4.03 Classification Metrics I
4.04 Classification Metrics II

Data Science Process

1. Define the problem
2. Gather data
3. Explore data
4. Model with data
5. Evaluate model
6. Answer problem

Framing

Remember the regression metrics lesson from last week, where we explored different methods for evaluating the performance of **regression models**.

We'll do the same thing today, but for **classification models**.

- In regression, we quantify the performance of our model by comparing predicted and observed values in some capacity.
- We'll do the same thing in classification... but predicted and observed are categories, so it's slightly different.

We're going to focus on **binary classification problems**.

Evaluating Our Model

Suppose you build a logistic regression model predicting whether or not people will vote. You make predictions for 100 people, then check to see if they actually voted.

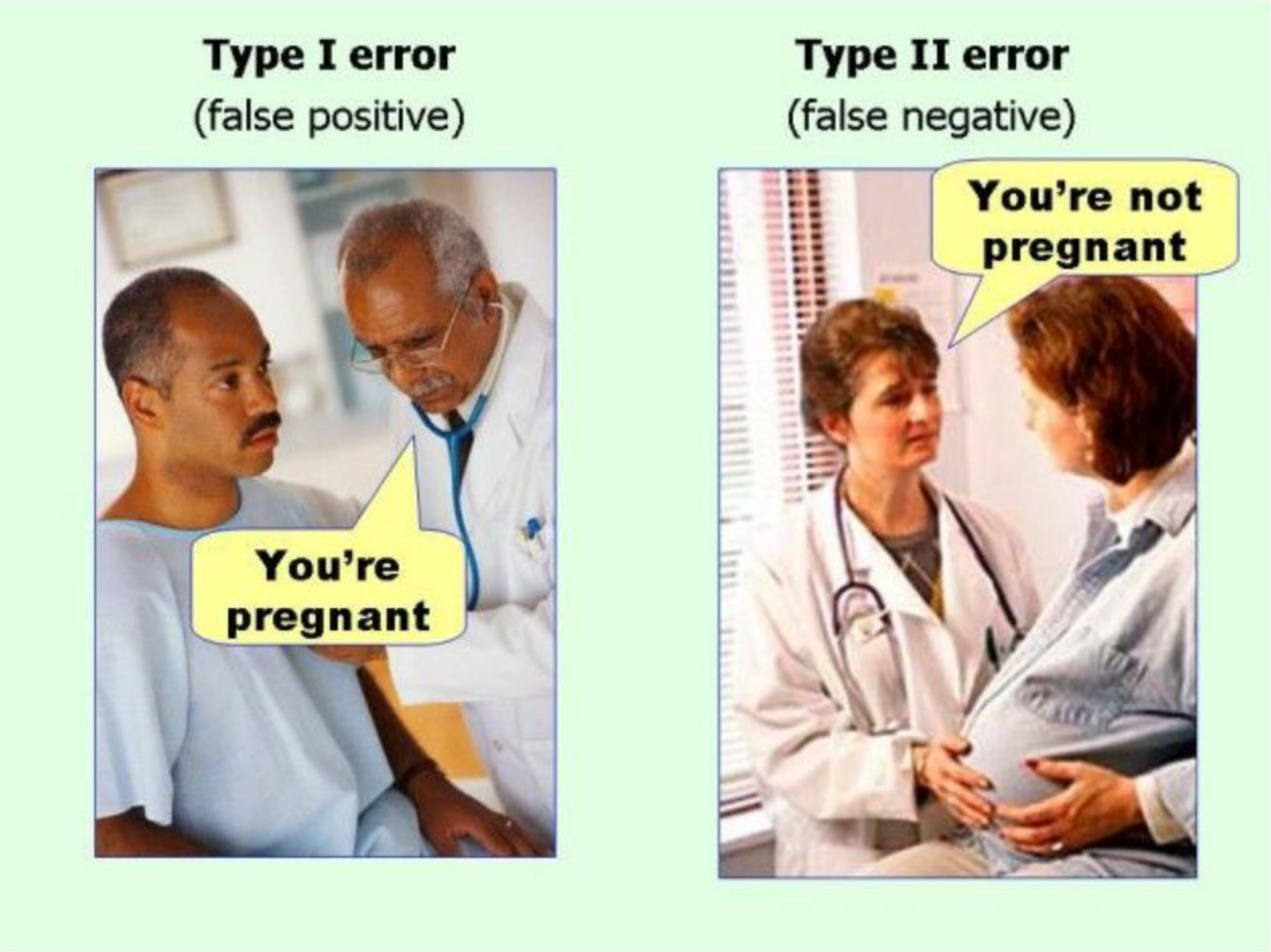
- There are 40 people you predicted to vote who did vote.
- There are 20 people you predicted to vote who didn’t vote.
- There are 15 people you predicted to stay home who did vote.
- There are 25 people you predicted to stay home who didn’t vote.

Evaluating Our Model

Suppose you build a logistic regression model predicting whether or not people will vote. You make predictions for 100 people, then check to see if they actually voted.

- There are 40 people you predicted to vote who did vote.
  - These are called **true positives**.
- There are 20 people you predicted to vote who didn't vote.
  - These are called **false positives**.
- There are 15 people you predicted to stay home who did vote.
  - These are called **false negatives**.
- There are 25 people you predicted to stay home who didn't vote.
  - These are called **true negatives**.

Evaluating Our Model



Evaluating Our Model

How do I keep true positives/true negatives/false positives/false negatives straight?

- First word (true/false): Was I right?
- Second word (positive/negative): What did I predict?

Evaluating Our Model

How do I keep true positives/true negatives/false positives/false negatives straight?

- First word (true/false): Was I right?
- Second word (positive/negative): What did I predict?

What is it called if I correctly predicted that someone does not vote?

Evaluating Our Model

How do I keep true positives/true negatives/false positives/false negatives straight?

- First word (true/false): Was I right?
- Second word (positive/negative): What did I predict?

What is it called if I incorrectly predicted that someone does vote?

Confusion Matrix

It's helpful for us to list out the number of each category in a 2x2 grid called a **confusion matrix**.

	Actual Positive	Actual Negative
Predicted Positive		
Predicted Negative		

The axes or ordering of "Yes" vs. "No" may be rearranged!

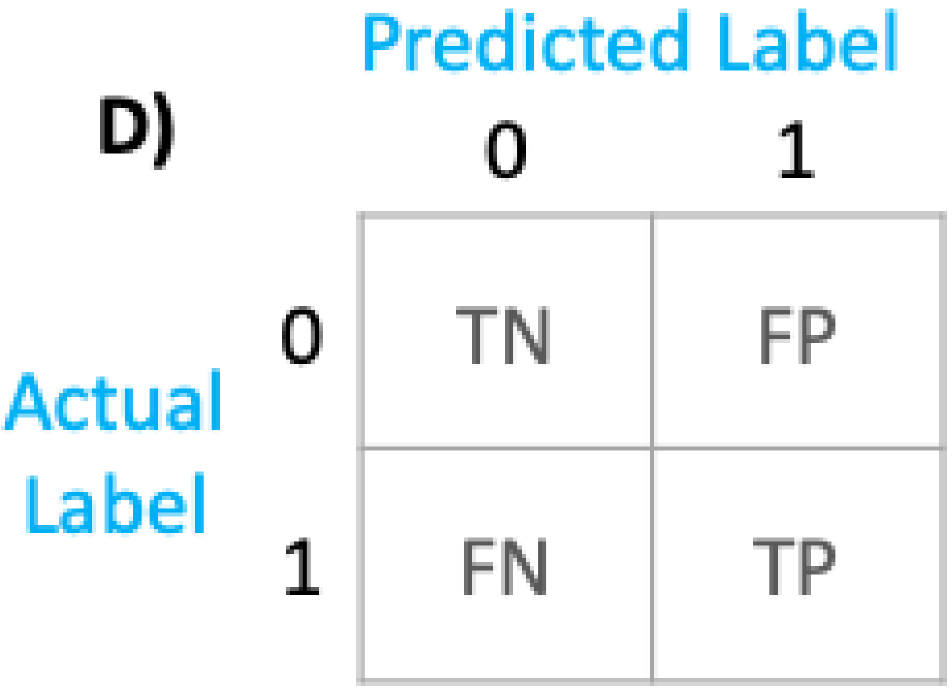
Be clear what "Yes" / "Positive" means.

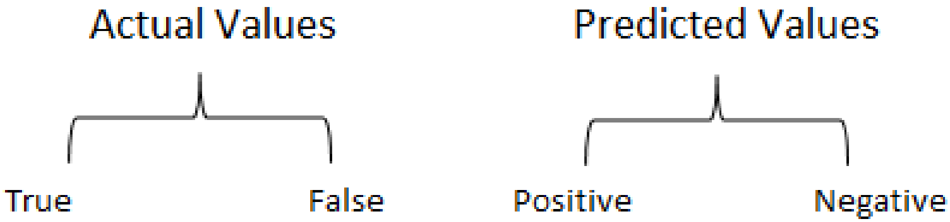
Confusion Matrix

- Available in Scikit-Learn Metrics Function
- A convenient way to visualise Classification Models' performance
- This matrix will reveal how well our predictions line up with the actuals across the positive and negative subsets of our data

Confusion Matrix

D)





Confusion Matrix

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Interpretations of TP, FP, FN, TN

- **True Positive:**
  - Interpretation: You predicted positive and it’s true
- **True Negative:**
  - Interpretation: You predicted negative and it’s true
- **False Positive: (Type 1 Error)**
  - Interpretation: You predicted positive and it’s false
- **False Negative: (Type 2 Error)**
  - Interpretation: You predicted negative and it’s false

Complementing Confusion Matrix

- Confusion Matrix is a convenient way to visualise model performance

- However, there are metrics to summarise model performance with a single number. These include:
  - Accuracy
  - Misclassification Rate
  - Sensitivity
  - Specificity
  - Precision

Accuracy

Interpretation: What percentage of observations did I **correctly** predict?

$$\text{Accuracy} = \frac{\textit{All Correct}}{\textit{All Predictions}} = \frac{TP + TN}{TP + FP + TN + FN}$$

	Actual Positive	**Actual Negative**
Predicted Positive	40	20
Predicted Negative	15	25

Misclassification Rate

Interpretation: What percentage of observations did I **incorrectly** predict?

$$\text{Misclassification Rate} = \frac{\textit{All Incorrect}}{\textit{All Predictions}} = \frac{FP + FN}{TP + FP + TN + FN} = 1 - \text{Acc}$$

	Actual Positive	**Actual Negative**
Predicted Positive	40	20
Predicted Negative	15	25

Sensitivity

Interpretation: Among those who will vote, how many did I get correct?

$$\text{Sensitivity} = \frac{\textit{True Positives}}{\textit{All Positives}} = \frac{TP}{TP + FN} = \frac{TP}{P}$$

a.k.a. True Positive Rate, Recall

	Actual Positive	**Actual Negative**
Predicted Positive	40	20
Predicted Negative	15	25

Specificity

Interpretation: Among those who will not vote, how many did I get correct?

$$\text{Specificity} = \frac{\textit{True Negatives}}{\textit{All Negatives}} = \frac{TN}{TN + FP} = \frac{TN}{N}$$

a.k.a. True Negative Rate

	Actual Positive	**Actual Negative**
Predicted Positive	40	20
Predicted Negative	15	25

Precision

Interpretation: Among those I predicted to vote, how many did I get correct?

$$\text{Precision} = \frac{\textit{True Positives}}{\textit{Predicted Positives}} = \frac{TP}{TP + FP}$$

a.k.a. Positive Predictive Value

	Actual Positive	**Actual Negative**
Predicted Positive	40	20
Predicted Negative	15	25

Classification Metrics Summary

- **Accuracy** = (True Positives + True Negatives) / Total Predictions
- **Precision** = True Positives / (True Positives + False Positives)
- **Recall** = True Positives / (True Positives + False Negatives)
- **F1-Score**
  - (2 * Recall * Precision) / (Recall + Precision)
  - Offers a better overall measure of performance
- **Support** = True Positives (or True Negatives) that lie in that class

Example

Suppose I'm working on a fraud analytics team and our goal is to detect fraudulent credit card transactions. I take a random sample of 500 transactions. Of these transactions, 50 are fraudulent. I predict 100 total fraudulent transaction, 45 of which are correct.

1. Identify the TP, TN, FP, FN and construct a confusion matrix.
2. Calculate the Classification metrics found on the previous slides

Example

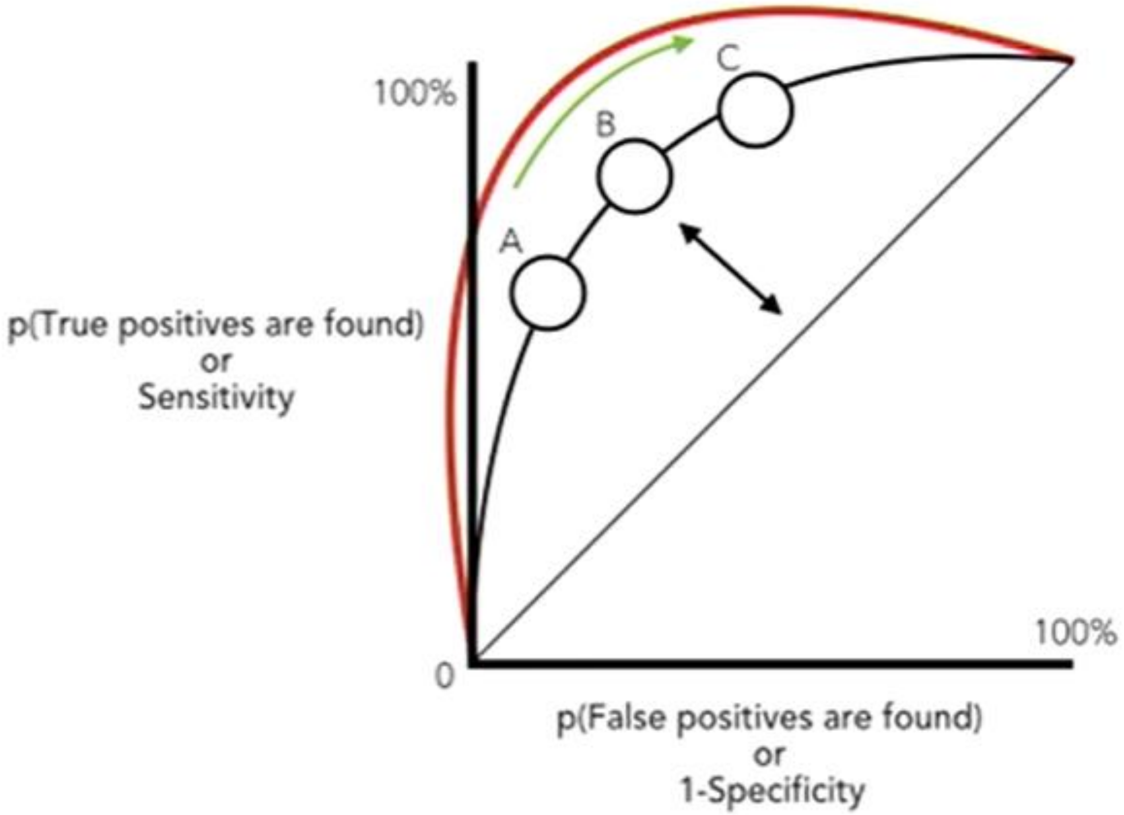
Suppose I'm working on a fraud analytics team and our goal is to detect fraudulent credit card transactions. I take a random sample of 500 transactions. Of these transactions, 50 are fraudulent. I predict 100 total fraudulent transaction, 45 of which are correct.

When building my classification model, I want to optimize one of the above metrics. Given the use-case of identifying fraudulent transactions, which metric should I optimize as I build my model?

Area Under The Curve (AUC)

- Each logistic regression model produces an AUC

  Perfect score = 1.0



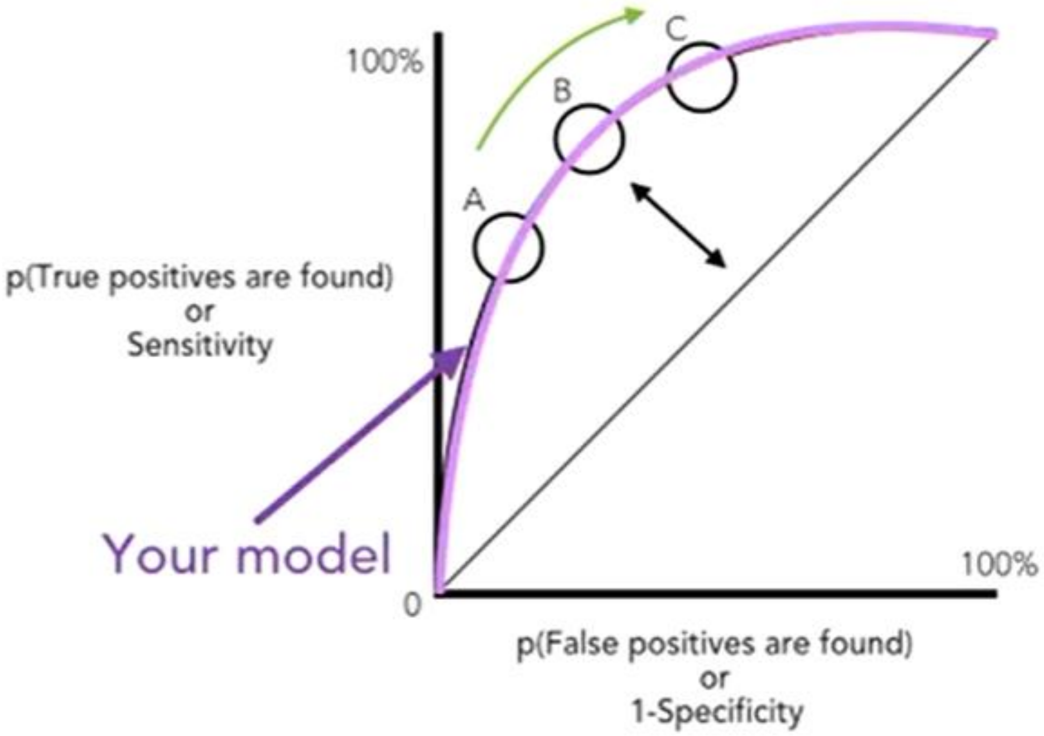
It is the plot between the TPR(y-axis) and FPR(x-axis). Since our model classifies the patient as having heart disease or not based on the probabilities generated for each class, we can decide the threshold of the probabilities as well.

Area Under The Curve (AUC)

- Each logistic regression model produces an AUC

  Perfect score = 1.0

- Your model is plotted, and the area under the curve is measured



Area Under The Curve (AUC)

- Each logistic regression model produces an AUC

  Perfect score = 1.0

- Your model is plotted, and the area under the curve is measured

  Will be measure between 0.5 and 1.0

  The higher, the better