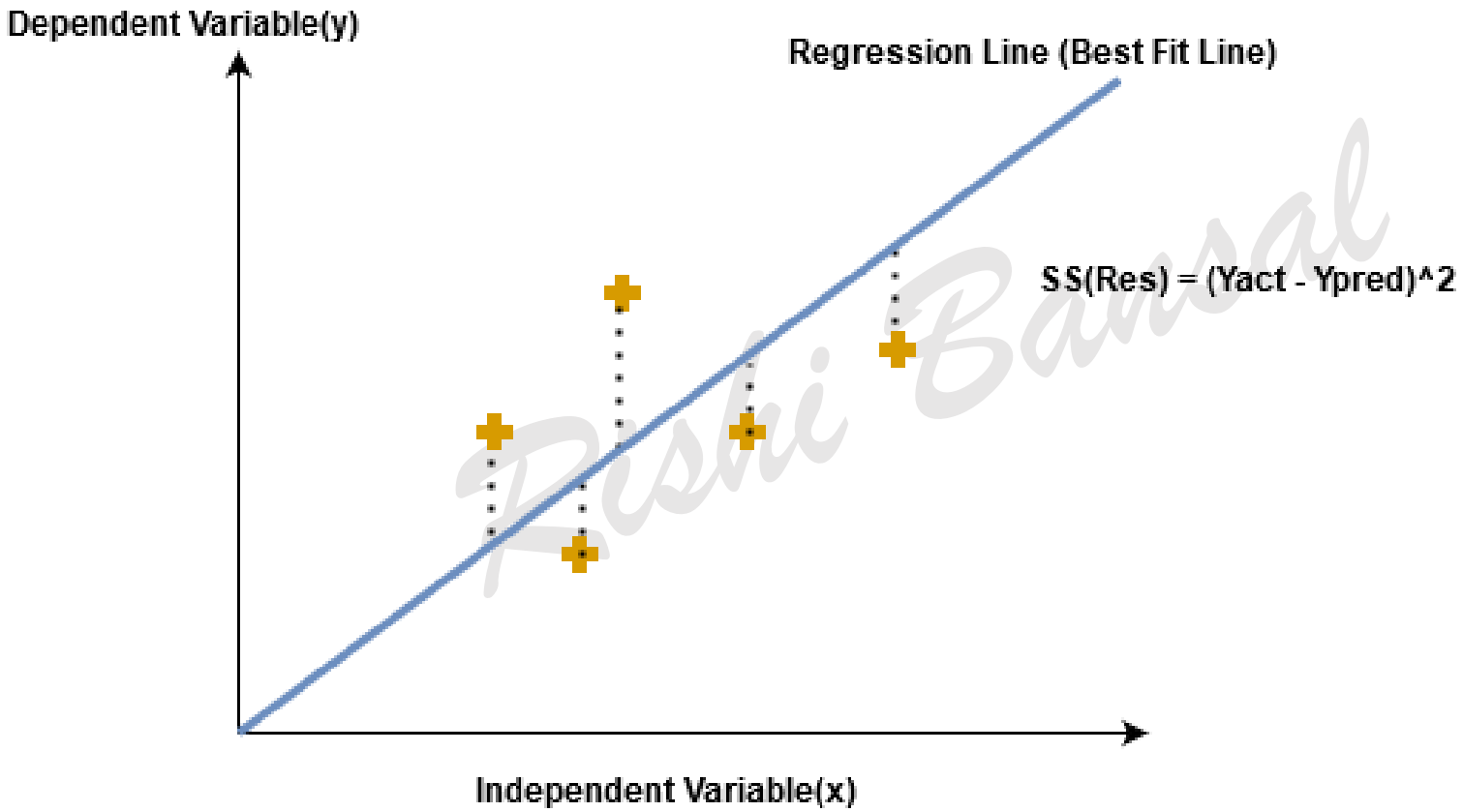
Regression

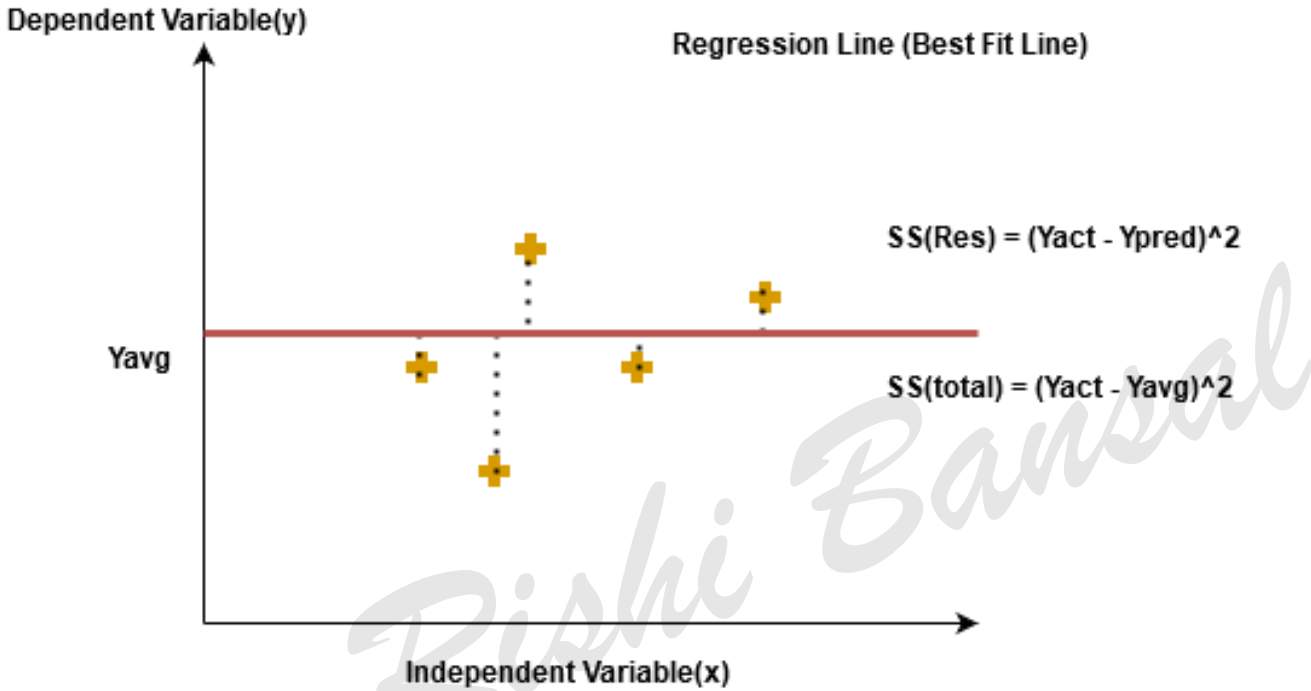
Regression Model Performance(r^2)

- It tells how well regression equation explains the data.



Regression Model Performance

- R^2



$$r^2 = 1 - \frac{Sum\,of\,Squares\,of\,Errors(SSE)}{Total\,Sum\,of\,Squares(SST)}$$

Regression Model Performance

- A value of R^2 = 1 means regression predictions perfectly fit/explains the data.

Question: Can $r2$ be negative?

- Ans: When: (Sum of Square Errors(SSE) > {Total Sum of Squares(SST)})
- This means when our predicted model performs worst than average line which is a very rare case.

$$r^2 = 1 - \frac{Sum of Squares of Errors(SSE)}{Total Sum of Squares(SST)}$$

Confusion Matrix

Describe the performance of a classification model

- Accuracy: Is fraction of correct predictions in all prediction made by model
- Precision: Is fraction of correct positive predictions in all positive predictions made by the model
- Recall: Is fraction of correct positive predictions made in actual positive data

$$\text{Precision} = \frac{TP}{PredictedYes} = 65/73 = 0.89$$

$$\text{Sensitivity/Recall} = \frac{TP}{ActualYes} = 65/68 = 0.96$$

$$\text{Accuracy} = \frac{TN+TP}{Total} = 89/100 = 0.89$$

$$\text{Error Rate} = \frac{FN+FP}{Total} = 11/100 = 0.11$$

n=100		Predicted: Positive	Predicted: Negative	
Actual	Positive	TP=65	FN=3	68
Actual	Negative	FP=8	TN=24	32
		73	27	

Relevance of Confusion Matrix

- Spam Filter (positive class: spam): Optimize for precision or specificity because false negatives (spam goes to inbox) are more acceptable than false positives (non-spam caught by spam filter).

$$\text{Precision} = \frac{TP}{PredictedYes} = 65/73 = 0.89$$

- Fraudulent transaction detector (positive class: fraud): Optimize for sensitivity because false positives (normal transactions that are flagged as possible fraud) are more acceptable than false negatives (fraudulent transactions that are not detected)

$$\text{Sensitivity/Recall} = \frac{TP}{ActualYes} = 65/68 = 0.96$$

- Type I Error: FP
- Type II Error: FN

n=100		Predicted: Positive	Predicted: Negative	
Actual	Positive	TP=65	FN=3	68
Actual	Negative	FP=8	TN=24	32
		73	27	