PERFORMANCE PARAMETER

PRESENTED BY:
NITISH VIG

PERFORMANCE PARAMETER

Basics

There are various parameters in Machine Learning that are useful to compare performance or score of different algorithms

Syntax : *from sklearn import metrics*
from sklearn import stats

The parameters are used for :

(a) Correlation between different features of the same dataset.
(b) Comparing score of dataset with itself or other algorithms
(c) Analyzing Statistical values of univariate, bivariate

COEFFICIENT

Basics

ICTACADEMY

They are the estimators of unknown population parameters and describe relationship between predictor variable and response.

Sign indication :
" + " : With increase in predictor value, the response increases
" - " : With decrease in predictor value, the response decreases

Note, this relationship holds if others variables remain constant

Coefficient Example

Predictors			Actual Y
X1	X2	X3	Y
10	2.5	7.6	4.4125
2	1.5	2.4	2.1275
6	7.5	9.6	10.5975
4	3.5	3.7	7.9575
6	4.5	5	10.7825
6	1.5	2.3	7.1675
4	3.5	6.2	2.9575
0	7.5	7.3	7.9375
8	2.5	4.9	7.3925
2	9.5	11.2	8.5675

Coefficient

$$X1*1.21 + X2*3.005 - X3*2$$

INTERCEPTS

Basics

Practically, it doesn't apply anything so don't try to focus much on questions why y = 0 or y is negative.

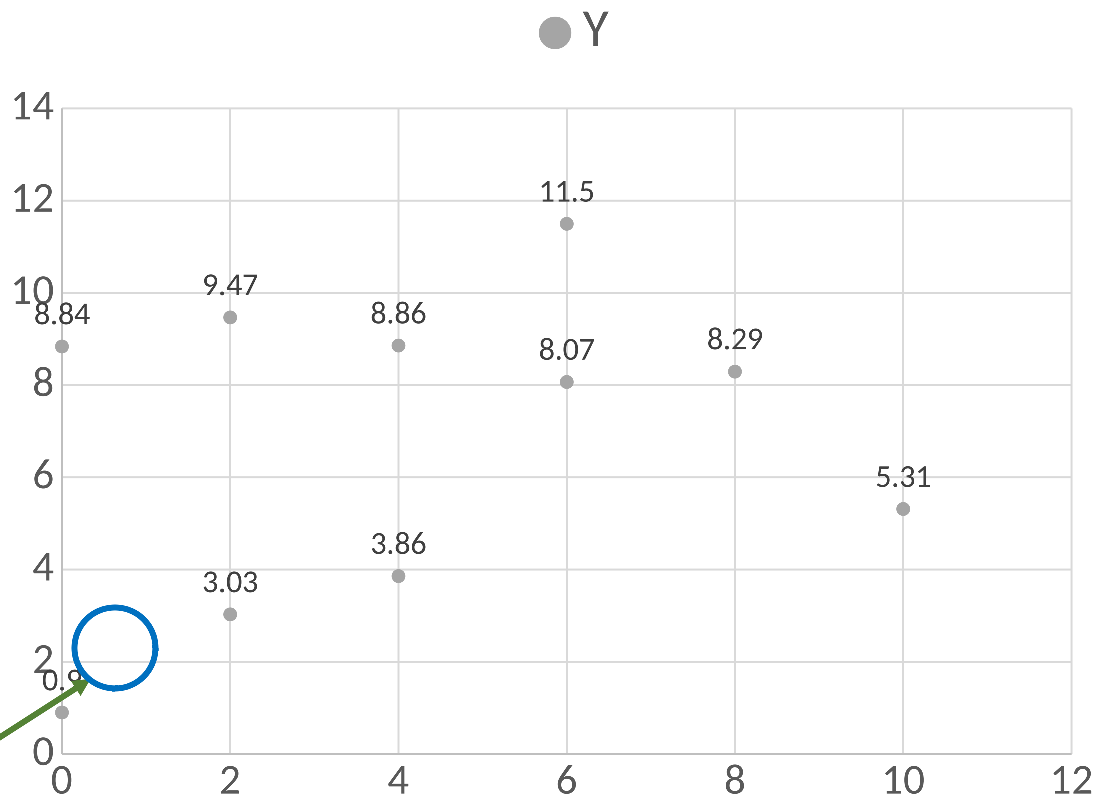
One of the purpose of this constant is that it forces the residuals to have the crucial zero mean.

If we omit this constant, then every regression line will pass through (0 , 0) forcing other variables to 0 as well.

Intercept Example

ICTACADEMY

X1	X2	X3	Y
10	2.5	7.6	5.3125
2	1.5	2.4	3.0275
6	7.5	9.6	11.4975
4	3.5	3.7	8.8575
0	0	0	0.9
6	1.5	2.3	8.0675
4	3.5	6.2	3.8575
0	7.5	7.3	8.8375
8	2.5	4.9	8.2925
2	9.5	11.2	9.4675



Intercept

0.9 + X1*1.21 + X2*3.005-X3*2

SUM OF SQUARES

Basics

Sum of Squares (SS) is simply defined as Sum of squared differences between two values

Formula : *SS = (columnName1 – columnName2) ^ 2*

In itself, it doesn't have any information.
But, when they are applied to find errors or mean difference, they imply a significant role to find the resultant.

For SS, you just have to omit denominator part.

SS Example

Formula : *SS = (columnName1 – columnName2) ^ 2*

X	µ	Y - µ	(Y - µ)^2
13	15	-2	4
16	15	1	1
10	15	-5	25
20	15	5	25
13	15	-2	4
18	15	3	9
			68

$$\text{Variance} = \frac{\overbrace{\Sigma(Y-\mu)^2}^{SS}}{n} = \frac{68}{6} = 11.33$$

RESIDUAL

SUM OF ERROR

(RSE)

Basics

Residual Sum of Errors (RSE) is defined as Sum of errors between two features.

Formula : *RSE = (y_test – y_predicted)*

A residual is defined as the error caused due to difference between 2 features.
RSE is sum of residuals.

RSE Example

ICTACADEMY

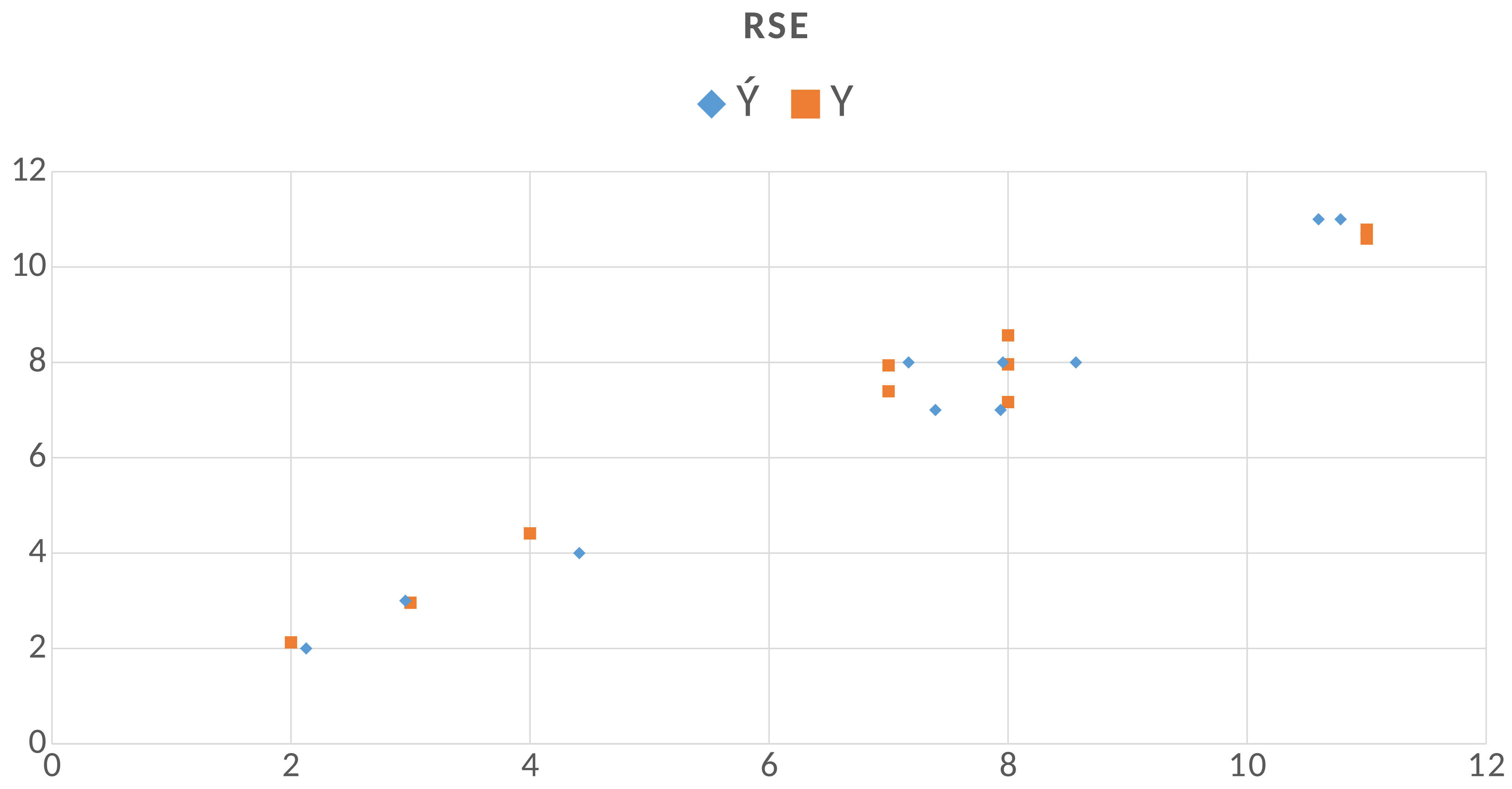
Syntax : *RSE = Σ(columnName1 – columnName2)*

Predictors			Actual Y	Predicted Y	Error
X1	X2	X3	Y	Ý	Y - Ý
10	2.5	7.6	4.4125	4	0.4125
2	1.5	2.4	2.1275	2	0.1275
6	7.5	9.6	10.598	11	-0.4025
4	3.5	3.7	7.9575	8	-0.0425
6	4.5	5	10.783	11	-0.2175
6	1.5	2.3	7.1675	8	-0.8325
4	3.5	6.2	2.9575	3	-0.0425
0	7.5	7.3	7.9375	7	0.9375
8	2.5	4.9	7.3925	7	0.3925
2	9.5	11.2	8.5675	8	0.5675
					0.9 **Total**

Residual

Residual Sum of Error (RSE)

RSE Example



SUM OF SQUARED ERRORS

RSS Basics

ICTACADEMY

Residual Sum of Squares (RSS) is defined as Sum of squared errors between the actual and predicted response.

Formula : *RSS = (y_test – y_predicted) ^ 2*

RSS is used to find MSE and RMSE.

Lower the value, better is the ML model.

It is the quantity used in defining an error which is not explained by the model.
We try to minimize the error as small as possible.

ESS Basics

Explained Sum of Squares (ESS) is defined as Sum of squared errors between the actual and mean response.

Formula : *ESS = (y_test – mean(y_test)) ^ 2*

Higher the value, better is the ML model.

It is the quantity used in describing the error (or irregularity or variance) which is explained by the model.

TSS Basics

ICTACADEMY

Total Sum of Squares (TSS) is defined as sum of squared residual errors and sum of squared explained errors.

Syntax : *TSS = RSS + ESS*

It is used to find r-squared to get the accuracy measure of model

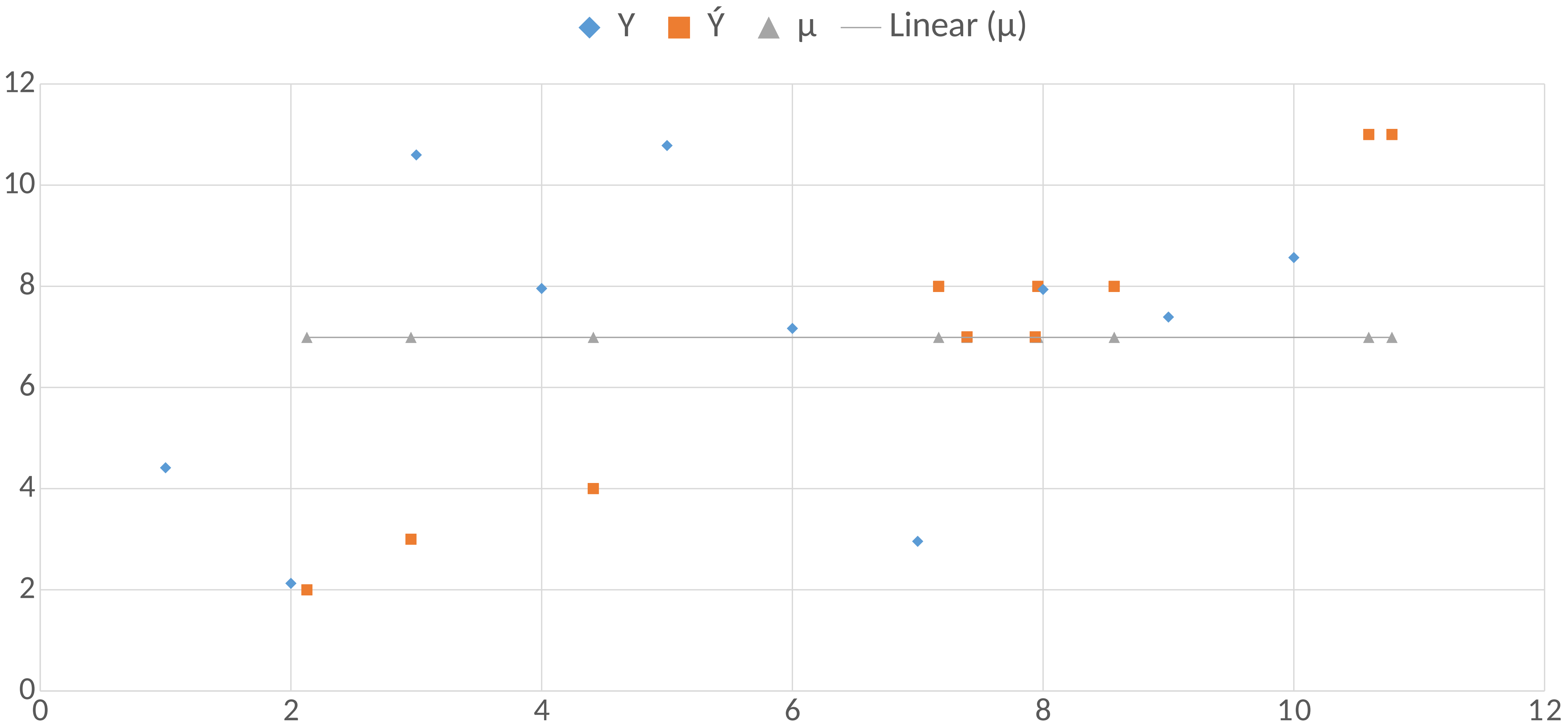
SSE Example

ICTACADEMY

Predictors			Actual Y	Mean Y	Expained Y	Expained Y^2	Predicted Y	Error	Error^2
X1	X2	X3	Y	µ	Y - µ	(Y - µ)^2	Ý	Y - Ý	(Y - Ý)^2
10	2.5	7.6	4.4125	6.99	-2.5775	6.64350625	4	0.4125	0.17016
2	1.5	2.4	2.1275	6.99	-4.8625	23.64390625	2	0.1275	0.01626
6	7.5	9.6	10.598	6.99	3.6075	13.01405625	11	-0.4025	0.16201
4	3.5	3.7	7.9575	6.99	0.9675	0.93605625	8	-0.0425	0.00181
6	4.5	5	10.783	6.99	3.7925	14.38305625	11	-0.2175	0.04731
6	1.5	2.3	7.1675	6.99	0.1775	0.03150625	8	-0.8325	0.69306
4	3.5	6.2	2.9575	6.99	-4				
0	7.5	7.3	7.9375	6.99	0				
8	2.5	4.9	7.3925	6.99	0				
2	9.5	11.2	8.5675	6.99	1				

RSS	ESS	TSS
Σ(Y - Ý)^2	Σ(Y - µ)^2	RSS + ESS
2.4474125	78.4614125	80.908825

SSE Example



R SQUARED (R^2)

R^2 Basics

It is also known as the coefficient of determination.
It simply measures how close the data are to the fitted regression line, ie, it is the percentage of the response variable explained by the model.

It is the statistical tool used for indicating the variance of the dependent variable that is to be predicted from the independent variable.

Formula : $R^2 =$

n = observations

Syntax : *from sklearn import metrics*

metrics.r2_score(y_test – y_predicted)

R^2 Basics

R^2 varies from 0% to 100%.
0% means that model explains none of the variability of response data around it's mean.
100% means that model explains all variability of response data around it's mean.

In general, higher the R^2 better the model fits the data.

Limitation :

1. R^2 can't determine if coefficient estimator and predictor are biased. This is because, R^2 increases with increase in number of predictors. (so, watch residual plots also before making a decision)

2. We can have low R^2 for good model or high R^2 for not a good model.

CONFUSION MATRIX

Basics

ICTACADEMY

Confusion Matrix gives the metrics output that describes the complete performance of the model.

It helps to determine in predicting the few parameters that defines accuracy of the model.

Syntax : *from sklearn import metrics*
metrics.confusion_matrix(y_test , y_predicted)

	Predicted 'Yes'	**Predicted No'**
Actual 'Yes'	True Positive	False Negative
Actual 'No'	False Positive	True Negative

Basics

Important Terms :

True Positives : The cases in which we predicted YES and the actual output was also YES.

True Negatives : The cases in which we predicted NO and the actual output was NO.

False Positives (Type 1 Error) : The cases in which we predicted YES and the actual output was NO.

False Negatives (Type 2 Error) : The cases in which we predicted NO and the actual output was YES.

Confusion Matrix Example

ICTACADEMY

Age	Education	Gender	Pocket Money	Hobby	Will Invest	Predicted
19	Graduation	Girl	100	Cooking	Yes	Yes
20	Graduation	Boy	200	Cooking	Yes	Yes
20	Graduation	Boy	100	Jogging	No	No
16	Graduation	Girl	100	Jogging	No	No
16	Graduation	Girl	300	Jogging	No	No
19	Graduation	Boy	100	Jogging	Yes	Yes
20	Graduation	Girl	200	Dancing	No	Yes
19	Graduation	Girl	500	Cooking	Yes	No
22	Post Graduation	Girl	200	Jogging	No	Yes
22	Post Graduation	Boy	500	Dancing	Yes	Yes
22	Post Graduation	Girl	500	Dancing	Yes	No
19	Graduation	Boy	500	Jogging	No	Yes
19	Graduation	Girl	200	Dancing	No	Yes
16	Graduation	Girl	100	Cooking	No	Yes
22	Post Graduation	Boy	200	Dancing	No	No
16	Graduation	Boy	300	Dancing	Yes	Yes

Confusion Matrix Example

ICTACADEMY

Age	Education	Gender	Pocket Money	Hobby	Will Invest	Predicted
19	Graduation	Girl	100	Cooking	Yes	Yes
20	Graduation	Boy	200	Cooking	Yes	Yes
20	Graduation	Boy	100	Jogging	No	No
16	Graduation	Girl	100	Jogging	No	No
16	Graduation	Girl	300	Jogging	No	No
19	Graduation	Boy	100	Jogging	Yes	Yes
20	Graduation	Girl	200	Dancing	No	Yes
19	Graduation	Girl	500	Cooking	Yes	No
22	Post Graduation	Girl	200	Jogging	No	Yes
22	Post Graduation	Boy	500	Dancing	Yes	Yes
22	Post Graduation	Girl	500	Dancing	Yes	No
19	Graduation	Boy	500	Jogging	No	Yes
19	Graduation	Girl	200	Dancing	No	Yes
16	Graduation	Girl	100	Cooking	No	Yes
22	Post Graduation	Boy	200	Dancing	No	No
16	Graduation	Boy	300	Dancing	Yes	Yes

True Positive

Confusion Matrix Example

ICTACADEMY

Age	Education	Gender	Pocket Money	Hobby	Will Invest	Predicted
19	Graduation	Girl	100	Cooking	Yes	Yes
20	Graduation	Boy	200	Cooking	Yes	Yes
20	Graduation	Boy	100	Jogging	No	No
16	Graduation	Girl	100	Jogging	No	No
16	Graduation	Girl	300	Jogging	No	No
19	Graduation	Boy	100	Jogging	Yes	Yes
20	Graduation	Girl	200	Dancing	No	Yes
19	Graduation	Girl	500	Cooking	Yes	No
22	Post Graduation	Girl	200	Jogging	No	Yes
22	Post Graduation	Boy	500	Dancing	Yes	Yes
22	Post Graduation	Girl	500	Dancing	Yes	No
19	Graduation	Boy	500	Jogging	No	Yes
19	Graduation	Girl	200	Dancing	No	Yes
16	Graduation	Girl	100	Cooking	No	Yes
22	Post Graduation	Boy	200	Dancing	No	No
16	Graduation	Boy	300	Dancing	Yes	Yes

True Negative

Confusion Matrix Example

ICTACADEMY

Age	Education	Gender	Pocket Money	Hobby	Will Invest	Predicted
19	Graduation	Girl	100	Cooking	Yes	Yes
20	Graduation	Boy	200	Cooking	Yes	Yes
20	Graduation	Boy	100	Jogging	No	No
16	Graduation	Girl	100	Jogging	No	No
16	Graduation	Girl	300	Jogging	No	No
19	Graduation	Boy	100	Jogging	Yes	Yes
20	Graduation	Girl	200	Dancing	No	Yes
19	Graduation	Girl	500	Cooking	Yes	No
22	Post Graduation	Girl	200	Jogging	No	Yes
22	Post Graduation	Boy	500	Dancing	Yes	Yes
22	Post Graduation	Girl	500	Dancing	Yes	No
19	Graduation	Boy	500	Jogging	No	Yes
19	Graduation	Girl	200	Dancing	No	Yes
16	Graduation	Girl	100	Cooking	No	Yes
22	Post Graduation	Boy	200	Dancing	No	No
16	Graduation	Boy	300	Dancing	Yes	Yes

False Positive

Confusion Matrix Example

ICTACADEMY

Age	Education	Gender	Pocket Money	Hobby	Will Invest	Predicted
19	Graduation	Girl	100	Cooking	Yes	Yes
20	Graduation	Boy	200	Cooking	Yes	Yes
20	Graduation	Boy	100	Jogging	No	No
16	Graduation	Girl	100	Jogging	No	No
16	Graduation	Girl	300	Jogging	No	No
19	Graduation	Boy	100	Jogging	Yes	Yes
20	Graduation	Girl	200	Dancing	No	Yes
19	Graduation	Girl	500	Cooking	Yes	No
22	Post Graduation	Girl	200	Jogging	No	Yes
22	Post Graduation	Boy	500	Dancing	Yes	Yes
22	Post Graduation	Girl	500	Dancing	Yes	No
19	Graduation	Boy	500	Jogging	No	Yes
19	Graduation	Girl	200	Dancing	No	Yes
16	Graduation	Girl	100	Cooking	No	Yes
22	Post Graduation	Boy	200	Dancing	No	No
16	Graduation	Boy	300	Dancing	Yes	Yes

False Negative

CLASSIFICATION REPORT

Basics

Classification Report is used to measure the quality of predictions.

The report shows the main classification metrics precision, recall and f1-score on a per-class basis. The metrics are calculated by using true and false positives, true and false negatives.

Syntax : *from sklearn import metrics*

metrics.classification_report(y_test , y_predicted)

Basics

ICTACADEMY

Age	Education	Gender	Pocket Money	Hobby	Will Invest	Predicted
19	Graduation	Girl	100	Cooking	Yes	Yes
22	Post Graduation	Boy	500	Dancing	Yes	Yes
22	Post Graduation	Girl	500	Dancing	Yes	No
19	Graduation	Boy	500	Jogging	No	Yes
19	Graduation	Girl	200	Dancing	No	Yes
16	Graduation	Girl	100	Cooking	No	Yes
22	Post Graduation	Boy	200	Dancing	No	No
16	Graduation	Boy	300	Dancing	Yes	Yes

	Predicted 'Yes'	Predicted No'
Actual 'Yes'	TP = 5	FN = 2
Actual 'No'	FP = 5	TN = 4

Type II Error

Type I Error

Classification Report Parameters

ICTACADEMY

(a) Accuracy : It is defined as the total number of correct predictions to the total number of sample size

Age	Educa
19	Gradu
20	Gradu
20	Gradu
16	Gradu
16	Gradu
19	Gradu
20	Gradu
19	Gradu
22	Post Gra
22	Post Gra
22	Post Gra
19	Gradu
19	Gradu
16	Gradu
22	Post Gra
16	Gradu

	Predicted 'Yes'	Predicted No'		
Actual 'Yes'	5	2	7	TP + FN
Actual 'No'	5	4	9	FP + TN
	10	6	**16**	
	TP + FP	FN + TN		

ccuracy

Accuracy / Classification Rate / Support

$$\frac{TP + TN}{TP + TN + FP + FN} = \frac{9}{16} \qquad \mathbf{0.56}$$

Classification Report Parameters

ICTACADEMY

(b) Misclassification Rate : It is defined as the total number of incorrect predictions to the total number of sample size

Age	Education
19	Graduation
20	Graduation
20	Graduation
16	Graduation
16	Graduation
19	Graduation
20	Graduation
19	Graduation
22	Post Graduati
22	Post Graduati
22	Post Graduati
19	Graduation
19	Graduation
16	Graduation
22	Post Graduati
16	Graduation

	Predicted 'Yes'	Predicted No'		
Actual 'Yes'	5	2	7	TP + FN
Actual 'No'	5	4	9	FP + TN
	10	6	**16**	
	TP + FP	FN + TN		

sclassification Rate

Misclassification Rate

FP + FN	=	7	**0.44**
TP + TN + FP + FN		16	

Classification Report Parameters

ICTACADEMY

(c) Recall : It is defined as the fraction of actual positives that were correctly identified.

Age	Educa
19	Grad
20	Grad
20	Grad
16	Grad
16	Grad
19	Grad
20	Grad
19	Grad
22	Post Gr
22	Post Gr
22	Post Gr
19	Grad
19	Grad
16	Grad
22	Post Gr
16	Grad

	Predicted 'Yes'	Predicted No'		
Actual 'Yes'	5	2	7	TP + FN
Actual 'No'	5	4	9	FP + TN
	10	6	**16**	
	TP + FP	FN + TN		

Recall

Recall

$$\frac{TP}{TP + FN} = \frac{5}{7} \quad \mathbf{0.71}$$

Classification Report Parameters

ICTACADEMY

(d) Precision : It is defined as accuracy of positive predictions.

Age	Educ
19	Gra
20	Gra
20	Gra
16	Gra
16	Gra
19	Gra
20	Gra
19	Gra
22	Post (
22	Post (
22	Post (
19	Gra
19	Gra
16	Gra
22	Post (
16	Gra

	Predicted 'Yes'	Predicted No'		
Actual 'Yes'	5	2	7	TP + FN
Actual 'No'	5	4	9	FP + TN
	10	6	**16**	
	TP + FP	FN + TN		

➤ **Precision**

Precision

$$\frac{TP}{TP + FP} = \frac{5}{10} \quad \mathbf{0.50}$$

THANK YOU

Reach out to me at: **nitish@ictacademy.in**