Example: Confusion Matrix

Satishkumar L. Varma

Dataset, N = 20 {4Class, 5Item}		Actual Class (Condition Given)				Y+N	Performance Measures
		Y		N			
Predicted Class (Outcome)	P	TP Correct Decision Prob = 1 - α	2	1	FP Type I Errors Prob = α	3	P: PPV = TP/(TP+FP) = 2 / (2+1) **0.667**
	N	FN Type II Errors Prob = β	3	14	TN Correct Decision Prob = 1- β	17	NPV = TN/(TN+FN) = 3 / (3+14) **0.176**
Total in Dataset	N+P	**5**		**15**		**20**	
Performance Measures		R: Sensitivity = TP/(TP+FN) = 2 / (2+3) **0.400**		Specificity = TN/(TN+FP) = 14 / (14+1) **0.933**			A=(TP+TN)/(TP+FP+TN+FN) = (2+14) / (2+1+3+14) **0.8**

Evaluation Metric	Formula	Value	Meaning
Precision P (PPV)	TP / (TP+FP)	0.667	Proportion of the retrieved documents that are correct
Recal R (Sensitivity)	TP / (TP+FN)	0.400	Proportion of the positives that the model retrieved
Accuracy (Recognition Rate)	(TP+TN) / Total	0.800	Proportion of total number of predictions that were correct
Error Rate	1 - Accuracy	0.200	Error with respect to recognition rate
Fall-out	FP / (FP+TN)	0.067	Proportion of non-relevant retrieved of all non-relevant.
F-Measure (F1 or F-Score)	2*(P*R)/(P+R)	0.500	Weighted harmonic mean of precision and recall.
Specificity (TN Recognition Rate)	TN / (TN+FP)	0.933	Proportion of actual -ve cases which are correctly identified.
NPV	TN / (TN+FN)	0.176	Proportion of -ve cases that were correctly identified.

Inference	Example	Meaning
Type I Error (α) or Level of Statistical Significance	For eg: α = 0.05; then the researcher has set 5% as the maximum chance of incorrectly rejecting the Ho.	Prob. of rejecting the Ho when it is actually true. (Falsely rejecting a Ho)
Type II Error (β)	For eg: β is set at 0.10, then the investigator has decided willing to accept a 10% chance of missing an association of a given effect size.	Prob. of failing to reject the Ho when it is actually false. (Falsely accepting a Ho)
Power(1 - β)	If β is set at 0.10, then it represents a power of 0.90, i.e., a 90% chance of finding an association of that size.	Prob. of observing an effect in the sample (if any), of a specified effect size or greater exists in the population.
P Value	Ho is rejected in favor of the Ha if P value < α, the predetermined level of statistical significance.	Prob. of obtaining the study results by chance if the Ho is true.

Confusion Matrix: Example 2

IR System	Retrieved (Ret) POSITIVE (P)	Not Retrieved (nRet) NEGATIVE (N)	Total in Dataset
Relevant (Rel)	Rel_Ret (TP) **12**	Rel_nRet (FN) **8**	Total_Rel **20**
Non-Relevent (nRel)	Rel_nRet (FP) **3**	nRel_nRet (TN) **77**	Total_nRel **80**
Total in Dataset	Total_Ret **15**	Total_nRet **85**	Total **100**

Metric	Formula	Value	Meaning
Precision P (Positive Predictive Value)	TP / (TP+FP)	0.80	Proportion of the retrieved documents that are correct
Recal R (Sensitivity)	TP / (TP+FN)	0.60	Proportion of the positives that the model retrieved
Accuracy (Recognition Rate)	(TP+TN) / Total	0.89	Proportion of total number of predictions that were correct
Error Rate	1 - Accuracy	0.11	Error with respect to recognition rate
Fall-out	FP / (FP+TN)	0.04	Proportion of non-relevant retrieved of all non-relevant.
F-Measure (F1 or F-Score)	2*(P*R)/(P+R)	0.69	Weighted harmonic mean of precision and recall.
Specificity (TN Recognition Rate)	TN / (FN+TN)	0.91	Proportion of actual -ve cases which are correctly identified.
Negative Predictive Value	TN / (FP+TN)	0.96	Proportion of -ve cases that were correctly identified.

Confusion Matrix: Example with 100 samples

Confusion Matrix		Target Performance: Positive (Retrieved)		Target Performance: Negative (Not Retrieved)		Performance Metrics	
System Model N=100	Relevant	12	TP	FN	8	Sensitivity (Recall)	TP/(TP+FN)= 0.60
	Irrelevant	3	FP	TN	77	Negative Predictive Value	TN/(TN+FP)= 0.96
Performance Metrics		Positive Predictive Value (Precision)		Specificity (True Negative Rate)		Accuracy, Recognition Rate	(TP+TN)/N= 0.89
		TP/(TP+FP)= 0.80		TN/(FN+TN)= 0.91			

Confusion Matrix: Example with 20 samples

N = 20		Target Performance: Positive (Retrieved)		Target Performance: Negative (Not Retrieved)		Performance Metrics	
System Model	Relevant	3	TP	FN	2	Sensitivity (Recall)	TP/(TP+FN)= 0.60
	Irrelevant	1	FP	TN	14	Negative Predictive Value	TN/(TN+FP)= 0.93
Performance Metrics		Positive Predictive Value (Precision)		Specificity (True Negative Rate)		Accuracy, Recognition Rate	(TP+TN)/N= 0.85
		TP/(TP+FP)= 0.75		TN/(FN+TN)= 0.88			

Confusion Matrix: Example with 20 samples

IR System	Retrieved (Ret) POSITIVE (P)	Not Retrieved (nRet) NEGATIVE (N)	Total in Dataset
Relevant (Rel)	Rel_Ret (TP) 3	Rel_nRet (FN) 2	Total_Rel 5
Non-Relevent (nRel)	Rel_nRet (FP) 1	nRel_nRet (TN) 14	Total_nRel 15
Total in Dataset	Total_Ret 4	Total_nRet 16	Total (All) 20

Metric	Formula	Value	Meaning
Precision P (Positive Predictive Value)	TP / (TP+FP)	0.75	Proportion of the retrieved documents that are correct
Recal R (Sensitivity)	TP / (TP+FN)	0.60	Proportion of the positives that the model retrieved
Accuracy (Recognition Rate)	(TP+TN) / All	0.85	Proportion of total number of predictions that were correct
Error Rate	1 - Accuracy	0.15	Error with respect to recognition rate
Fall-out	FP / (FP+TN)	0.07	Proportion of non-relevant retrieved of all non-relevant.
F-Measure (F1 or F-Score)	2*(P*R)/(P+R)	0.67	Weighted harmonic mean of precision and recall.
Specificity (TN Recognition Rate)	TN / (FN+TN)	0.88	Proportion of actual -ve cases which are correctly identified.
Negative Predictive Value	TN / (FP+TN)	0.93	Proportion of -ve cases that were correctly identified.

Example: Multiclass Confusion Matrix

Irris Dataset		Actual Class (Condition Given)		
		SETOSA	VERSICOLR	VIRGINICA
Predicted Class (Outcome)	SETOSA	16	0	0
	VERSICOLR	0	17	0
	VIRGINICA	0	1	11

◉ To apply a classifier model to Iris dataset to classify the given instance as Versicolor or Virginia, or Setosa flower.
◉ With the help of petal length, petal width, sepal length, and sepal width, the model has to classify the given instance as Versicolor or Virginia, or Setosa flower.
◉ The dataset has 3 classes; hence we get a 3 X 3 confusion matrix.

◉ **Let us calculate the TP, TN, FP, and FN values for the class Setosa using the Above tricks:**

TP: The actual value and predicted value should be the same.
So concerning Setosa class, the value of cell 1 is the TP value.

FN: The sum of values of corresponding rows except for the TP value.
FN = (cell 2 + cell3) = (0 + 0)= 0

FP: The sum of values of the corresponding column except for the TP value.
FP = (cell 4 + cell 7)= (0 + 0)= 0

TN: The sum of values of all columns and rows except the values of that class.
TN = (cell 5 + cell 6 + cell 8 + cell 9)= 17 +0 + 1 + 11= 29

◉ **Similarly, for the Versicolor class, the values/metrics are calculated as below:**

TP: 17 (cell 5)

FN : 0 + 1 = 1 (cell 4 +cell 8)

FP : 0 + 0 = 0 (cell 2 + cell 6)

TN : 16 +0 +0 + 11 =27 (cell 1 + cell 3 + cell 7 + cell 9).

Satishkumar L. Varma, PCE, New Panvel

Dataset, N = 20 {4Class, 5Item} Truth in the population		Actual Class (Condition Given) P			N	P+N	Performance Measures
Predicted Class (Outcome)	P Reject Ho	TP Correct Decision Prob = 1 - α	2	1	FP Type I Errors Prob = α	3	PPV = TP/(TP+FP) = 2 / (2+1) **0.667**
	N Fail to reject Ho	FN Type II Errors Prob = β	3	14	TN Correct Decision Prob = 1- β	17	NPV = TN/(TN+FN) = 3 / (3+14) **0.176**
Total in Dataset	**N+P**	**5**		**15**		**20**	
Performance Measures		Sensitivity = TP/(TP+FN) = 2 / (2+3) **0.400**		Specificity = TN/(TN+FP) = 14 / (14+1) **0.933**			A=(TP+TN)/(TP+FP+TN+FN) = (2+14) / (2+1+3+14) **0.8**

Evaluation Metric	Example	Meaning
Type I Error (α) or Level of Statistical Significance	For eg: α = 0.05; then the researcher has set 5% as the maximum chance of incorrectly rejecting the Ho.	Prob. of rejecting the Ho when it is actually true. (Falsely rejecting a Ho)
Type II Error (β)	For eg: β is set at 0.10, then the investigator has decided willing to accept a 10% chance of missing an association of a given effect size.	Prob. of failing to reject the Ho when it is actually false. (Falsely accepting a Ho)
Power(1 - β)	If β is set at 0.10, then it represents a power of 0.90, i.e., a 90% chance of finding an association of that size.	Prob. of observing an effect in the sample (if any), of a specified effect size or greater exists in the population.
P Value	Ho is rejected in favor of the Ha if P value < α, the predetermined level of statistical significance.	Prob. of obtaining the study results by chance if the Ho is true.

Plus '+' (a)

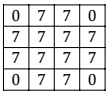


Image ID: 1

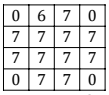


Image ID: 2

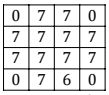


Image ID: 3

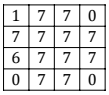


Image ID: 4

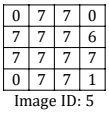


Image ID: 5

Minus '-' (b)

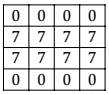


Image ID: 6

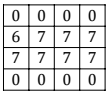


Image ID: 7

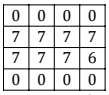


Image ID: 8

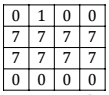


Image ID: 9

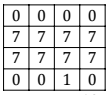


Image ID: 10

Slace '\' (c)

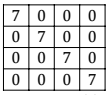


Image ID: 11

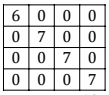


Image ID: 12

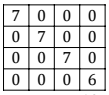


Image ID: 13

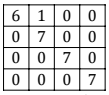


Image ID: 14

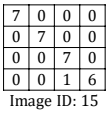


Image ID: 15

Exclaim '!' (d)

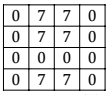


Image ID: 16

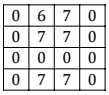


Image ID: 17

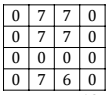


Image ID: 18

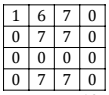


Image ID: 19

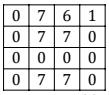
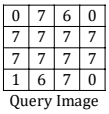


Image ID: 20

Class	Image ID	Image (I) Feature Vector								Dist(I, Query) = (I - Q)^2							Euclidean
		0	1	2	3	4	5	6	7	3	1	0	0	0	0	10	sqrt(Dist)
Plus '+'	ID: 1	4	0	0	0	0	0	0	12	1	1	0	0	0	0	4	4	3.162
	ID: 2	4	0	0	0	0	0	1	11	1	1	0	0	0	0	1	1	2.000
	ID: 3	4	0	0	0	0	0	1	11	1	1	0	0	0	0	1	1	2.000
	ID: 4	3	1	0	0	0	0	1	11	0	0	0	0	0	0	1	1	1.414
	ID: 5	3	1	0	0	0	0	1	11	0	0	0	0	0	0	1	1	1.414
Minus '-'	ID: 6	8	0	0	0	0	0	0	8	25	1	0	0	0	0	4	4	5.831
	ID: 7	8	0	0	0	0	0	0	1	7	25	1	0	0	0	0	1	9	6.000
	ID: 8	8	0	0	0	0	0	0	1	7	25	1	0	0	0	0	1	9	6.000
	ID: 9	7	1	0	0	0	0	0	0	8	16	0	0	0	0	0	4	4	4.899
	ID: 10	7	1	0	0	0	0	0	0	8	16	0	0	0	0	0	4	4	4.899
Slace '\'	ID: 11	12	0	0	0	0	0	0	4	81	1	0	0	0	0	4	36	11.045
	ID: 12	12	0	0	0	0	0	0	1	3	81	1	0	0	0	0	1	49	11.489
	ID: 13	12	0	0	0	0	0	0	1	3	81	1	0	0	0	0	1	49	11.489
	ID: 14	11	1	0	0	0	0	0	1	3	64	0	0	0	0	0	1	49	10.677
	ID: 15	11	1	0	0	0	0	0	1	3	64	0	0	0	0	0	1	49	10.677
Exclaim '!'	ID: 16	10	0	0	0	0	0	0	0	6	49	1	0	0	0	0	4	16	8.367
	ID: 17	10	0	0	0	0	0	0	1	5	49	1	0	0	0	0	1	25	8.718
	ID: 18	10	0	0	0	0	0	0	1	5	49	1	0	0	0	0	1	25	8.718
	ID: 19	9	1	0	0	0	0	0	1	5	36	0	0	0	0	0	1	25	7.874
	ID: 20	9	1	0	0	0	0	0	1	5	36	0	0	0	0	0	1	25	7.874



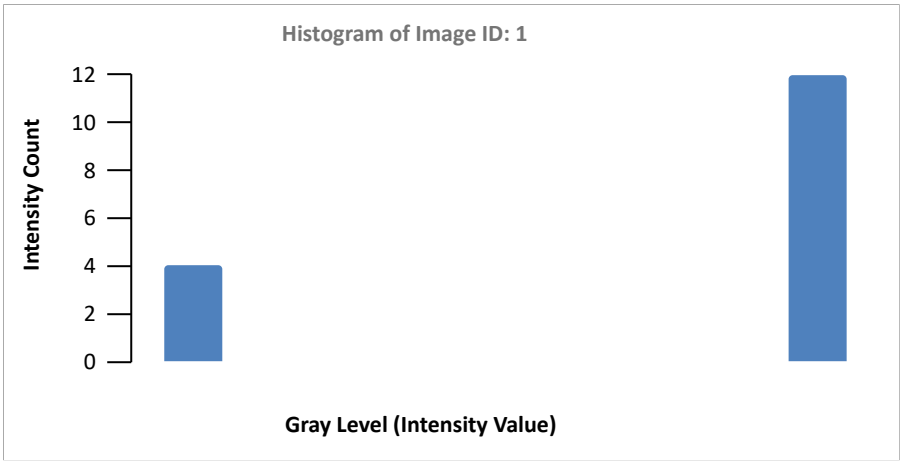
Query Image

Gray Levels in Query Image:	0	1	2	3	4	5	6	7
Gray Level Count:	3	1	0	0	0	0	2	10
Elements in Feature Vector:								

0	7	7	0
7	7	7	7
7	7	7	7
0	7	7	0

Image ID: 1

Parameter	Elements of Image ID: 1								# Elements in Vector
Gray Levels	0	1	2	3	4	5	6	7	8
Count	4	0	0	0	0	0	0	12	8
Feature Vector 1	4	0	0	0	0	0	0	12	8
Feature Vector 2	4		0		0		12		4
Feature Vector 3	4				12				2
Feature Vector 4	16								1



0	7	6	0
7	7	7	7
7	7	7	7
1	6	7	0

Query Image

Parameter	Elements of Query Image								# Elements in Vector
Gray Levels	0	1	2	3	4	5	6	7	8
Count	3	1	0	0	0	0	2	10	8
Feature Vector 1	3	1	0	0	0	0	2	10	8
Feature Vector 2	4		0		0		12		4
Feature Vector 3	4				12				2
Feature Vector 4	16								1