GENERAL ASSEMBLY

COMMUNICATING RESULTS

Mason Gallo

Data Scientist

LEARNING OBJECTIVES

- Explain the trade-offs between the precision and recall of a model while articulating the cost of false positives vs. false negatives
- Describe the difference between visualization for presentations vs. exploratory data analysis
- Identify the components of a concise, convincing report and how they relate to specific audiences/stakeholders

OPENING

COMMUNICATING RESULTS

WE BUILT A MODEL! NOW WHAT?

- We’ve built our model, but there is still a gap between your Notebook with plots/figures and a slideshow needed to present your results.

- Classes so far have focused on two core concepts:
    - developing consistent practices
    - interpreting metrics to evaluate and improve model performance

- But what does that mean to your audience?

WE BUILT A MODEL! NOW WHAT?

- Imagine how a non-technical audience might respond to the following statements:
    - The predictive model I built has an accuracy of 80%.
    - Logistic regression was optimized with L2 regularization.
    - Gender was more important than age in the predictive model because it has a larger coefficient.

WE BUILT A MODEL! NOW WHAT?

- Who is your audience? Are they technical? What are their concerns?
- Remember: in a business setting, you may be *the only person* who can interpret what you’ve built.
- Some people may be familiar with basic visualization, but you will likely have to do a lot of “hand holding”.
- You need to be able to efficiently explain your results in a way that makes sense to **all** stakeholders (technical or not).

WE BUILT A MODEL! NOW WHAT?

- Today, we'll focus on communicating results for "simpler" problems, but this applies to any type of model you may work with.

- First, let's talk classification metrics, review our knowledge, and talk about how we might communicate what we know.

REVIEW OR INTRODUCING

THE CONFUSION MATRIX

BACK TO THE CONFUSION MATRIX

- Confusion matrices allow for the interpretation of correct and incorrect predictions for *each class label*.

- It is the first step for the majority of classification metrics and goes deeper than just accuracy.

BACK TO THE CONFUSION MATRIX

Hypothesized class \ True class	p	n
Y	True Positives	False Positives
N	False Negatives	True Negatives
Column totals:	P	N

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

ACTIVITY: KNOWLEDGE CHECK

EXERCISE

ANSWER THE FOLLOWING QUESTIONS

1. What’s the intuition and uses for the following?
    a. Accuracy
    b. True positive rate
    c. False positive rate

DELIVERABLE

Answers to the above questions

PRECISION AND RECALL

PRECISION AND RECALL

- Our previous metrics were primarily designed for less biased data problems: we could be interested in both outcomes, so it was important to generalize our approach.

- For example, we may be interested if a person will vote for a Republican or Democrat. This is a binary problem, but we're interested in both outcomes.

PRECISION AND RECALL

- Precision and recall, metrics built from the confusion matrix, focus on *information retrieval*, particularly when one class is more interesting than the other.

- For example, we may want to predict if a person will be a customer. We care much more about people who will be a customer of ours than people who won't.

PRECISION AND RECALL

- *Precision* aims to product a high amount of relevancy instead of irrelevancy.
- Precision asks, “Out of all of our positive predictions (both true positive and false positive), how many were correct?”
- *Recall* aims to see how well a model returns specific data (literally, checking whether the model can *recall* what a class label looked like).
- Recall asks, “Out of all of our positive class labels, how many were correct?”

ACTIVITY: KNOWLEDGE CHECK

EXERCISE

ANSWER THE FOLLOWING QUESTIONS

1. If the goal of the “recall” metric is to identify specific values of a class correctly, what other metric performs a similar calculation?

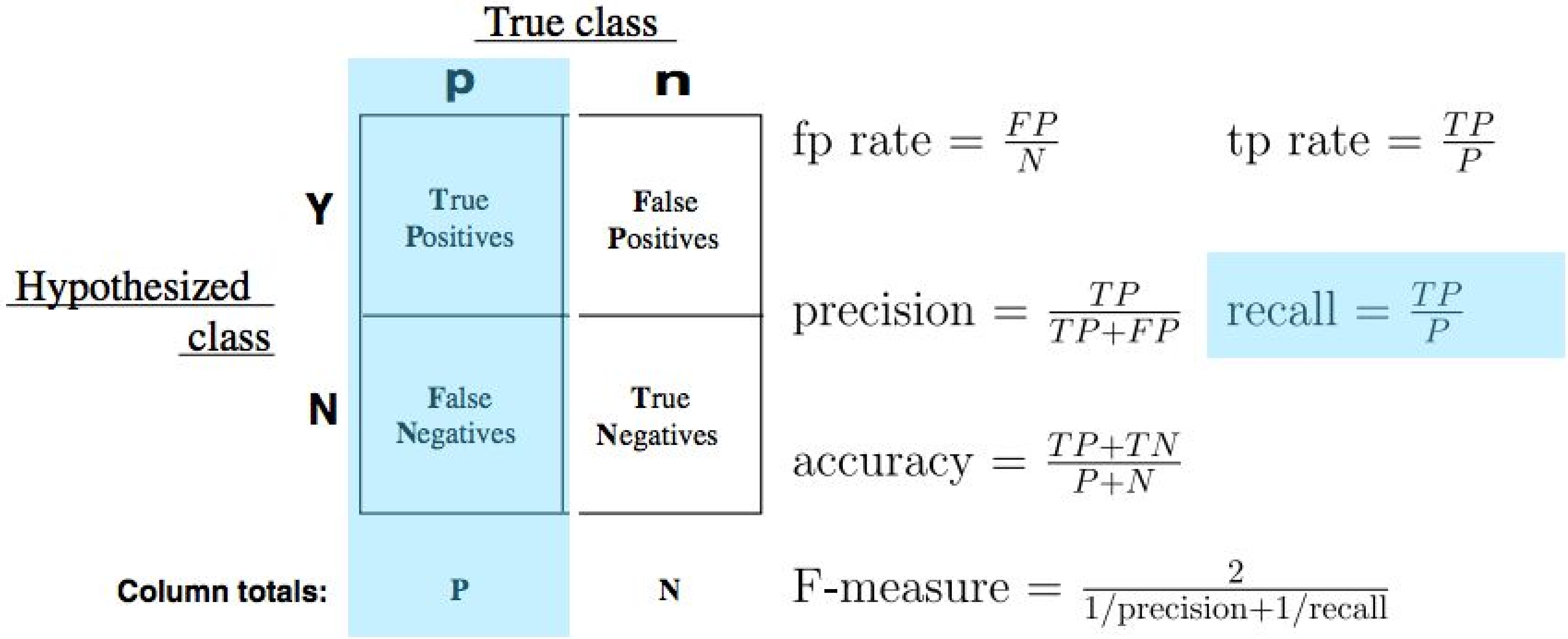
DELIVERABLE

Answers to the above question

THE MATH FOR RECALL

- Recall is the count of predicted *true positives* over the total count of that class label.

- This is the same as True Positive Rate or *sensitivity*.



THE MATH FOR RECALL

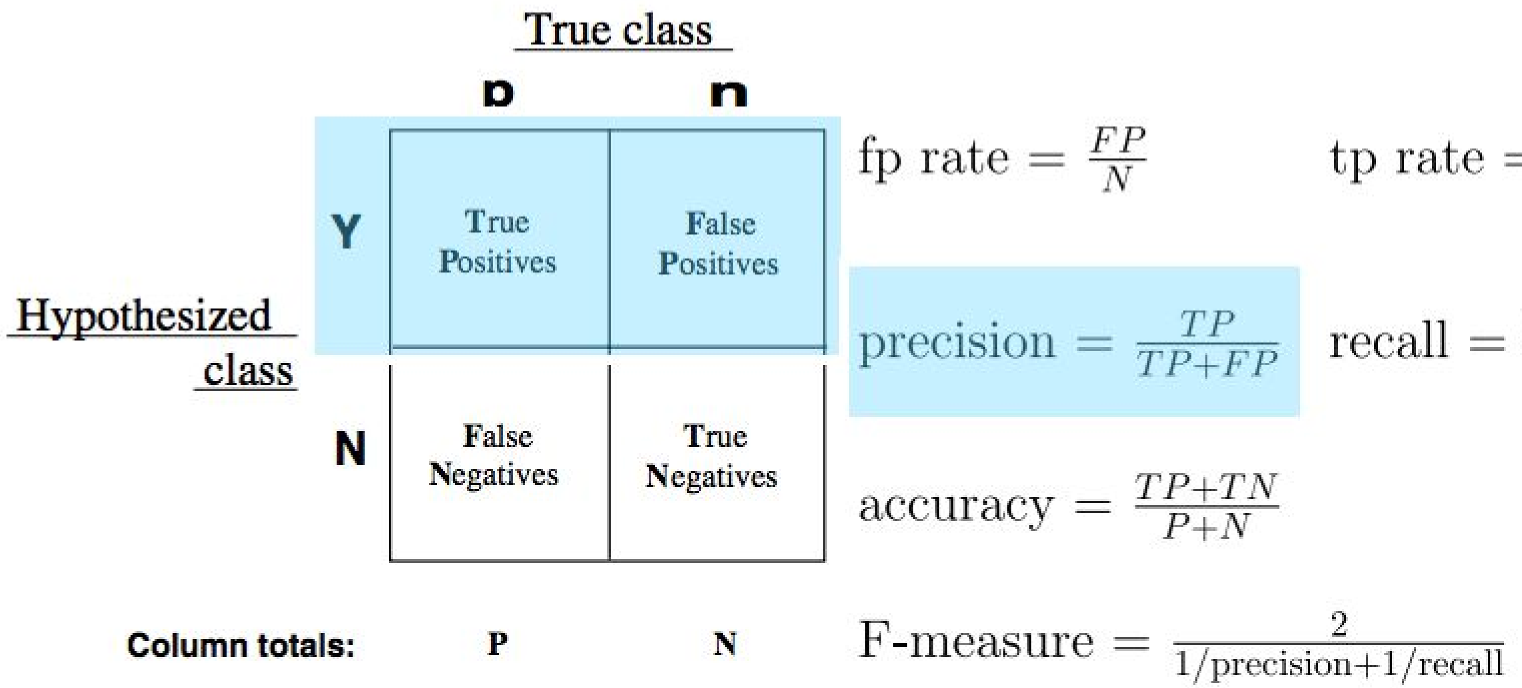
- Imagine predicting the color of a marble as either red or green. There are 10 of each.

- If the model identifies 8 of the green marbles as green, the recall is 8 / 10 = 0.80.

- However, this says nothing of the number of *red* marbles that are also identified as green.

THE MATH FOR PRECISION

- Precision, or positive predicted value, is calculated as the count of predicted true positives over the count of all values predicted to be positive.



THE DIFFERENCE BETWEEN PRECISION AND RECALL

- The key difference between the two is the attribution and value of error.
- Should our model be more picky in avoiding false positives (precision)?
- Or should it be more picky in avoiding false negatives (recall)?
- The answer should be determined by the problem you’re trying to solve.

GUIDED PRACTICE

COST BENEFIT ANALYSIS

ACTIVITY: COST BENEFIT ANALYSIS

EXERCISE

One tool that complements the confusion matrix is cost-benefit analysis, where you attach a *value* to correctly and incorrectly predicted data.

Like the Precision-Recall trade off, there is a balancing point to the *probabilities* of a given position in the confusion matrix, and the *cost* or *benefit* to that position. This approach allows you to not only add a weighting system to your confusion matrix, but also to speak the language of your business stakeholders (i.e. communicate your values in dollars!).

ACTIVITY: COST BENEFIT ANALYSIS

EXERCISE

DIRECTIONS

Consider the following marketing problem:

As a data scientist working on marketing spend, you've build a model that reduces user churn--the number of users who decide to stop paying for a product--through a marketing campaign. Your model generates a confusion matrix with the following probabilities (these probabilities are calculated as the value in that position over the sum of the sample):

```
| TP: 0.2 | FP: 0.2 |
---------------------
| FN: 0.1 | TN: 0.5 |
```

ACTIVITY: COST BENEFIT ANALYSIS

EXERCISE

In this case:

- The *benefit* of a true positive is the retention of a user ($10 for the month)
- The *cost* of a false positive is the spend of the campaign per user ($0.05)
- The *cost* of a false negative (someone who could have retained if sent the campaign) is, effectively, 0 (we didn't send it... but we certainly didn't benefit!)
- The *benefit* of a true negative is 0: No spend on users who would have never retained.

To calculate Cost-Benefit, we'll use this following function:

```
(P(TP) * B(TP)) + (P(TN) * B(TN)) + (P(FP) * C(FP)) + (C(FN) * C(FN))
```

which for our marketing problem, comes out to this:

```
(.2 * 10) + (.5 * 0) - (.2 * .05) - (.1 * 0)
```

or $1.99 per user targeted.

SHOWING WORK

SHOWING WORK

- We’ve spent a lot of time exploring our data and building a reasonable model that performs well.
- However, if we look at our visuals, they are most likely:
    - Statistically heavy: Most people don’t understand histograms.
    - Overly complicated: Scatter matrices produce too much information.
    - Poorly labeled: Code doesn’t require adding labels, so you may not have added them.

SHOWING WORK

- In order to convey important information to our audience, make sure our charts are:
  - Simplified
  - Easily interpretable
  - Clearly labeled

SIMPLIFIED

- At most, you'll want to include figures that either explain a variable on its own or explain that variable's relationship with a target.
- If your model used a data transformation (like natural log), just visualize the original data.
- Try to remove any unnecessary complexity.

EASILY INTERPRETABLE

- Any stakeholder looking at a figure should be seeing the exact same thing you're seeing.

- A good test for this is to share the visual with others less familiar with the data and see if they come to the same conclusion.

- How long did it take them?

CLEARLY LABELED

- Take the time to clearly label your axis, title your plot, and double check your scales - especially if the figures should be comparable.
- If you're showing two graphs side by side, they should follow the same Y axis.

QUESTION TO ASK

- When building visuals for another audience, ask yourself these questions:
  - **Who**: Who is my target audience for the visual?
  - **What**: What do they already know about this project? What do they need to know?
  - **How**: How does my project affect this audience? How might they interpret (or misinterpret) the data?

CONCLUSION

TOPIC REVIEW

REVIEW AND NEXT STEPS

- What do precision and recall mean? How are they similar and different to True Positive Rate and False Positive Rate?
- How does cost benefit analysis play a role in building models?
- What are at least two very important details to consider when creating visuals for a project's stakeholders?

COURSE

BEFORE NEXT CLASS

HOMEWORK

- Check the GitHub page for this class

LESSON

EXIT TICKET

DON'T FORGET TO FILL OUT YOUR EXIT TICKET