

GroupG Practice

- Consider a classifier trained on examples $(\mathbf{x}, y)$ in the first two columns of the table below that makes the predictions on training data in the third column.

$\mathbf{x}$	y	$\hat{y}$
(0, 2)	1	1
(1, 1)	0	0
(2, 3)	0	0
(4, 1)	0	1

Complete the corresponding confusion matrix:

actual y	predicted $\hat{y}$	
	0	1
0		
1		

ANSWER:

actual y	predicted $\hat{y}$	
	0	1
0	2	1
1	0	1

- A classifier is evaluated on test data yielding this confusion matrix:

actual y	predicted $\hat{y}$	
	0	1
0	3	2
1	1	4

- What is the precision on the test data?

ANSWER:

The matrix is

actual y	predicted $\hat{y}$	
	0	1
0	TN	FP
1	FN	TP

so

$$precision = \frac{TP}{TP + FP} = \frac{4}{4 + 2} = \frac{2}{3} \approx 0.667$$

- What is the recall on the test data?

$$recall = \frac{TP}{TP + FN} = \frac{4}{4 + 1} = \frac{4}{5} = 0.8$$

- What is the accuracy on the test data?

$$accuracy = \frac{TP + TN}{TN + FN + FP + TP} = \frac{4 + 3}{3 + 1 + 2 + 4} = \frac{7}{10} = 0.7$$

3. Consider a classifier for which we can have either false positive rate 0 with true positive rate 0.5 or false positive rate 0.5 with true positive rate 1. What is the AUC?

Hint: Draw the curve. Find AUC as a sum (or difference) of rectangular areas.



The classifier's ROC curve goes through $(\text{FPR}, \text{TRR}) = (0, 0.5)$ and $(0.5, 1)$. The area under it is $7/8 = 0.875$.

4. For each situation, indicate whether precision or recall should be optimized:

- (a) A bank is doing fraud detection where a fraudulent transaction ("positive") that is missed is expensive but a valid transaction labeled fraudulent is inexpensive.

☐ Precision

☒ Recall ANSWER: ●

We want low FN and high TP, so we want high recall.

- (b) A doctor is screening patients for a disease in which an ill patient ("positive") infects others and dies if the disease is not diagnosed.

☐ Precision

☒ Recall ANSWER: ●

We want high TP (sick people diagnosed as sick) and low FN (sick people undiagnosed), so we want high recall. (FP are relatively less harmful here.)

- (c) A marketing campaign invests considerable expense in a prospective customer when it classifies that customer as likely to make a purchase ("positive").

☐ Precision ANSWER: ●

We want high TP and low FP, so we want high precision. (FN are just lost opportunities, which is ok with a limited marketing budget.)

☐ Recall