Date :
Page No.

Ans 1: $\theta = (X^T X)^{-1} X^T Y$

Let a be a non-zero vector.

~~$\Rightarrow a^T (X^T X) a$~~
~~$\Rightarrow (Xa)^T$~~

$\Rightarrow a^T (X^T X) a$
$\Rightarrow (Xa)^T (Xa)$
$\Rightarrow \|Xa\|^2 \geq 0$

$X^T X$ is semi-positive definite.

\# It is because:
$b^T M b \geq 0$; then M is semi-positive definite.

Now let $X_{m \times n}$.

Now, $Xa = 0$; only when X is linearly independent as ~~...~~ a is non-zero vector.

~~Now, as given X is linearly independent,~~
~~$\|Xa\| > 0$~~

So, $\|Xa\| > 0$, when ~~X~~ columns of X is independent.

Claim: Every positive definite ~~...~~ matrix is ~~...~~ invertible.

From the above claim, ~~X~~ $X^T X$ is invertible as it is positive definite.

~~Thus, proving that columns of X~~

Thus, $X^T X$ is invertible when columns ~~...~~ of X are linearly independent.

Ans2: $\hat{Y} = X\hat{\theta}$

$Y = X\theta + \varepsilon$

$Y - \hat{Y} = \hat{\varepsilon}$

From question:

$$\hat{\sigma}^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{N-M}$$

$$= \frac{E[\hat{\varepsilon}^T \hat{\varepsilon}]}{N-M}$$

$$\hat{\varepsilon} = Y - \hat{Y}$$
$$= Y - X\hat{\theta}$$
$$= Y - X(X^TX)^{-1}X^TY \quad ; \quad \hat{\theta} = (X^TX)^{-1}X^TY$$
$$= [I - X(X^TX)^{-1}X^T]Y$$

let $I - X(X^TX)^{-1}X^T = P$.

$$\hat{\varepsilon} = PY.$$
$$= (I - X(X^TX)^{-1}X^T)(X\theta + \varepsilon)$$
$$= X\theta + \varepsilon - X(X^TX)^{-1}X^TX\theta + X(X^TX)^{-1}X^T\varepsilon$$
$$= [I - X(X^TX)^{-1}X^T]\varepsilon$$
$$\hat{\varepsilon} = P\varepsilon$$

$$E[\hat{\varepsilon}^T\hat{\varepsilon}] = (P\varepsilon)^T(P\varepsilon)$$
$$= \varepsilon^T P^T P \varepsilon$$

Date :

Page No.

$$P^T P = [I - X(X^T X)^{-1} X^T][I - X(X^T X)^{-1} X^T]$$
$$= P$$

$$\cancel{E[\varepsilon^T} \; E[\hat{\varepsilon}^T \hat{\varepsilon}] = \varepsilon^T P \varepsilon^T$$

$$E[\hat{\varepsilon}^T \hat{\varepsilon}] = E[\varepsilon^T P \varepsilon^T]$$
$$= E[Tr(\varepsilon^T P \varepsilon^T)] \quad (1)$$

$$\cancel{Tr(\varepsilon^T P \varepsilon^T) = \varepsilon^T P \varepsilon^T = Tr(}$$

If A is semi positive definite matrix:

$$Tr(X^T A X) = X^T A X = \cancel{Tr(AX^TX)} \; Tr(A X X^T)$$

From 1;

$$E[\hat{\varepsilon}^T \hat{\varepsilon}] = E[\varepsilon^T P \varepsilon^T]$$, As P is semipositive definite
$$= E[Tr(\varepsilon^T M \varepsilon)]$$
$$= E[Tr(M \varepsilon \varepsilon^T)]$$
$$= [Tr(M)] . E[\varepsilon \varepsilon^T]$$

$$Tr(M) = Tr[I_{N \times N}] - Tr[X[X^T X]^{-1} X^T]$$

$$Tr[X[X^T X]^{-1} X^T] = Tr[(X^T X)^{-1} X X^T]$$
$$= Tr(I_{M \times M})$$
$$= M$$

$$Tr(M) = N - M.$$

Date :

Page No.

~~Ans~~ $E[\hat{\varepsilon}^T \hat{\varepsilon}] = \cancel{(N-M)}\ (N-M)\, E[\varepsilon \varepsilon^T]$

$\cancel{E[\sigma^2]}\ E[\varepsilon \varepsilon^T] = \sigma^2 I$

$= \sigma^2$

$\cancel{E[\hat{\sigma}^2]} = E[\hat{\sigma}^2] = \sigma^2$

Hence proved.

Date :
Page No.

Ans 3: So, the 5 points form ~~X~~:

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

$$Y = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 3 \\ 2 \end{bmatrix}$$

So, the best fit is when

$$J = [2 - (\theta_1 + \theta_2)]^2 + [1 - (\theta_1 + 2\theta_2)]^2 + [2 - (\theta_1 + 3\theta_2)]^2 + [3 - (\theta_1 + 3\theta_2)]^2 + [2 - (\theta_1 + 4\theta_2)]^2$$

$$= (4 + 1 + 4 + 9 + 4) - 2[2(\theta_1 + \theta_2)] - 2[(\theta_1 + 2\theta_2)] - 2[2(\theta_1 + 3\theta_2)] - 2[3(\theta_1 + 3\theta_2)] - 2[2(\theta_1 + 4\theta_2)] + (\theta_1 + \theta_2)^2 + (\theta_1 + 2\theta_2)^2 + (\theta_1 + 3\theta_2)^2 + (\theta_1 + 3\theta_2)^2 + (\theta_1 + 4\theta_2)^2$$

$$\frac{dJ}{d\theta} = 0$$

We will choose that θ that gives us best fit.

Date :

Page No.

~~$\frac{dJ}{d\theta_1} = 0$~~

$$22 - 4[\theta_1 + \theta_2] - 2[\theta_1 + 2\theta_2] - 4[\theta_1 + 3\theta_2)$$
$$-6[\theta_1 + 3\theta_2) - 4[\theta_1 + 4\theta_2) + (\theta_1^2 + 2\theta_1\theta_2 + \theta_2^2)$$
$$+(\theta_1^2 + 4\theta_1\theta_2 + 4\theta_2^2) + 2(\theta_1^2 + 6\theta_1\theta_2 + 9\theta_2^2)$$
$$+(\theta_1^2 + 8\theta_1\theta_2 + \cancel{\theta_2}\, 16\theta_2^2) = 0$$

$\frac{dJ}{d\theta_1} = 0$

$$\Rightarrow -4 - 2 - 4 - 6 - 4 + 2\theta_1 + 2\theta_2$$
$$+ 2\theta_1 + 4\theta_2 + \cancel{2}\, 4\theta_1 + 12\theta_2 + 2\theta_1 + 8\theta_2 = 0$$

$$\Rightarrow -20 + 10\theta_1 + 26\theta_2 = 0$$
$$\Rightarrow 5\theta_1 + 13\theta_2 = 10$$

$\frac{dJ}{d\theta_2} = 0$

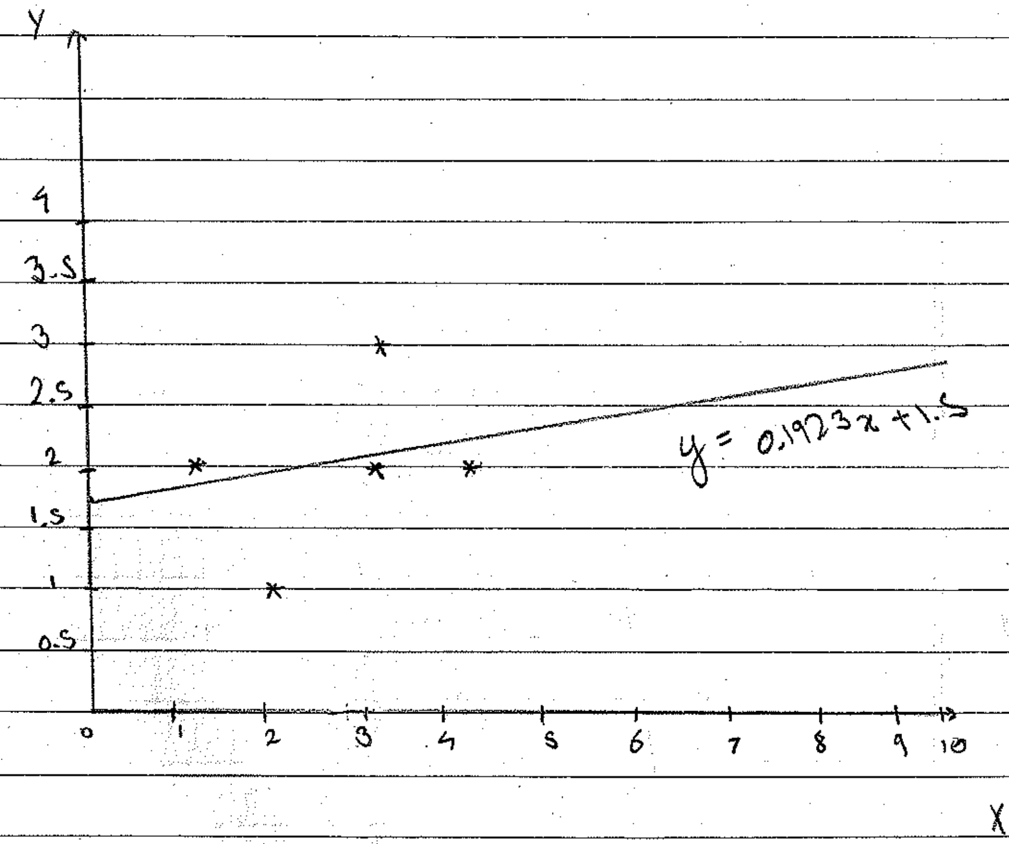
$$\Rightarrow -4 - 4 - 12 - 18 - 16 + \cancel{2\theta_2}\,2\theta_1 + 2\theta_2$$
$$+ 4\theta_1 + 8\theta_2 + 12\theta_1 + \cancel{18}\, 36\theta_2 +$$
$$8\theta_1 + 32\theta_2 = 0$$

-54

$$\Rightarrow -54 + 26\theta_1 + 78\theta_2 = 0$$
$$\Rightarrow 13\theta_1 + 39\theta_2 = 27$$

Date :
Page No.

So, $\theta_1 = 1.5$
$\theta_2 = 0.1923$



~~Plot with~~ $\theta = (X^T X)^{-1} X^T Y$ also gives θ same θ.
Plot attached in report (drawn from ~~math~~ matlab).