AI

Linear Regression:

→ $Y = mX + b + \epsilon$ Simple LR

→ $Y = m_1X_1 + m_2X_2 + m_nX_n + b + \epsilon$ Multiple LR

- $\beta = \begin{bmatrix} b \\ m_1 \\ m_2 \\ m_n \end{bmatrix} = (X^TX)^{-1}X^TY$
- X: Metrics of independent variables, Y: Vector of dependent variables.

  $SSE = \Sigma(Y_i - \hat{Y}_i)^2$: Sum of squared errors,

  ↳ Y_i = Actual Value, $\hat{Y}_i$ = Predicted Value.

→ Assumptions of Linear Regression:

1. Linearity: Relationship b/w dependent & independent variable is linear.
   ↳ Check: Scatter or Residual plot.
2. Independence: Observations should be independent (no autocorrelation).
   ↳ Check: Durbin-Watson test.
3. Homoscedasticity: The variance of residuals should be constant across all levels of independent variable.
   ↳ Residual vs Predicted values plot.
4. Residuals' Normality: Residuals should be normally distributed.
   ↳ Histogram of residuals or Q-Q plot.
5. No Multicolinearity: Independent variables should not be highly correlated.
   ↳ Check: Variance Inflation Factor (VIF)

→ Model Evaluation:

i) R-Squared : $R^2 = 1 - \left(\frac{SS_{res}}{SS_{tot}}\right)$

↳ measures how much variation in Y is explained by X.

↳ values b/w 0 & 1 (higher is better)

ii) Adjuste R-Squared: $R^2_{adj} = 1 - \left(\frac{(1-R^2)(n-1)}{n-p-1}\right)$

↳ Adjusted for number of predictors.

↳ Usefull when dealing with multiple regression.

iii) Mean Squared Error (MSE): $MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$

↳ Average squared difference b/w y_i & $\hat{y}_i$.

↳ Lower the better.

iv) Root MSE : $RMSE = \sqrt{MSE}$

↳ Give MSE in actual units.

v) Mean Absolute Error : $MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$

↳ Measures absolute (average) differece b/w y_i & $\hat{y}_i$.

→ Limitations:

- Sensitive to outliers.
- Highly non-linear.
- Not ideal for complex relationships.