

Precision vs recall, finally clear.

D8A.ACADEMY/RESOURCES/PRECISION-RECALL-F1

THE 4 OUTCOMES

	PREDICTED POSITIVE	PREDICTED NEGATIVE
actually positive	TP · true positive	FN · missed one
actually negative	FP · false alarm	TN · true negative

PRECISION

TP / (TP + FP)	of my alarms, how many were real?
optimize when	false alarms cost you (spam filter)

RECALL

TP / (TP + FN)	of the real ones, how many caught?
optimize when	a miss costs you (cancer screening)

F1 SCORE

2PR / (P + R)	harmonic mean: punishes imbalance
use when	one number needed, classes skewed

THE THRESHOLD DIAL

raise it	precision up, recall down
lower it	recall up, precision down

→ precision asks "was I right?", recall asks "did I catch them all?"

WORTH MEMORIZING

01 Two questions

Every metric argument in ML is one of these two questions.

PRECISION

- Of my positive calls, how many were right?
- Pain: the false alarm
- Spam, fraud review queues

RECALL

- Of the true positives, how many did I find?
- Pain: the missed case
- Screening, safety, search

→ You rarely get both. The threshold decides which one you buy with the other.

02 Trap: 99% accurate, useless

On data with 1% fraud, "always predict legit" scores 99% accuracy and catches nothing.

accuracy	right calls / all calls. Blind to imbalance.
recall here	0%. The model caught zero fraud.
the fix	report precision + recall (or F1), never accuracy alone

→ First question for any classifier: what is the class balance?