Introduction to Machine Learning:
Binary Classification

ECE 580
Spring 2022

Stacy Tantum, Ph.D.

Goal of Classifiers

Correctly classify a previously unseen data instance (with high probability)

Weight [g]	Wingspan [cm]	Webbed Feet?	Back Color	Species
1000.1	125.0	No	Brown	Buteo jamaicensis
3000.7	200.0	No	Gray	Sagittarius serpentarius
4100.0	136.0	Yes	Black	Gavia immer
3.0	11.0	No	Green	Calothorax lucifer
570.0	75.0	No	Black	Campephilius principalis
4.3	14.8	No	Green	??? *Calothorax lucifer*
600.0	80.0	No	Black	??? *Campephilius principalis*
785.0	100.0	No	Dark Brown	??? ???

Goal of Classifiers

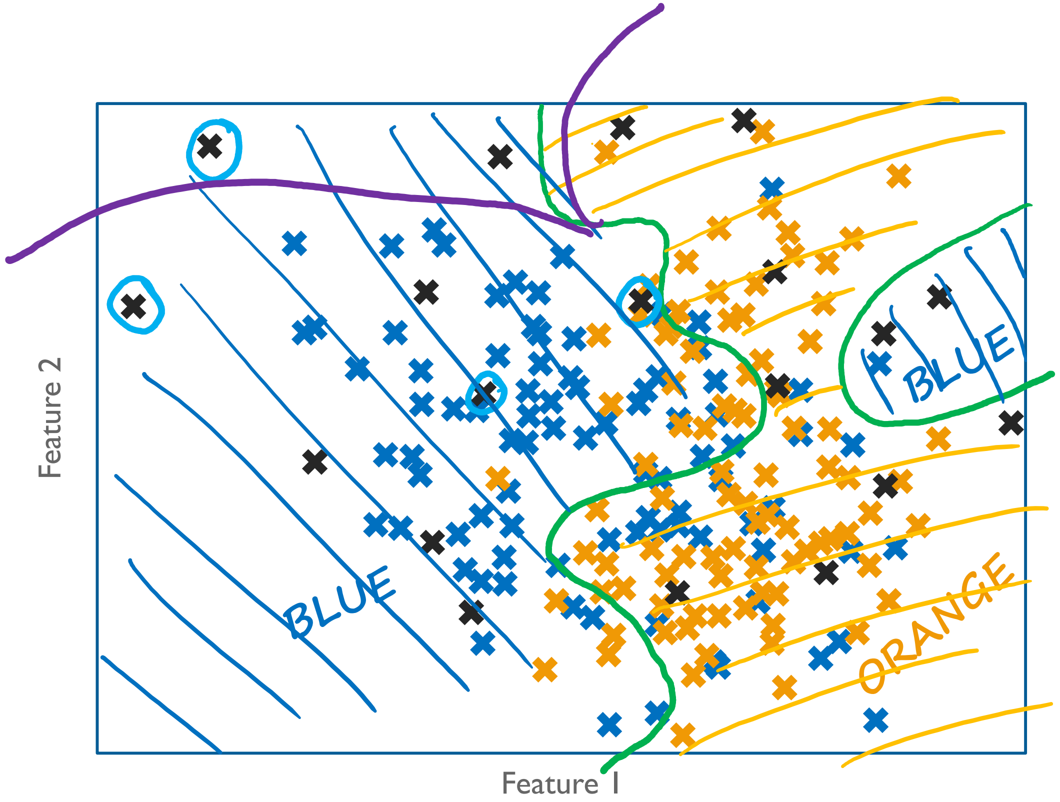
Correctly classify a previously unseen data instance (with high probability)

✖ = class 0

✖ = class 1

✖ = test data

(don't know a priori if truly blue or orange)



$$t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}$$ N target variables (0 or 1)

$$X = \begin{bmatrix} — x_{n=1}^T — \\ — x_{n=2}^T — \\ \vdots \\ — x_{n=N}^T — \end{bmatrix}$$ N observations

D dimensions (D features)

Want majority of test data to fall in region that corresponds to the true class

Classification can be viewed as regression with $t \in \{0,1\}$ *or* $t \in \{-1,+1\}$

Define Your Problem!

Weight [g]	Wingspan [cm]	Webbed Feet?	Back Color	Species
1000.1	125.0	No	Brown	Buteo jamaicensis
3000.7	200.0	No	Gray	Sagittarius serpentarius
4100.0	136.0	Yes	Black	Gavia immer
3.0	11.0	No	Green	Calothorax lucifer
570.0	75.0	No	Black	Campephilius principalis





Black birds vs. non-black birds?

Swimming birds vs. non-swimming birds?

Small birds vs. big birds?

Colorful birds vs. drab birds?

Ivory-billed woodpecker vs. all others?

Ensure what you measure (performance) is meaningful to your goal !

(Binary) Decision Statistics

Classifiers transform the set of D-dimensional features for an observation to a scalar (a **decision statistic**) that forms the basis for making a decision

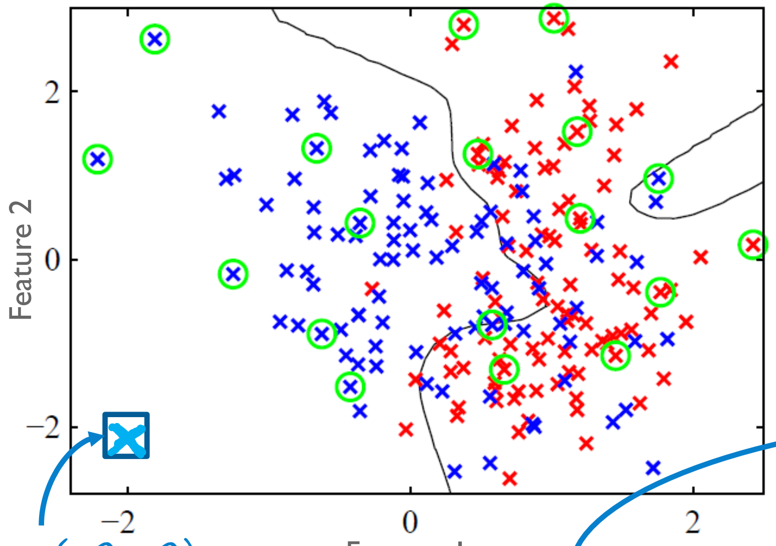
$$\text{decision statistic for an observation } x_n \text{ is } \lambda_n = f(x_{n1}, x_{n2}, \ldots x_{nD})$$

Compare λ to a threshold (a constant = β) to make a decision

$\lambda \geq \beta \rightarrow$ decide H_1 (red)

$\lambda < \beta \rightarrow$ decide H_0 (blue)

Feature Scatter Plot



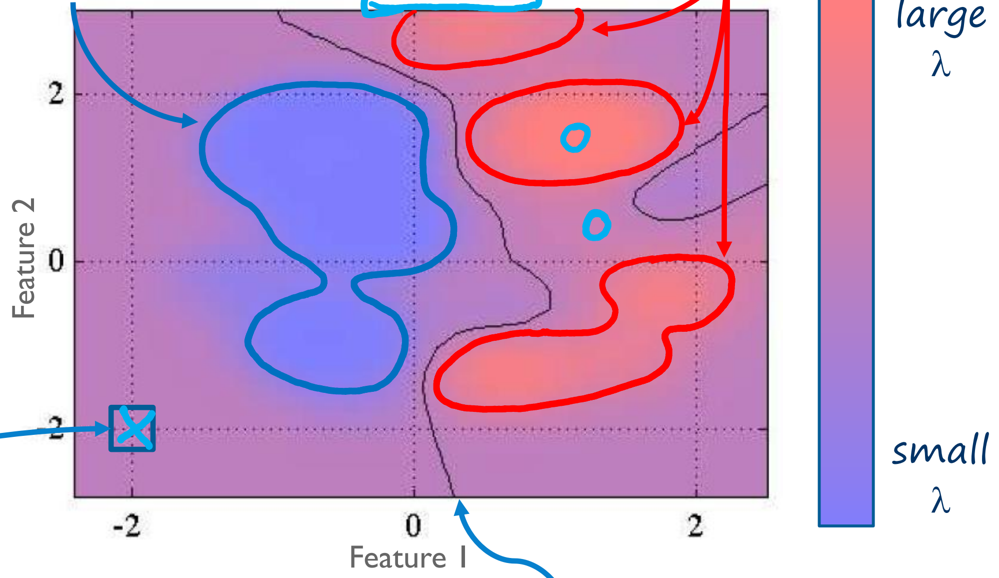
$(x_1, x_2) = (-2, -2)$

color denotes λ for the (x_1, x_2) features at this pixel location [(−2,−2) in this example]

contour for a smaller β

contours for a larger β

Decision $\boxed{\text{Statistic}}$ Surface



decide "blue" for λ<β ← → decide "red" for λ≥β

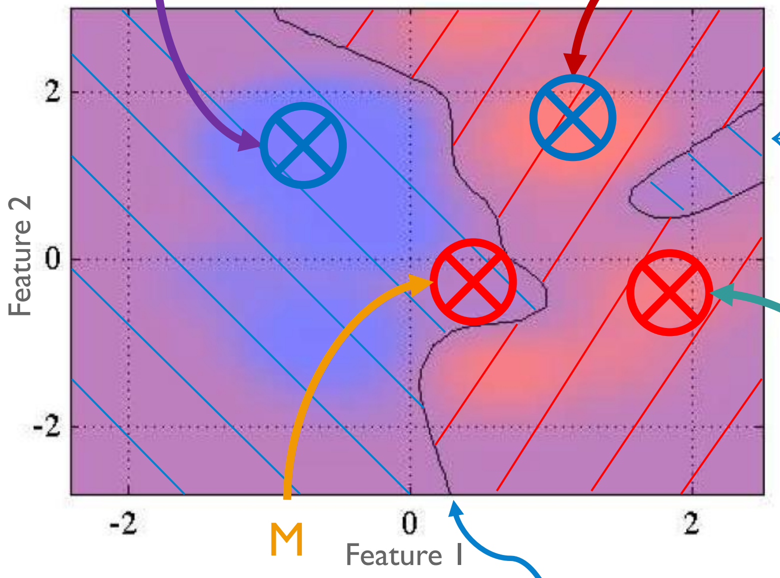
contour for a constant $\boxed{\beta}$

Binary Decision Outcomes: Confusion Matrix (for a **single** threshold β)

Binary Hypotheses

H_0: Data ***does not*** come from the class-of-interest

H_1: Data comes from the class-of-interest



Confusion Matrix (for a single β threshold)

Decision

Truth	H_0 (blue)	H_1 (red)
H_0 (blue)	Decide H_0 when truth is H_0 Correct Reject (CR) True Negative (TN)	Decide H_1 when truth is H_0 False Alarm (FA) False Positive (FP)
H_1 (red)	Decide H_0 when truth is H_1 Miss (M) False Negative (FN)	Decide H_1 when truth is H_1 Detection (D) True Positive (TP)

Probabilities of these 4 possibilities

Sensitivity (or Probability of Detection, P_D)

$$\text{Sensitivity (Recall)} = \text{True Positive Rate} = \text{TPR}$$

$$= \frac{TP}{TP + FN}$$

$$= p(\text{decide } H_1 \mid H_1 \text{ true})$$

$$= \text{probability of detection } (P_D)$$

		Decision	
		H_0	H_1
Truth	H_0	TN	FP
	H_1	FN	TP

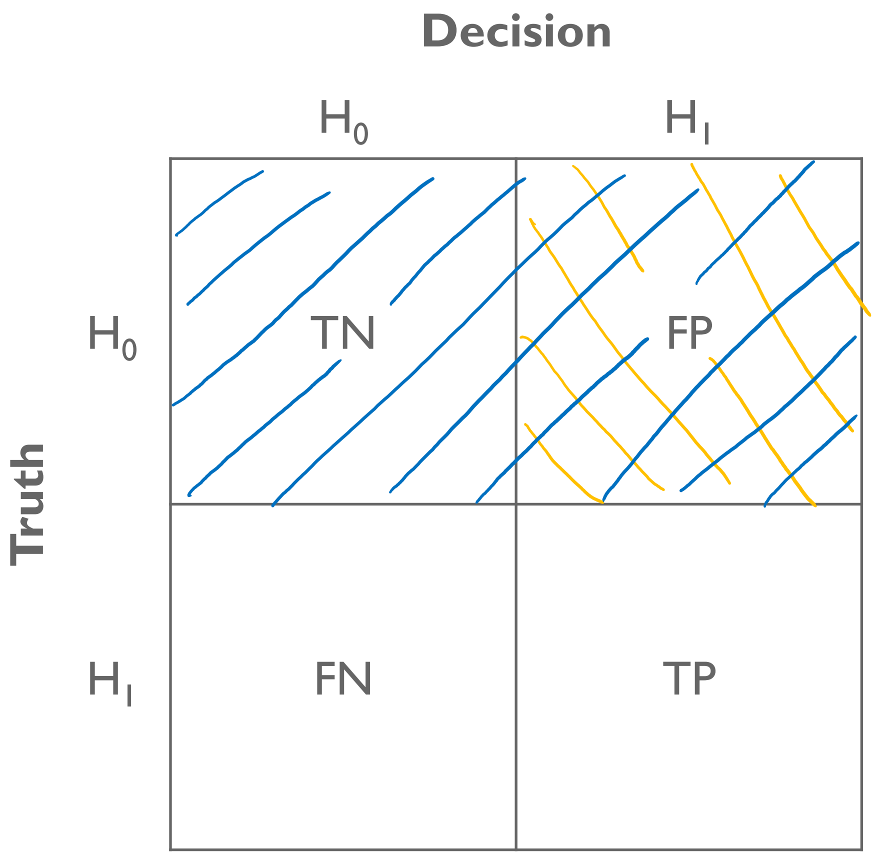
Fall-out (or Probability of False Alarm, P_{FA})

Fall-out = False Positive Rate = FPR

$$= \frac{FP}{FP + TN}$$

$$= p(\text{decide } H_1 \mid H_0 \text{ true})$$

$$= \text{probability of false alarm } (P_{FA})$$



Specificity (or Probability of Correct Rejection, P_{CR})

$$\text{Specificity} = \text{True Negative Rate} = TNR$$

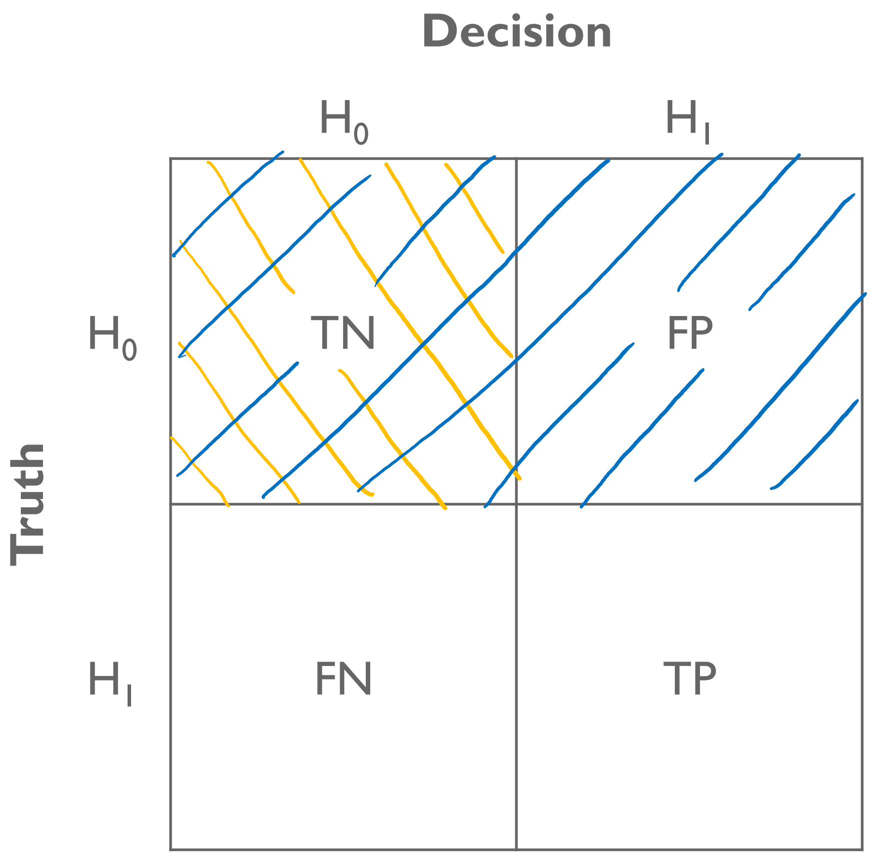
$$= \frac{TN}{TN + FP}$$

$$= p(\text{decide } H_0 \mid H_0 \text{ true})$$

$$= \text{probability of correct rejection } (P_{CR})$$

$$= 1 - \text{Fall-Out}$$

$$= 1 - P_{FA}$$

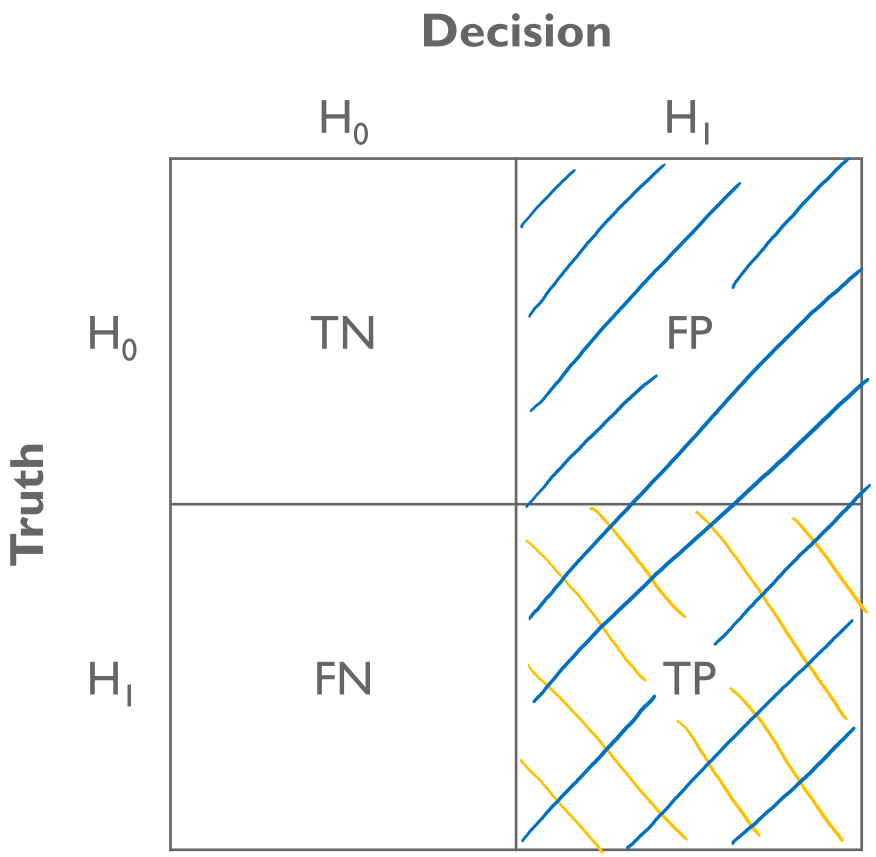


Precision

Precision = Positive Predictive Value = PPV

$$= \frac{TP}{TP + FP}$$

$$= p(H_1 \text{ true} \mid \text{decide } H_1)$$



Accuracy

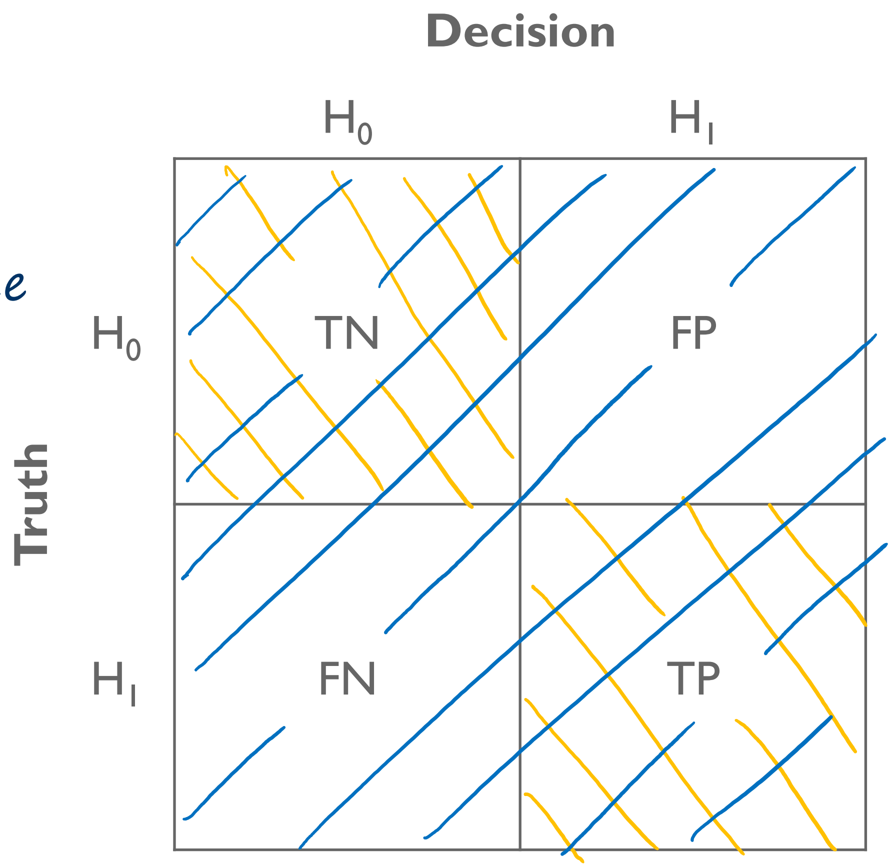
$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$= \text{probability of correct decision } (P_{CD})$$

Accuracy is an overall summary of performance

- Incorporates both types of "right" decisions <u>and</u> both types of "wrong" decisions
- A change in any of TN, FP, FN, TP changes accuracy

→ Accuracy captures <u>**everything**</u>



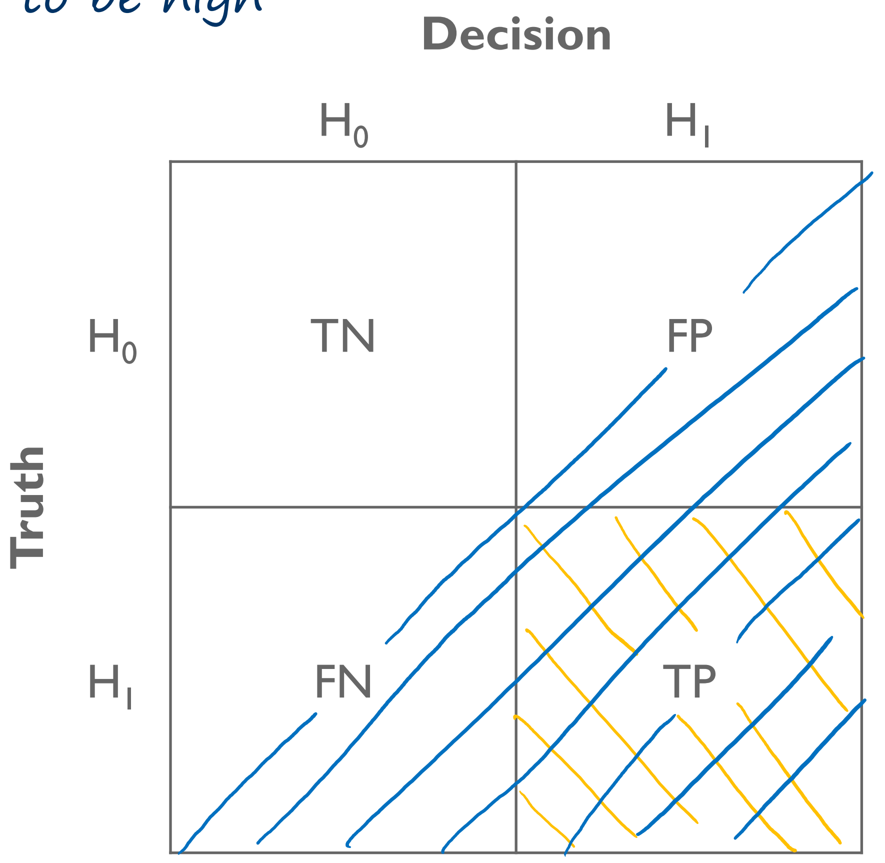
F1 Score

$$\frac{1}{\text{F1 Score}} = \frac{1}{2}\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)$$

→ need both precision **and** recall to be high for F1 to be high

$$F1 = \frac{TP}{TP + \frac{1}{2}\left(FP + FN\right)}$$

$$= \text{probability of ... ?}$$



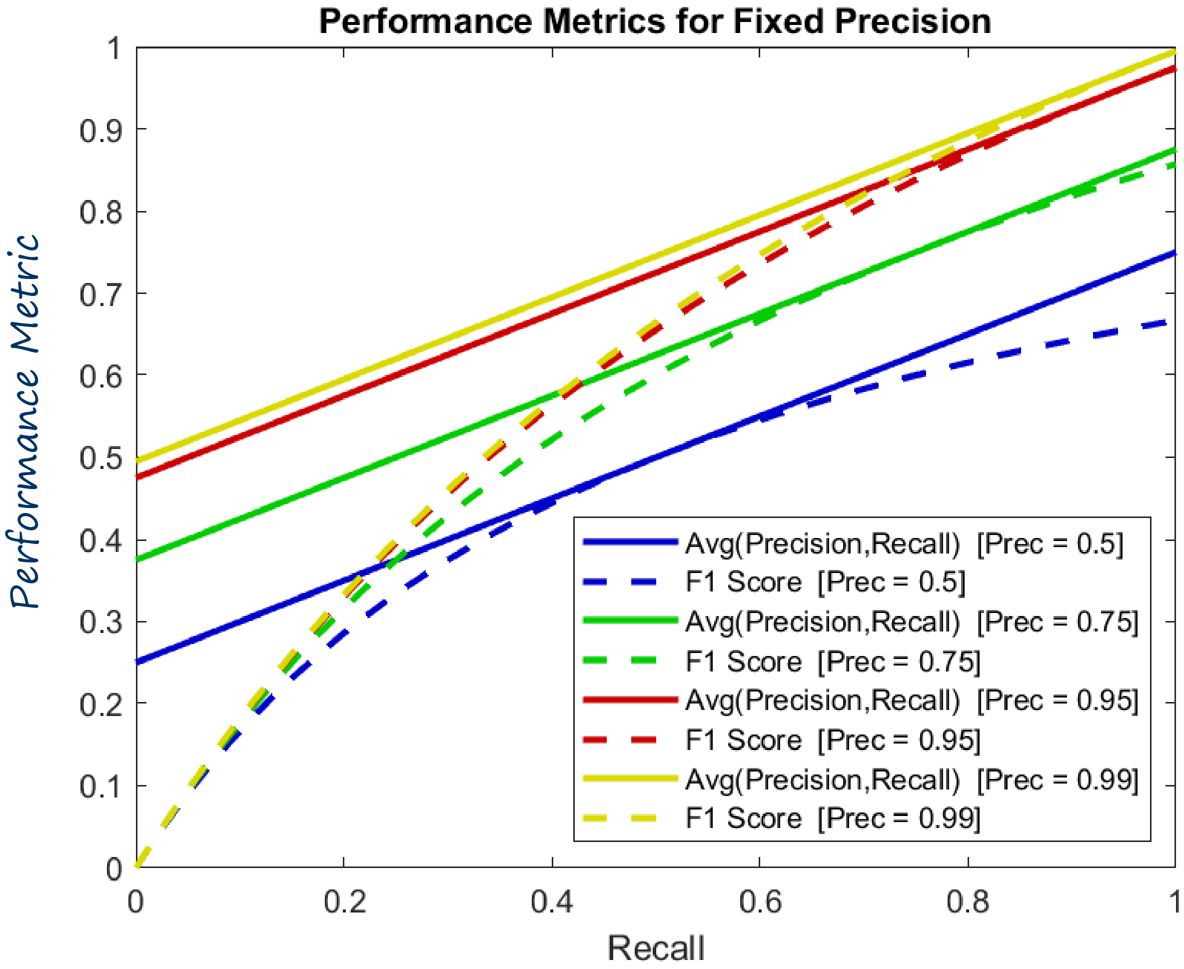
Overall summary of performance

What if TN changes? → F1 unchanged

→ F1 does not capture *everything*

(See Table 3.1 on page 107 in Machine Learning: A Quantitative Approach)

F1 Score vs Averaging Precision and Recall



F1 Score has a "stronger" trade-off between precision [$p(H_1 \text{ true} | \text{decide } H_1)$] and recall [$p(\text{decide } H_1 | H_1 \text{ true})$]