Nov-12:

x	y	$\hat{y}$
1	6	5
3	23	19
9	60	61
10	67	68

$\hat{y} \to 7x-2.$

$\hat{y} = 7x - 2.$

Equation given by SLR model.

Evaluation Metrics of Regression:

1. R2 ~~Square~~ score (or) R square value (R^2 value) - how good model is? (higher the better)
2. MSE (Mean Square Error)
3. MAE (Mean Absolute Error)
4. RMSE (Root Mean Square Error)
5. MAPE (Mean Absolute Percentage Error)

- how bad model is? (Lower the better)

1. R2 score / R^2 value:

$$\boxed{R^2 = 1 - \frac{LR}{AVG}}$$

where
LR: The total error of predicted model.
AVG: ~~Average~~ The total error of ~~actual~~ average model.

Total error of predicted model (LR) $= \sum_{i=1}^{n} (y_i - \hat{y})^2$

Total error of average ~~actual~~ model (AVG) $= \sum_{i=1}^{n} (\bar{y_i} - y)^2$

$LR = (6-5)^2 + (23-19)^2 + (60-61)^2 + (67-68)^2$

$= 1 + 16 + 1 + 1$

$\boxed{LR = 19}$

$\bar{y}_i = \frac{6+23+60+67}{4}$

$= \frac{156}{4}$

$\bar{y}_i = 39$

$AVG = (39-6)^2 + (39-23)^2 + (39-60)^2 + (39-67)^2$

$= (33)^2 + (16)^2 + (-21)^2 + (-28)^2$

$= 1089 + 256 + 441 + 784$

$\boxed{AVG = 2570}$

$\boxed{\text{Error} = \text{Actual} - \text{predicted}}$

$$R^2 = 1 - \frac{LR}{AVG}$$

$$= 1 - \frac{19}{2570}$$

$$= 1 - 0.0073$$

$$\boxed{R^2 = 0.9926}$$

R^2 values ranges from 0 to 1

if R^2 value is near to 1 it is best model.

if R^2 value is near to 0 it is worst model.

Here, R^2 value is near to 1 so the model is considered as best model.

Steps we follow in machine learning:

1. Data Analysis (pandas, numpy, matplot)
2. Data validation } sk learn (scikit)
3. Model building

- Build a model – Algorithm name
- Train a model – x, y
- Test a model – x
- Evaluate a model – result analysis.

2. Mean Square Error (MSE):

MSE tells on avg how many units or

MSE measure how far predictions are from actual values.

$\hat{y} = 7x - 2$

x	y	$\hat{y}$
1	6	5
3	23	19
9	60	61
10	67	66

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

$$= \frac{1}{4}\left[(6-5)^2 + (23-19)^2 + (60-61)^2 + (67-66)^2\right]$$

$$= \frac{1}{4}(19)$$

$$\boxed{MSE = 4.75}$$

MSE ranges from 0 to ∞, the lower the value the better the model is.

Here, MSE is less hence the model is good & have good accuracy.

3. Mean Absolute Error (MAE): MAE tells on avg how many units the model is differing.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$= \frac{1}{4}\left[|6-5| + |23-19| + |60-61| + |67-66|\right]$$

$$= \frac{1}{4}[1+4+1+1]$$

$$= \frac{1}{4}(7)$$

$$\boxed{MAE = 1.75}$$

MAE ranges from 0 to ∞

MAE value near to ∞ means greater error & vice versa.

Here, MAE value is less hence model is good.

4. Root Mean Square Error: It indicates error in actual units of y

&.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$RMSE = \sqrt{4.75}$

$\boxed{RMSE = 2.179}$

Larger RMSE means worst accuracy.

Here, RMSE is smaller hence model is having good accuracy

5. MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y - \hat{y}}{y} \right|$$

$$= \frac{100}{4} \left[\left| \frac{6-5}{6} \right| + \left| \frac{23-19}{23} \right| + \left| \frac{60-61}{60} \right| + \left| \frac{67-66}{67} \right| \right]$$

$$= \frac{100}{4} \left[\frac{1}{6} + \frac{4}{23} + \frac{1}{60} + \frac{1}{67} \right]$$

$$= \frac{100}{4} \left[0.1667 + 0.1739 + 0.0167 + 0.0149 \right]$$

$$= \frac{100}{4} \left[0.3722 \right]$$

$\boxed{MAPE = 9.30\%}$

MAPE give percentage error

If % error

0% - perfect fit

0-10% - Highly accurate

10-20% - Good

20-50% - Reasonable

> 50% - poor

Here our MAPE value lies in 0-10%. Hence the model is highly accurate.