



Confusion Matrix 101

Confusion Matrix

Overview



		Predicted Class	
		Yes	No
Actual Class	Yes	True Pos (Hit)	False Neg (Type II Error)
	No	False Pos (Type I Error)	True Neg (Correct Rejection)

 True Outcome
 Errors

Confusion Matrix

Example

		Event Occured?	
		Yes	No
Event Predicted ?	Yes	True Pos (Hit)	False Pos (Type I Error) (False Alarm)
	No	False Neg (Type II Error) (Miss)	True Neg (Correct Rejection)

 True Outcome
 Errors

Confusion Matrix

Example: Tsunami

		Event Occured?	
		Yes	No
Event Predicted ?	Yes	Tsunami was observed, when it actually happened	A tsunami was predicted, but there was none
	No	There was a tsunami, but it was not predicted.	No tsunami occurred and nothing was Predicted

True Outcome

Less-critical Error

Critical Error

Confusion Matrix

Performance Measures: Accuracy

Numerator

		Effect Exists?	
		Yes	No
Effect Observed?	Yes	True Pos	False Pos
	No	False Neg	True Neg

Denominator

		Effect Exists?	
		Yes	No
Effect Observed?	Yes	True Pos	False Pos
	No	False Neg	True Neg

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Usually compared to baseline result or to compare models

Confusion Matrix

Example

		Predicted		
		Predicted Negative	Predicted Positive	
Actual	Actual Negative	TN = 50	FP = 10	Pred. Cond. Neg = 60
	Actual Positive	FN = 5	TP = 100	Pred. Cond. Pos = 105
		Condition Neg = 55	Condition Pos = 110	Total = 165

FPR
Specificity
TPR
Accuracy

The diagram illustrates the calculation of four key performance indicators from the confusion matrix:

- FPR (False Positive Rate):** Calculated as $\frac{FP}{FP + TN} = \frac{10}{10 + 50} = \frac{1}{6}$. A red arrow points from FP = 10 to the numerator, and a green arrow points from TN = 50 to the denominator.
- Specificity:** This is another term for FPR, representing the proportion of actual negatives correctly identified.
- TPR (True Positive Rate):** Calculated as $\frac{TP}{TP + FN} = \frac{100}{100 + 5} = \frac{20}{21}$. A yellow arrow points from TP = 100 to the numerator, and a blue arrow points from FN = 5 to the denominator.
- Accuracy:** Calculated as $\frac{TP + TN}{Total} = \frac{100 + 50}{165} = \frac{150}{165}$. A blue arrow points from TP = 100 and a yellow arrow points from TN = 50 to the numerator, and a blue arrow points from Total = 165 to the denominator.

TP...True Positive
FP...False Positive
TN...True Negative
FN...False Negative

Confusion Matrix

Performance Measures

Accuracy = $(TN+TP)/Total = (100+50)/165 = 0.91$

...measures correct classifier

Error Rate = $1 - Accuracy = 1 - 0.91 = 0.09$

...how often is classifier wrong?

Specificity = $TN / Actual\ NO = 50 / 60 = 0.83$

...when actually no,
how often predicted no?

TPR = $TP/Actual\ Pos = 100/105 = 0.95$

...when actually yes, how often predicted yes

FPR = $FP/Actual\ Neg = 10/60 = 0.17$

...when actually no, how often predicted yes

Precision = $TP / Predicted\ Pos = 100/110 = 0.91$

...when predicted yes,
how often correct?