONSTANT EST.}} - \underbrace{\rho_{X,Y}^2\,\sigma_Y^2}_{\text{REDUCTION IN MINIMUM MSE DUE TO } X}$$

IF X HAS NO 'RELATION' (LINEAR) WITH Y, THEN $\rho_{X,Y} = 0$

$$= \sigma_Y^2 \left(1 - \rho_{X,Y}^2\right)$$

SUMMARY OF MINIMUM MSE ESTIMATION :

- UNCONSTRAINED ESTIMATORS OF THE FORM : $g(X)$

  → MIMIMUM MSE UNCONSTRAINED ESTIMATOR $g^*(X) = E(Y \mid X)$

  → MSE OF $g^*(X) = E(Y^2) - E[(E(Y \mid X))^2]$

A SPECIAL CASE OF UNCONSTRAINED ESTIMATORS

- LINEAR ESTIMATORS OF THE FORM : $L(X) = aX + b$

  → MIMIMUM MSE LINEAR ESTIMATOR $L^*(X) = \mu_Y + \rho_{X,Y}\,\sigma_Y \left(\dfrac{X - \mu_X}{\sigma_X} \right)$

  → MSE OF $L^*(X) = \sigma_Y^2 (1 - \rho_{X,Y}^2)$

A SPECIAL CASE OF LINEAR ESTIMATORS $L(X) = aX + b$, WHERE $a = 0$ AND $b = \delta$

- CONSTANT ESTIMATORS OF THE FORM : δ

  → MIMIMUM MSE CONSTANT ESTIMATOR $\delta^* = \mu_Y$

  → MSE OF $\delta^* = \sigma_Y^2$

NOTE THAT

$$\underbrace{E(Y^2) - E[(E(Y \mid X))^2]}_{\text{MSE OF } g^*(X)} \le \underbrace{\sigma_Y^2 (1 - \rho_{X,Y}^2)}_{\text{MSE OF } L^*(X)} \le \underbrace{\sigma_Y^2}_{\text{MSE OF } \delta^*}$$