Classification Metrics I

GENERAL ASSEMBLY

Data Science Process

1. Define the problem
2. Gather data
3. Explore data
4. Model with data
5. Evaluate model ← classification metrics
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

1 = positive class
Ø = negative class

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

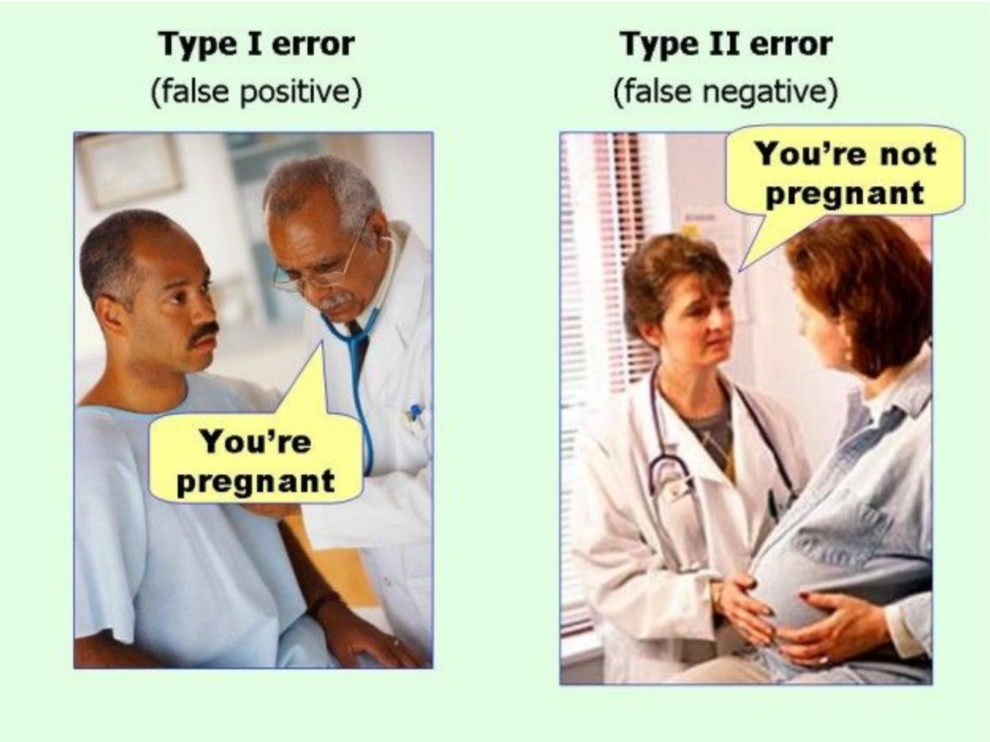
Suppose you build a logistic regression model predicting whether or not people will vote. You make predictions for 100 people, then check to see if they actually voted.

- 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

Type I

False Positive



Type II

False Negative

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?

true positive

false negative

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?

true negative

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?

False positive

Type I

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	40	20
Predicted Negative	15	25

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

A confusion matrix is a convenient way for us to visualize how our model performs.

However, there are metrics that can help us to summarize performance with one number.

- Accuracy
- Misclassification Rate
- Sensitivity
- Specificity
- Precision

Accuracy

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

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

65%

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

$$\frac{40 + 25}{40 + 25 + 20 + 15}$$

.65

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

$1 - .65$

$\boxed{.35}$

Sensitivity

There is "p" in the word

Only "p"'s in the math

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

$$\text{Sensitivity} = \frac{\textit{True Positives}}{\textit{All Positives}} = \frac{TP}{TP + FN} = \boxed{\frac{TP}{4\ P}}$$

$\boxed{.72}$

$$\frac{40}{40+15}$$

a.k.a. True Positive Rate, $\boxed{\text{Recall}}$

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

Specificity

There is a "p" in the word
No p's in the math

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

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

$\boxed{.56}$

$$\frac{25}{25+20}$$

a.k.a. True Negative Rate

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

Precision

percentage of correct positive predictions

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

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

$$\frac{40}{40 + 20} \quad \boxed{.67}$$

a.k.a. Positive Predictive Value

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

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 accuracy, misclassification rate, positive predictive value, recall, and true negative rate.

	Pred Neg	Pred Pos
Actual Neg	395 TN	FP 55
Actual Pos	5 FN	TP 45

Accuracy: $\frac{\text{Correct}}{\text{Total}}$ $\frac{(395+45)}{500} = .88$

Misclassification: $1 - \text{Acc} = .12$

Sensitivity: $\frac{TP}{\text{Actual } P} = \frac{45}{45+5} = .9$

Specificity: $\frac{TN}{AN} = \frac{395}{450} = \sim .878$

Precision: $\frac{TP}{TP+FP} = \frac{45}{100} = \boxed{.45}$

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?

Final Notes

We explored binary classification problems today.

We can construct confusion matrices for 3+ categories and calculate a lot of these metrics (accuracy, misclassification error, etc.), but they get a lot more complicated.

These get *especially* complicated when working with **ordinal data**.

