Solutions

1) a) $X = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$

b) $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ Design Matrix

c) $Y = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 7 \end{bmatrix}$

d) $\hat{\beta} = (A^T A)^{-1} (A^T Y)$

$\beta_0 = 0.75, \ \beta_1 = 2.5, \ \beta_2 = 3.5$

e) $\beta_0, \beta_1, \beta_2 \rightarrow$ Computed by solving eqns in numpy.

f) $RSS = \sum_{i=1}^{4} (X\beta - Y)^2 = 0.25$

$TSS = RSS + ESS \Rightarrow 18.5$

g) TSS ✓ } Using numpy program. $= 18.75$

h) R^2 ✓ } $R^2 = \frac{TSS - RSS}{TSS} = 0.98666...$

- What portion of variance in y explained by x?? R^2

- Predict value of $x = (0.5, 0.5) = \beta_0 + 0.5\beta_1 + 0.5\beta_2$

$= 3.75$ ✓

Ans 2) Can be done easily on programming assignment.

i) $RSS = \sum_{i=1}^{n} (X\beta - Y)^2 = 14517.551760...$

ii) $TSS = \sum_{i=1}^{n} (Y - \bar{Y})^2 = 42716.295415 0...$

iii) $R^2 = \frac{TSS - RSS}{TSS} = 0.66014019...$

iv) What portion of variation in Y explained by X?

- Explained by R^2 i.e. 0.6601 or 66.01 %.

Ans 3) Can have many answers.

Features → Amount of Rainfall, Amount of fertilizer, average temperature, # sunny days.

Output Variable Y → Crop yield in Kg (or any unit of measure)
Or
Crop failed in Kg (or any unit of measure).

Ans 4) i) $f_1(b_1, b_2) = (b_1 - 4)^2 + (b_2 + 3)^2$

Python Code:-

```
b1 = 0
b2 = 0
for i in range(4):
    temp1 = b1 - 0.5 * 2 * (b1 - 4)
    temp2 = b2 - 0.5 * 2 * (b2 + 3)
    b1 = temp1
    b2 = temp2
    print(b1, b2)
```

$\frac{df_1(b_1, b_2)}{db_1}$

Output

```
4.0, -3.0
4.0, -3.0
4.0, -3.0
4.0, -3.0
```

Minimizing the function in the first iteration itself!!

ii) $f_2(b_1, b_2) = (4-b_1)^2 + 34\left((b_1+4)-(b_2-4)^2\right)^2$

$$\frac{df_2(b_1,b_2)}{db_1} = 2(4-b_1) + 68\left[(b_1+4)-(b_2-4)^2\right]$$

$$\frac{df_2(b_1,b_2)}{db_2} = 0 + 68\left((b_1+4)-(b_2-4)^2\right) \times (0 - 2(b_2-4))$$

$$= 68\left((b_1+4)-(b_2-4)^2\right)(-2b_2+8)$$

Python Code:-

$b_1, b_2 = 0, 0$

Derivative of f_2 w.r.t b_1 → $db_1 = 2*(4-b_1) + 68\left((b_1+4)-(b_2-4)*(b_2-4)\right)$

Derivative of f_2 w.r.t b_2 → $db_2 = 68*\left((b_1+4)-(b_2-4)*(b_2-4)\right)*(-2b_2+8)$

for i in range(4):

$temp1 = b_1 - 0.5*db_1$

$temp_2 = b_2 - 0.5 * db_2$

$b_1 = temp_1$

$b_2 = temp_2$

print(b_1, b_2)

Output :

404.0, 3264.0

808.0, 6528.0

1212.0, 9792.0

1616.0, 13056.0