task # Linear Regression.

Syeda Zainab Fatima
W25-PAK-INP-AI-01

Linear Regression:-

1 Problem Setup:-

It seeks to find the best-fit line that minimizes the difference between the observed (y_i) and the predicted values ($\hat{y}_i$).

Model Equation:-

$$y_i = \beta_0 + \beta_i x_i + \epsilon_i.$$

y_i = actual values
x_i = Independent variable
β_0 = Intercept (value of y when $x=0$)
β_1 = Slope (rate of change of y per unit change in x)
ϵ_i = Residual error.

2 Objective: Minimize the error:-

Sum of Squared Errors (SSE).

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

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

$$SSE = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2.$$

3 Optimal Coefficients (β_0, β_1):-

1 Expand the SSE

$$SSE = \sum_{i=1}^{n} \left(y_i^2 - 2y_i(\beta_0 + \beta_1 x_i) + (\beta_0 + \beta_1 x_i)^2 \right)$$

Expanding $(\beta_0 + \beta_1 x_i)^2$;

$$SSE = \sum_{i=1}^{n} (y_i^2 - 2y_i\beta_0 - 2y_i\beta_1 x_i + \beta_0^2 + 2\beta_0\beta_1 x_i + \beta_1^2 x_i^2).$$

2. Partial Derivatives:-

→ with respect to β_0 :-

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

Set $\frac{\partial SSE}{\partial \beta_0} = 0$:-

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i$$

$$\beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n}$$

→ with respect to β_1 :-

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

Set $\frac{\partial SSE}{\partial \beta_1} = 0$:-

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2$$

3. Solve for β_1 :-

Substitute β_0 into the equation for β_1, after simplification:-

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\bar{x} = \frac{\sum x_i}{n}$: mean of x

$\bar{y} = \frac{\sum y_i}{n}$: mean of y.

4) Solve for β_0

Substitute β_1 back into:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Final Equation:-

$$\boxed{\hat{y} = \beta_0 + \beta_1 x}$$