Assignment - 1
Written Part - Solutions

Ans1) a) This is a <u>Regression</u> problem and we are interested in <u>Inference</u>. Here, number of samples $(n) = 500$ & Number of predictors$(p) = 3$.

b) <u>Classification Problem</u>. We are interested in <u>Prediction</u>. $n = 20$, $p = 13$.

c) <u>Regression Problem</u>. We are interested in <u>Prediction</u>.

$n = 52$ (i.e. No. of weeks in year 2012) & $p = 3$

Ans.2) This question can have many different answers

a) GPA, time taken to graduate etc.

b) GPA (continuous). The answer depends on the type of variable we are considering in part a).

c) High-School GPA, SAT and ACT score etc.

d) College GPA v/s High School GPA can be modeled linearly. Also, slope would be positive because target variable $\propto$ predictor.

Ans 3) a) The sample means are :-

$$\bar{x} = \frac{1}{N} \sum x_i = \frac{10}{5} = 2$$

$$\bar{Y} = \frac{1}{N} \sum y_i = \frac{30}{5} = 6$$

b) $$S_x^2 = \frac{1}{N} \sum_i (x_i - \bar{x})^2 = 2$$

$$S_y^2 = \frac{1}{N} \sum_i (y_i - \bar{Y})^2 = 37.2$$

$$S_{xy} = \frac{1}{N} \sum_i (x_i - \bar{x})(y_i - \bar{Y}) = 8$$

c) $$\beta_1 = \frac{S_{xy}}{S_x^2} = 4$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{x} = -2$$

d) $\hat{y} = \beta_0 + \beta_1 x$

$\hat{y}$ at $x = 2.5$ is :

$$\hat{y} = -2 + 4(2.5) = 8$$

e) MSE (Mean Squared Error) $= \frac{1}{N} RSS$

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

$$= \frac{1}{N} \sum_{i=1}^{N} (\beta_0 + \beta_1 x_i - y_i)^2$$

$$= \frac{4+0+9+4+9}{5} = \frac{26}{5} = 5.2$$

f) R^2 (Confidence of Determination)

$$TSS = \sum_{i=1}^{N} (y_i - \bar{y})^2 = 186, \quad RSS = \sum_{i=1}^{N} (y_i - \hat{y})^2 = 26$$

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{26}{186} = 0.86$$

g) For the first change, there would be slight change in β_0 & β_1.

For the second change, there would be drastic change in the values.

Ans 4) a) $z(t) = z_0 e^{-\alpha t}$

Taking log both sides

$$\Rightarrow \ln(z(t)) = \ln(z_0 e^{-\alpha t})$$

Using $\ln(AB) = \ln A + \ln B$

& $\ln e^{x} = x$

$$\Rightarrow \ln(z(t)) = \ln z_0 - \alpha t$$

Let $\ln(z(t)) = y$, $\ln z_0 = \beta_0$ & $\beta_1 = -\alpha$

$$\Rightarrow \boxed{y = \beta_0 + \beta_1 t}$$

b) To compute least Square Solutions:-

$$Y_i = \ln(z(t_i)) \quad , \quad x_i = t_i$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad , \quad \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$S_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad , \quad S_Y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2$$

$$S_{xY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{Y})$$

$$\beta_1 = \frac{S_{xy}}{S_x^2} \quad \& \quad \beta_0 = \bar{Y} - \beta_1 \bar{x}$$

Substituting β_1 & β_0 in following equation to get α & Z_0.

$$\beta_1 = -\alpha \quad , \quad \beta_0 = \ln(Z_0)$$

c) <u>Python Code:-</u>

```
import numpy as np

x = t
y = np.log(Z)

xm = np.mean(x)

ym = np.mean(y)

Sxx = np.mean((x-xm)**2)

Sxy = np.mean((x-xm) * (y-ym))

beta1 = Sxy
        ---
        Sxx

beta0 = ym - (b1 * xm)

alpha = -beta1

Z0 = np.exp(beta0)
```

Ans 5) a) $RSS = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{N} (\beta x_i - y_i)^2$

b) $\frac{\partial(RSS)}{\partial \beta} = \frac{\partial \sum_{i=1}^{N} (\beta x_i - y_i)^2}{\partial \beta}$

$$\frac{\partial(RSS)}{\partial \beta} = 2 \sum_{i=1}^{N} (\beta x_i - y_i)\, x_i$$

For minima, Put $\frac{\partial(RSS)}{\partial \beta} = 0$

$$\Rightarrow 2 \sum_{i=1}^{N} (\beta x_i - y_i)\, x_i = 0$$

or $\sum_{i=1}^{N} (\beta x_i^2 - x_i y_i) = 0$

or $\boxed{\beta = \frac{\sum_{i=1}^{N} (x_i y_i)}{\sum_{i=1}^{N} (x_i^2)}}$