GENERAL ASSEMBLY

CLASSIFICATION METRICS I

Matt Brems

Data Science Immersive

DATA SCIENCE PROCESS

1. Define 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

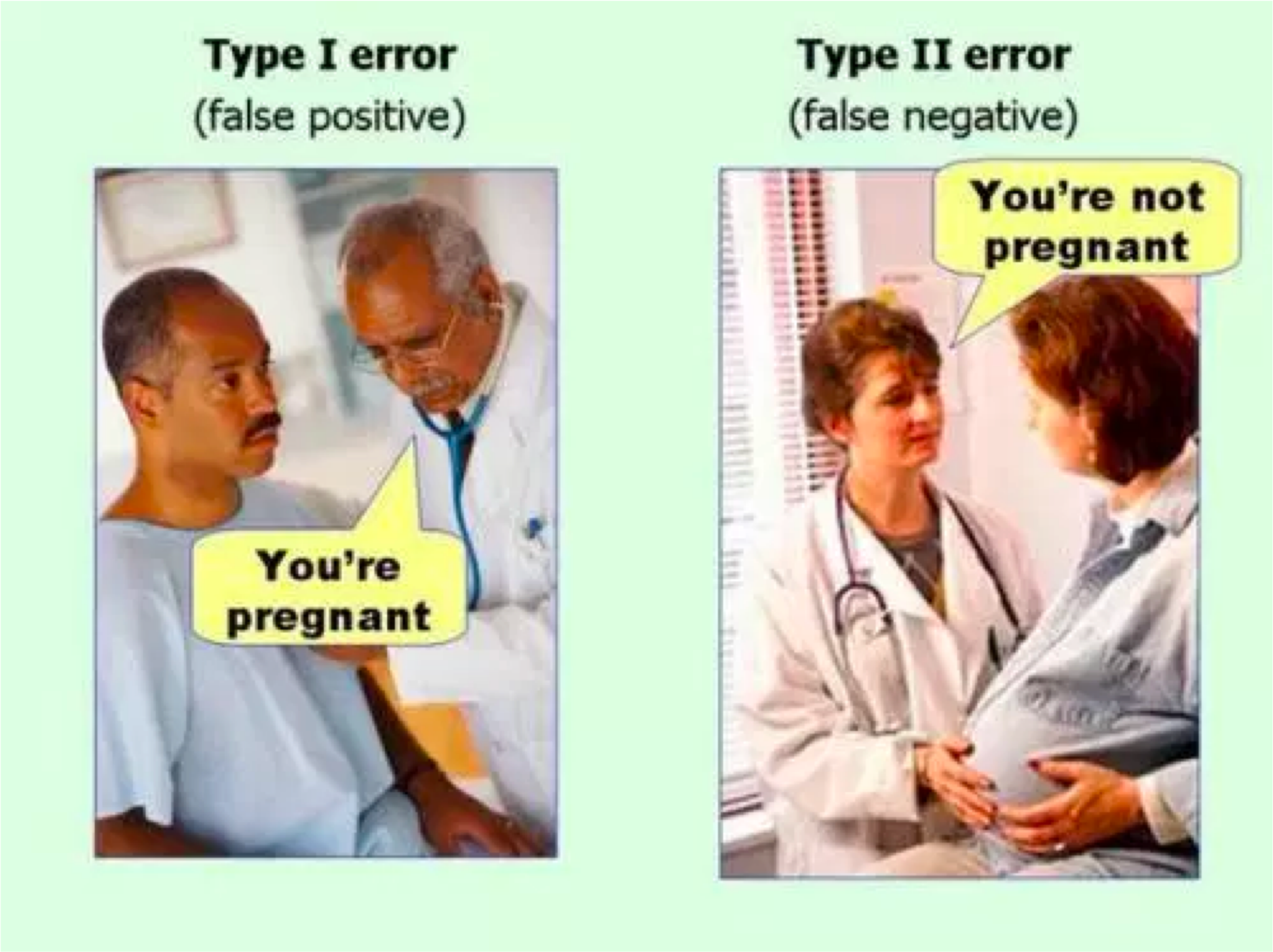
- Suppose you build a logistic regression model predicting whether or not people will vote. You make predictions for 100 people, then check to see if they actually voted.
    - There are 40 people you predicted to vote who did vote.
    - There are 20 people you predicted to vote who didn’t vote.
    - 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: Was I right?
  - Second word: What did I predict?

EVALUATING OUR MODEL

- How do I keep true positives/true negatives/false positives/false negatives straight?
  - First word: Was I right?
  - Second word: 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: Was I right?
  - Second word: 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		

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.

CLASSIFICATION METRICS

- 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

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

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

	ACTUAL POSITIVE	ACTUAL NEGATIVE
PREDICTED POSITIVE	40	20
PREDICTED NEGATIVE	15	25

MISCLASSIFICATION RATE

$$\text{Misclassification Rate} = \frac{all\ incorrect}{all} = \frac{FN + FP}{TP + FN + FP + TN} = 1 - \text{Acc}$$

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

	ACTUAL POSITIVE	ACTUAL NEGATIVE
PREDICTED POSITIVE	40	20
PREDICTED NEGATIVE	15	25

SENSITIVITY

$$\text{Sensitivity} = \frac{true\ positives}{all\ positives} = \frac{TP}{TP + FN} = \frac{TP}{P}$$

- Interpretation: Among those who will vote, how many did I get correct?
- a.k.a. True Positive Rate, Recall

	ACTUAL POSITIVE	ACTUAL NEGATIVE
PREDICTED POSITIVE	40	20
PREDICTED NEGATIVE	15	25

SPECIFICITY

$$\text{Specificity} = \frac{true\ negatives}{all\ negatives} = \frac{TN}{TN + FP} = \frac{TN}{N}$$

- Interpretation: Among those who will not vote, how many did I get correct?
- a.k.a. True Negative Rate

	ACTUAL POSITIVE	ACTUAL NEGATIVE
PREDICTED POSITIVE	40	20
PREDICTED NEGATIVE	15	25

PRECISION

$$\text{Precision} = \frac{true\ positives}{predicted\ positives} = \frac{TP}{TP + FP}$$

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

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?

BRIER SCORE

- One alternative when you have predicted probabilities:

$$\text{Brier score} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

- i.e. I have three observations.
  - A is predicted to have a 90% probability of voting and votes.
  - B is predicted to have a 50% probability of voting and votes.
  - C is predicted to have a 30% probability of voting and doesn’t vote.

- The Brier score is the MSE of our forecasts!

WRAP-UP

- 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**.