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.

Regression: Y is <u>continuous</u>.

↳ Predict the price of a house. $[0, \infty)$

↳ Estimate someone's height. $[0, \infty)$

↳ Estimate the probability that someone shows up to vote. $[0, 1]$

Classification: Y is <u>discrete</u>.

↳ Will a Reddit post go viral? $\{Y, N\}$

↳ Which day of the week sees highest sales?

↳ Will someone vote? $\{0, 1\}$

FRAMING

- Remember the regression metrics lesson from last week, where we explored different methods for evaluating the performance of **regression models**. R^2, R^2_{adj}, MSE, etc.

- 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. obs. pred. $y - \hat{y}$
  - 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**. two-class output

EVALUATING OUR MODEL

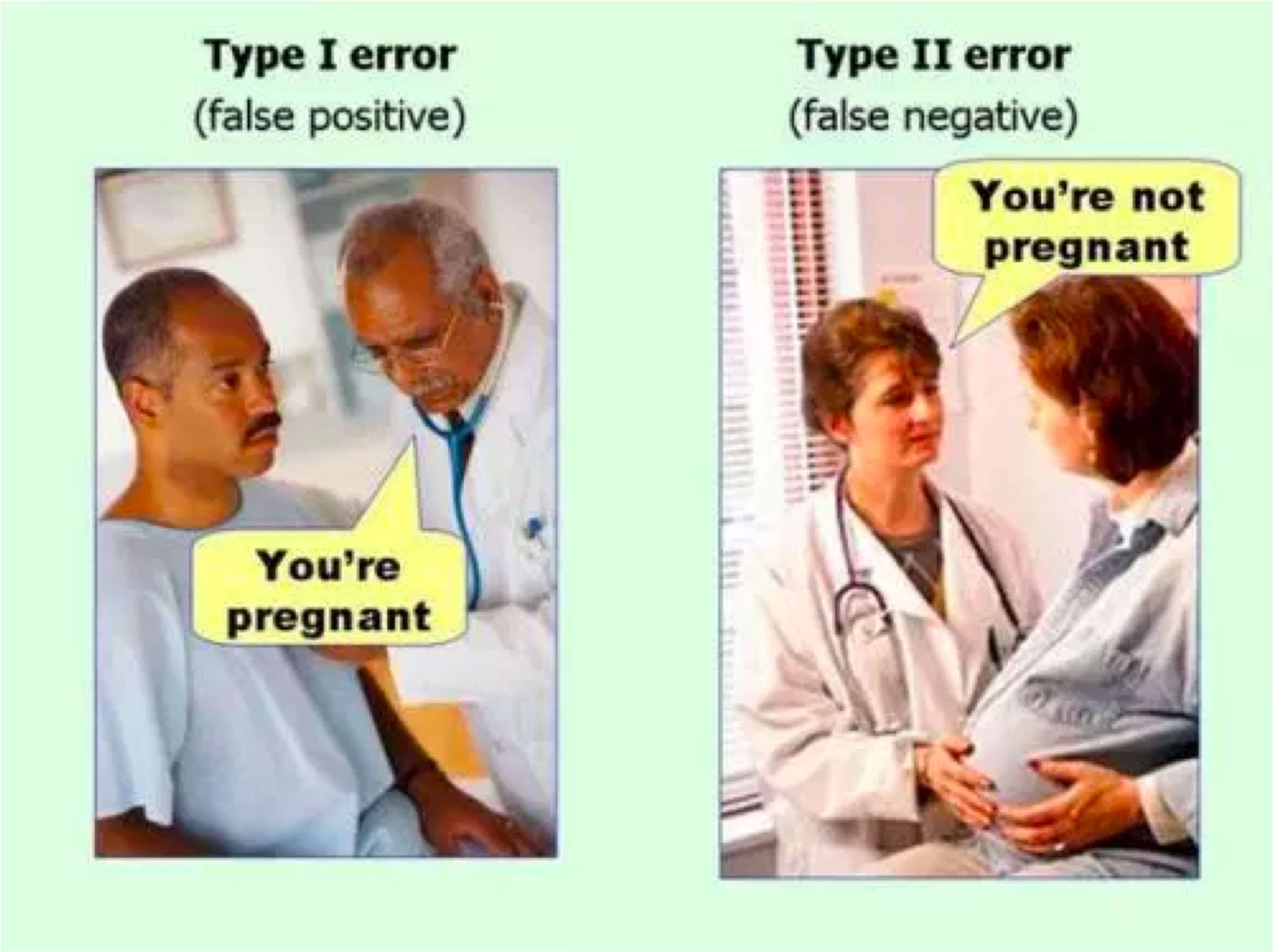
- Suppose you build a logistic regression model predicting whether or not people will vote. You make predictions for 100 people, then check to see if they actually voted.
  - There are 40 people you predicted to vote who did vote.
  - There are 20 people you predicted to vote who didn't vote.
  - There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

T/F	P/N
true = correct	pred. pos. = positive
false = incorrect	pred. neg. = negative

How do I know which is "positive?"

1) Context. (Usually the thing of interest.)

2) Be explicit where possible!

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?

positive = vote

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

T ✓ N ✓

right? predict?

True Negative

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?

FALSE POSITIVE

right? predict?

CONFUSION MATRIX

binary classes

- It's helpful for us to list out the number of each category in a 2x2 grid called a **confusion matrix**. → what we predict vs. what we observe

	ACTUAL POSITIVE	ACTUAL NEGATIVE
PREDICTED POSITIVE	40 TP	FP 20
PREDICTED NEGATIVE	15 FN	TN 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.

fraud detection
medical testing

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
  - Brier Score

ACCURACY

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

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

$$Acc = \frac{40+25}{100}$$

$$= \frac{65}{100}$$

$$= 65\%$$

	ACTUAL POSITIVE	ACTUAL NEGATIVE
PREDICTED POSITIVE	40 TP	FP 20
PREDICTED NEGATIVE	15 FN	TN 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?

$$MR = \frac{15 + 20}{100}$$

$$= 35\%$$

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

mnemonic

SENSITIVITY

$Sens = \frac{TP}{P}$

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

actual

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

$$Sens = \frac{TP}{P} = \frac{TP}{TP+FN} = \frac{40}{40+15} \approx 73\%$$

	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

$$Spec = \frac{TN}{N}$$

$$= \frac{TN}{TN+FP} = \frac{25}{25+20} \approx 56\%$$

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

$$\text{Sensitivity} = \frac{TP}{P}$$

Interpretation: "Of all who really are positive, how often was I right?"

$$\text{Specificity} = \frac{TN}{N}$$

Interpretation: "Of all who really are negative, how often was I right?"

PRECISION

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

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

- a.k.a. Positive Predictive Value

$$\text{Prec} = \frac{TP}{TP+FP} = \frac{40}{40+20} = 66.67\%$$

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

EXAMPLE

positive = fraudulent transactions

- 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.

	Predicted + fraud!	Pred − legit	
Actual +	45 TP	FN 5	50
Actual −	FP 55	TN 395	450
	100	400	500

$$TP = 45,\quad TN = 395,\quad FP = 55,\quad FN = 5$$

$$Acc = \frac{TP + TN}{all} = \frac{45 + 395}{500}$$

$$MR = 1 - Acc = 1 - \frac{45 + 395}{500} = \frac{60}{500}$$

$$Sens = \frac{TP}{P} = \frac{TP}{TP + FN} = \frac{45}{45 + 5} = \frac{45}{50}$$

$$Spec = \frac{TN}{N} = \frac{TN}{TN + FP} = \frac{395}{395 + 55} = \frac{395}{450}$$

$$Prec = \frac{TP}{TP + FP} = \frac{45}{45 + 55} = \frac{45}{100}$$

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?

False Positive: transaction predicted to be fraudulent, but was legitimate

→ halt the transaction

→ customer service

↳ what is the cost/badness of FP?

False Negative: transaction predicted to be legitimate, but was fraudulent

→ card was improperly used

→ $$$

↳ what is cost/badness of FN?

goal: minimize false negatives

$\hookrightarrow$ optimize sensitivity

$$\text{Sens} = \frac{TP}{P} = \frac{TP}{TP + \underbrace{FN}_{\to 0}}$$

1 if Sens = 100%

$\Rightarrow TP = TP + FN$

$\Rightarrow 0 = FN$

BRIER SCORE

- One alternative when you have predicted probabilities:

predicted probabilities

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

observed values

- 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

$$\text{Spec} = \frac{TN}{N}$$

- 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**.

Not Likely   Somewhat Likely   Very Likely

pred A ——————————— ~~2~~ 1.5 ——— actual A

pred B ——— 1 ——— actual B