

Evaluation Measures

Lesson



AI Academy

Limitation of Accuracy and Error

- Accuracy alone is not enough!
- It doesn't consider the balance of the class labels.
 - Same is true of Error Rate.

Example

Based on chemical analysis of the water trying to detect an oil slick in the sea:

- False positive: wrongly identifying an oil slick if there is none
- False negative: fail to identify an oil slick if there is one

Here, false negatives (environmental disasters) are much more costly than false positives (false alarms).

Model evaluation should take this into account

Cost Matrix

		PREDICTED CLASS	
ACTUAL CLASS	C(i j)	Class=Yes	Class= <u>No</u>
	<u>Class=Yes</u>	C(Yes Yes)	C(No Yes) +100
	Class=No	C(Yes No) +5	C(No No)

$C(i|j)$: Cost of misclassifying class j example as class i

Computing Cost of Classification

Fire Alarm

Cost Matrix		PREDICTED CLASS	
ACTUAL CLASS	C(i j)	+	-
	+	-1	100
	-	1	0

Model M ₁	PREDICTED CLASS	
ACTUAL CLASS	+	-
	150	40
	60	250

Accuracy = ?

Cost = ?

Model M ₂	PREDICTED CLASS	
ACTUAL CLASS	+	-
	250	45
	5	200

Accuracy = ?

Cost = ?

Computing Cost of Classification

Fire Alarm

		PREDICTED CLASS	
		C(i j)	
		+	-
ACTUAL CLASS	C(i j)	+1	-100
	+	-1	100
	-	1	0

		PREDICTED CLASS	
		Model M ₁	
		+	-
ACTUAL CLASS	+	150	40
	-	60	250

Accuracy = 80%

Cost = 3910

Accuracy = 90%

Cost = 4255

Cost: The lower, the better.

$$\text{Precision (p)} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall (r)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 Score} = \frac{2rp}{r + p} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}}$$



		PREDICTED	
ACTUAL		Yes	No
	Yes	(TP)	(FN)
	No	(FP)	(TN)

$$\text{Precision (p)} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

		PREDICTED	
ACTUAL		Yes	No
	Yes	(TP)	(FN)
	No	(FP)	(TN)

Precision: Of the items you said were positive ($TP+FP$), how many of them really were (TP).

$$\text{Recall (r)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

		PREDICTED	
ACTUAL		Yes	No
	Yes	(TP)	(FN)
	No	(FP)	(TN)

Recall: Of the items that really were positive ($TF+FN$), how many of them did you actually find (TP).

$$\text{F1 Score} = \frac{\underline{2rp}}{\underline{r + p}} = \frac{\underline{2TP}}{\underline{2TP + FN + FP}}$$

		PREDICTED	
ACTUAL		Yes	No
	Yes	(TP)	(FN)
	No	(FP)	(TN)

F1 Score: Harmonic mean of Precision and Recall.
Captures both measures in one score.

Learning Objectives: Evaluation Measures

You now should be able to:

- Use a cost matrix to compare model performance.
- Calculate Precision, Recall, and F1 Score from classification results



AI Academy
NC STATE



Evaluation Measures

Exercises



AI Academy

Question?

For the oil slick scenario: Based on chemical analysis of the water, try and detect an oil slick in the sea.

Do we want to maximize precision or recall?

Question?

For the oil slick scenario: Based on chemical analysis of the water, try and detect an oil slick in the sea.

We want to maximize recall (avoiding environmental disasters).

Maximizing precision (avoiding false alarms) is less important).

The F-measure can be used if precision and recall are equally important.

What is the accuracy, precision and recall?

Model M ₁	PREDICTED		
ACTUAL CLASS		+	-
	+	150	40
	-	60	250