GENERAL ASSEMBLY

INTRODUCTION TO LOGISTIC REGRESSION

Gus Ostow

Machine Learning Consultant, Papert Labs

LEARNING OBJECTIVES

- Build a Logistic regression classification model using the sklearn library
- Describe a sigmoid function, odds, and the odds ratio as well as how they relate to logistic regression
- Evaluate a model using metrics such as accuracy, precision, and recall.

PRE-WORK REVIEW

- Implement a linear model (LinearRegression) with sklearn
- Understand what a coefficient is
- Recall metrics such as accuracy and misclassification
- Recall the differences between L1 and L2 regularization

OPENING

INTRODUCTION TO LOGISTIC REGRESSION

INTRODUCTION TO LOGISTIC REGRESSION

EXERCISE

ANSWER THE FOLLOWING QUESTIONS

Read through the following questions and brainstorm answers for each:

1. What are the main differences between linear and KNN models? What is different about how they approach solving the problem?
   a. For example, what is *interpretable* about linear regression compared to what's *interpretable* in KNN?

1. What would be the advantage of using a linear model like linear regression to solve a classification problem, compared to KNN?
   a. What are some challenges for using linear regression to solve a classification problem (say, if the values were either 1 or 0)?

INTRODUCTION

LOGISTIC REGRESSION

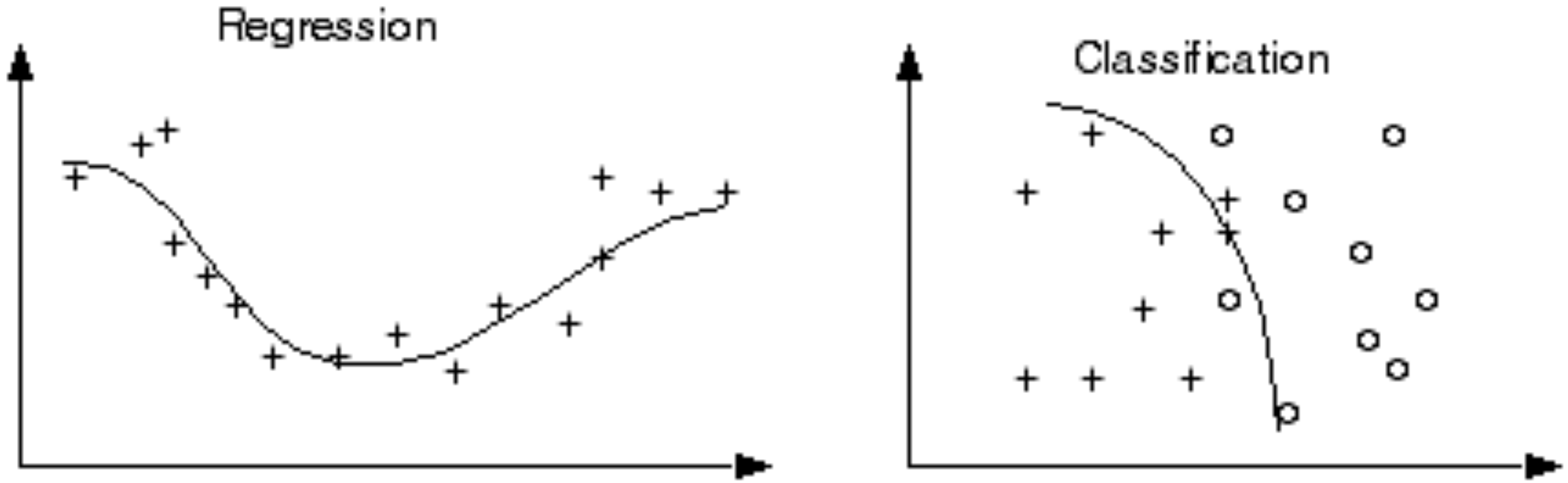
LOGISTIC REGRESSION

- Logistic regression is a *linear* approach to solving a *classification* problem.
- That is, we can use a linear model, similar to Linear regression, in order to solve if an item *belongs* or *does not belong* to a class label.

CHALLENGE! LINEAR REGRESSION RESULTS FOR CLASSIFICATION

- Regression results can have a value range from -∞ to ∞.

- Classification is used when predicted values (i.e. class labels) are not greater than or less than each other.



CHALLENGE! LINEAR REGRESSION RESULTS FOR CLASSIFICATION

- But, since most classification problems are binary (0 or 1) and 1 is greater than 0, does it make sense to apply the concept of regression to solve classification?

- How might we contain those bounds?

- Let’s review some approaches to make classification with regression feasible.

FIX 1: PROBABILITY

- One approach is predicting the probability that an observation belongs to a certain class.

FIX 1: PROBABILITY

- For example, suppose we know that roughly 700 of 2200 people from the Titanic survived. Without knowing anything about the passengers or crew, the probability of survival would be ~0.32 (32%).

- However, we still need a way to use a linear function to either increase or decrease the probability of an observation given the data about it.

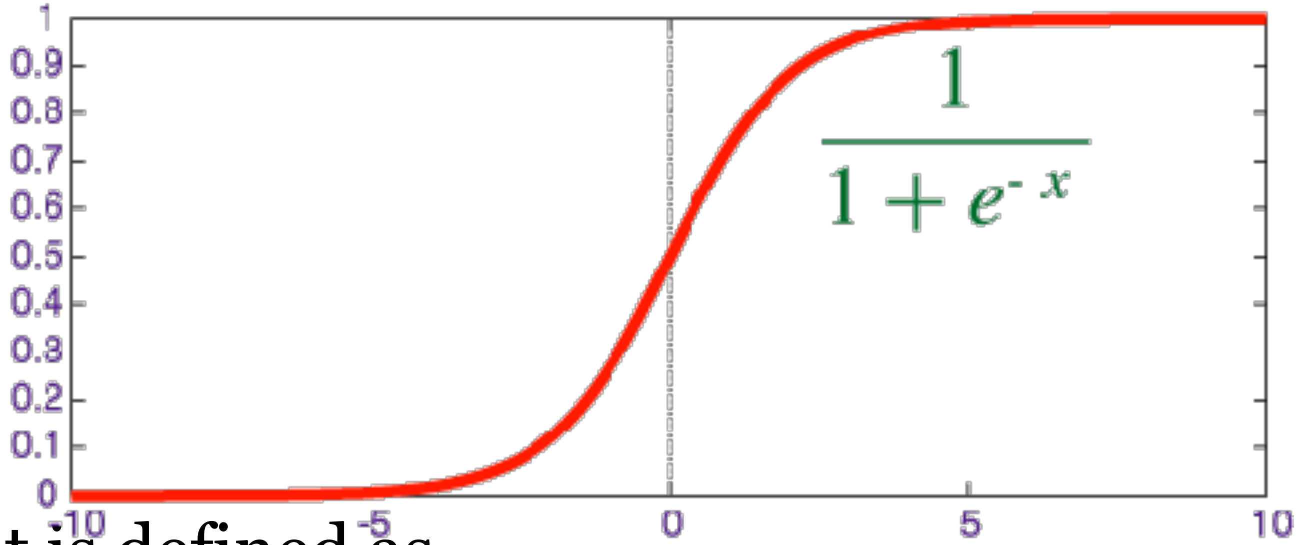
FIX 2: LINK FUNCTIONS AND THE SIGMOID FUNCTION

- We can use a linear model *like* linear regression to output regression *like* values, then convert them into probabilities for classification

- We use a link function to turn linear regression into something that can be used for classification

FIX 2: LINK FUNCTIONS AND THE SIGMOID FUNCTION

- A *sigmoid function* is a function that visually looks like an s.



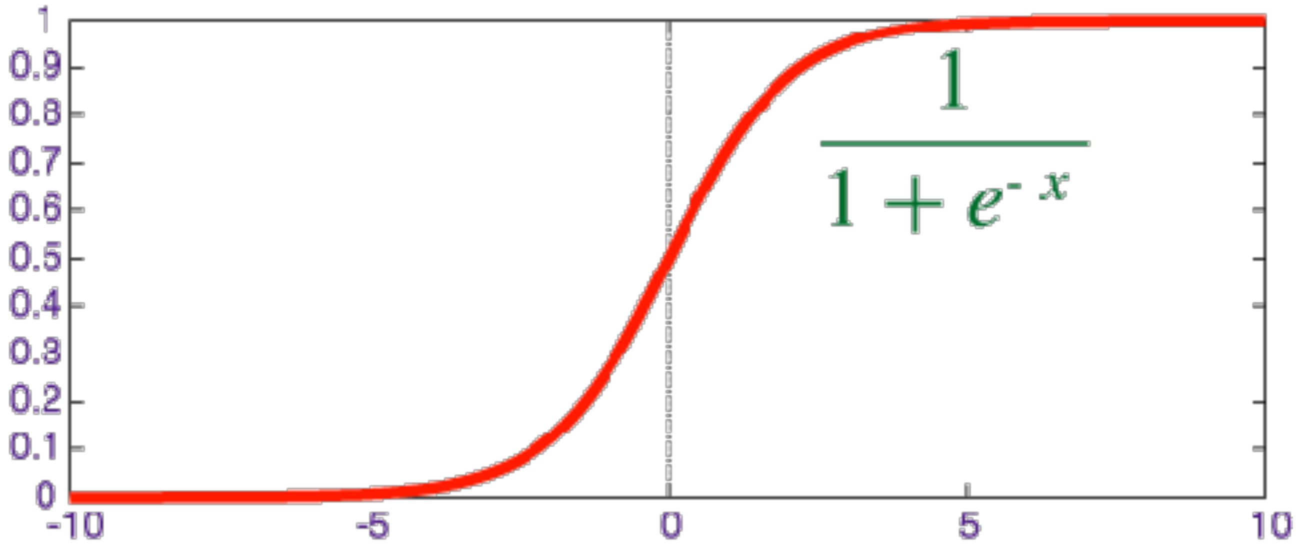
- Mathematically, it is defined as

$$f(x) = \frac{1}{1+\mathrm{e}^{-x}}$$

FIX 2: LINK FUNCTIONS AND THE SIGMOID FUNCTION

- As x increases, the results is closer to 1. As x decreases, the result is closer to 0.

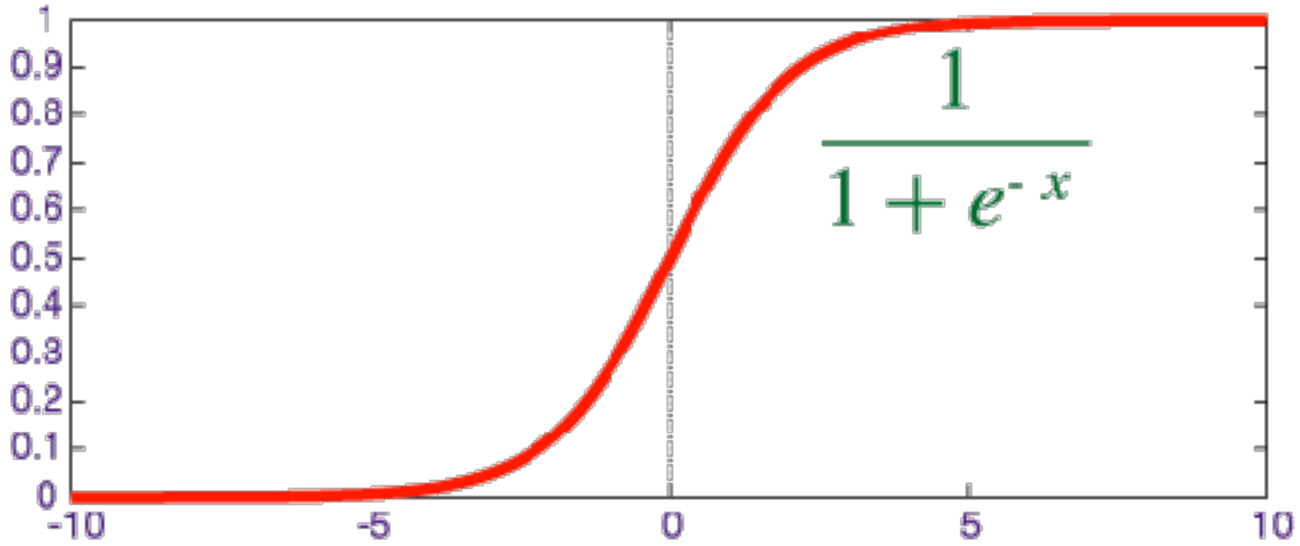
- When x = 0, the result is 0.5.



FIX 2: LINK FUNCTIONS AND THE SIGMOID FUNCTION

- Since x decides how to much to increase or decrease the value away from 0.5, x can be interpreted as something like a coefficient.

- X, here is the value a linear regression would output. Something between negative infinity and infinity



DEMO

PLOTTING A SIGMOID FUNCTION

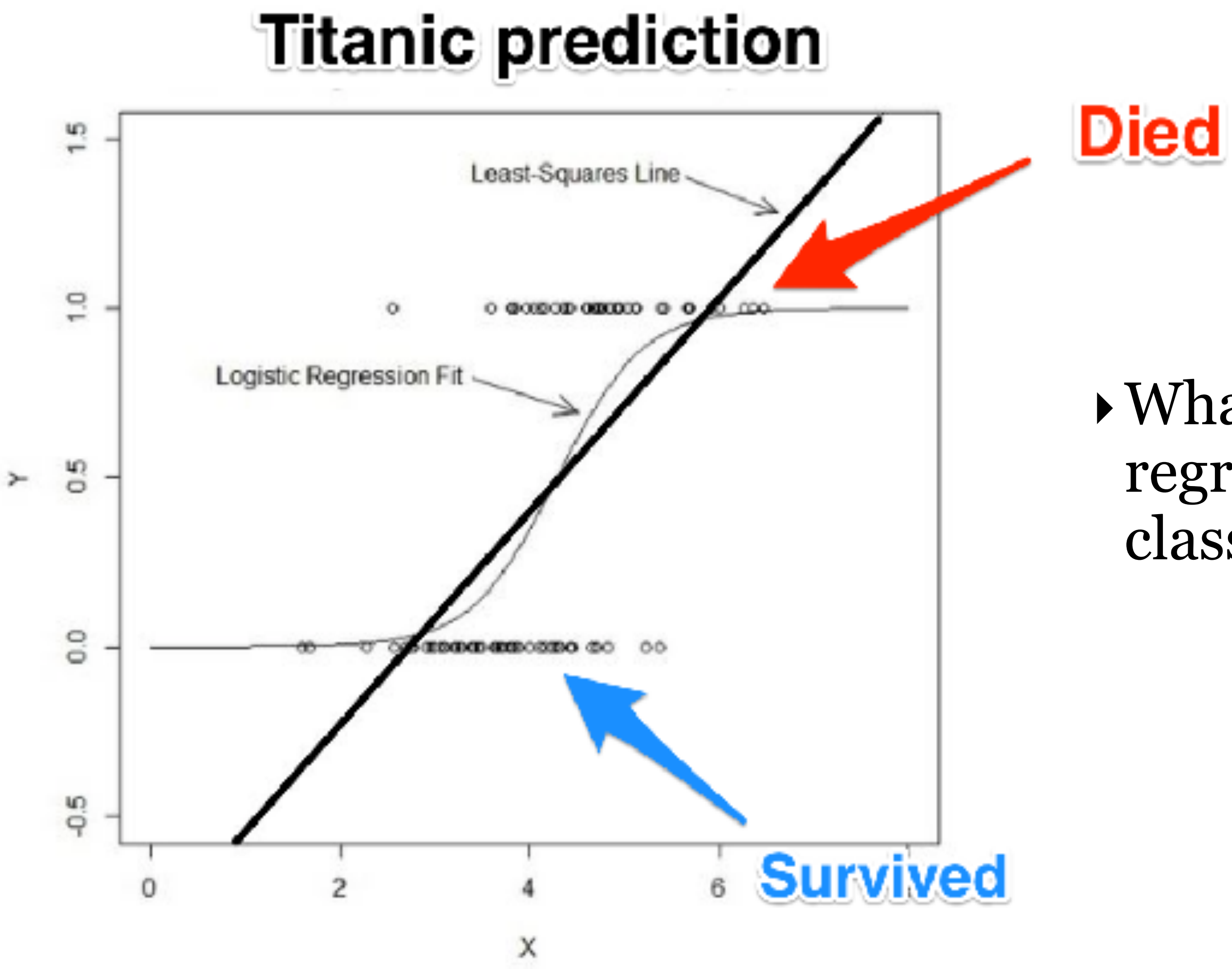
Learning check

- *Why is sigmoid well suited to turning numbers into a probabilities, especially for binary classification?*

FIX 3: ODDS AND LOG-ODDS

- The logit function allows for values between -∞ and ∞, but provides us probabilities between 0 and 1.

FIX 3: ODDS AND LOG-ODDS



- What does that straight, best-fit linear regression-like line represent in terms of classification?

FIX 3: ODDS AND LOG-ODDS

- While the logit value represents the *coefficients* in the logistic function, we can convert them into odds ratios that make them more easily interpretable.

$$\boxed{Ln\left(\frac{P}{1-P}\right)} = \beta_0 + \beta_1 X_1$$

Log odds

FIX 3: ODDS AND LOG-ODDS

- While the logit value represents the *coefficients* in the logistic function, we can convert them into odds ratios that make them more easily interpretable.

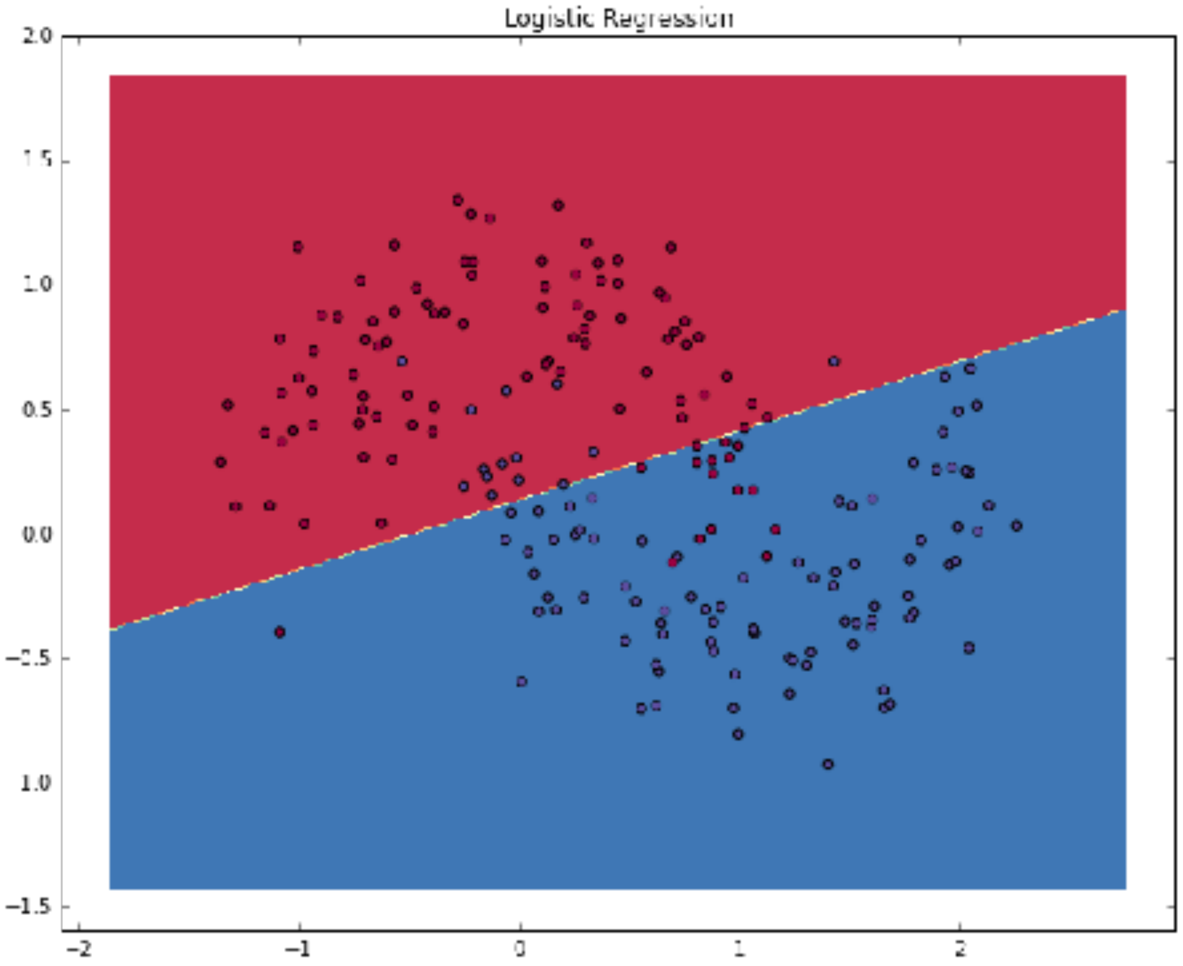
Odds

$$Ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1$$

- So how do we interpret the B1 coefficient? What can we use from what we know about linear regression?

FIX 3: ODDS AND LOG-ODDS

- With these coefficients, we get our overall probability: the logistic regression draws a linear *decision line* which divides the classes.



GUIDED PRACTICE

WAGER THOSE ODDS!

ACTIVITY: WAGER THOSE ODDS!

EXERCISE

DIRECTIONS (15 minutes)

1. Convert these odds ratios into probability that the favored team wins. Remember the formula for odds, you are going to have to invert it.

    a. Stanford : Iowa, 5:1
    b. Alabama : Michigan State, 20:1
    c. Clemson : Oklahoma, 1.1:1
    d. Houston : Florida State, 1.8:1
    e. Ohio State : Notre Dame, 1.6:1

DELIVERABLE

The desired probabilities

INDEPENDENT PRACTICE

Logistic Regression in sklearn

INTRODUCTION

ADVANCED CLASSIFICATION METRICS

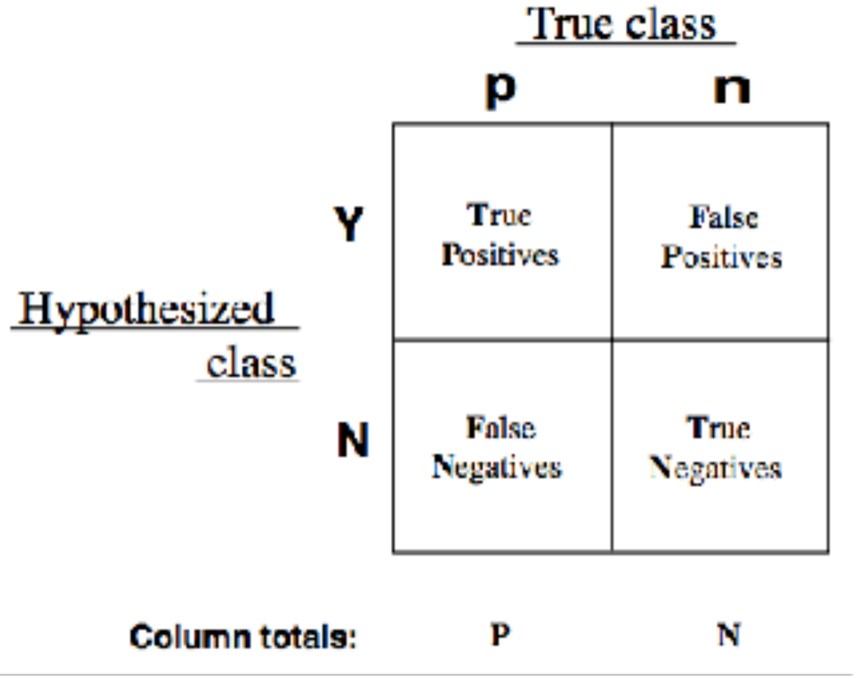
ADVANCED CLASSIFICATION METRICS

- Was it wrong across all labels?
- Did it just guess one class label for all predictions?
- It’s important to look at other metrics to fully understand the problem.

ADVANCED CLASSIFICATION METRICS

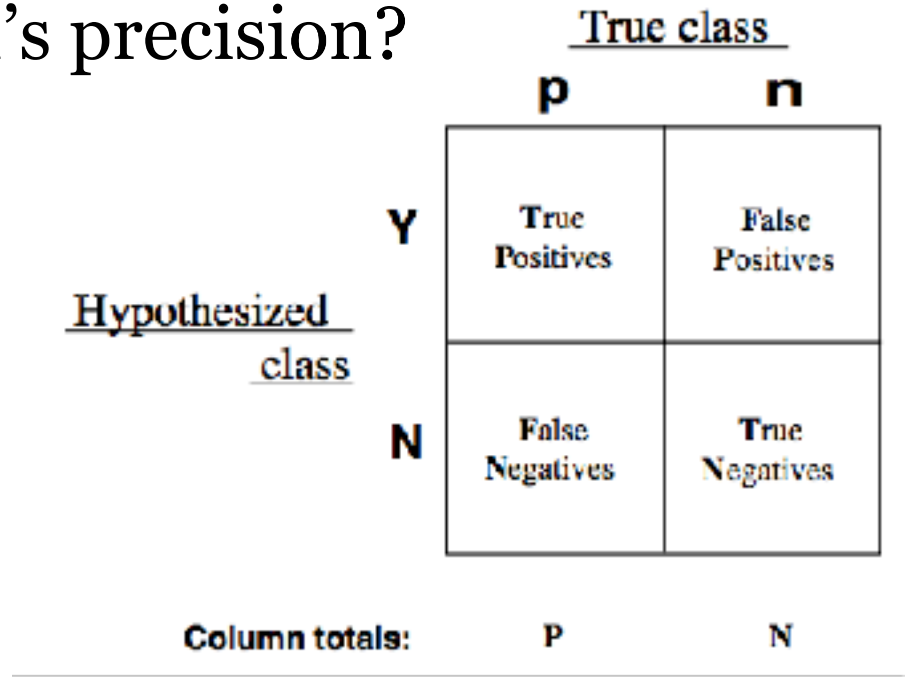
- We can split up the accuracy of each label by using the *true positive rate* and the *false positive rate*.

- For each label, we can put it into the category of a true positive, false positive, true negative, or false negative.



ADVANCED CLASSIFICATION METRICS

- **Precision:** “Of everything we classified as positive, what percentage were actually positive”
- A very precise model doesn’t have many false positives.
- What situations would you care about a model’s precision?

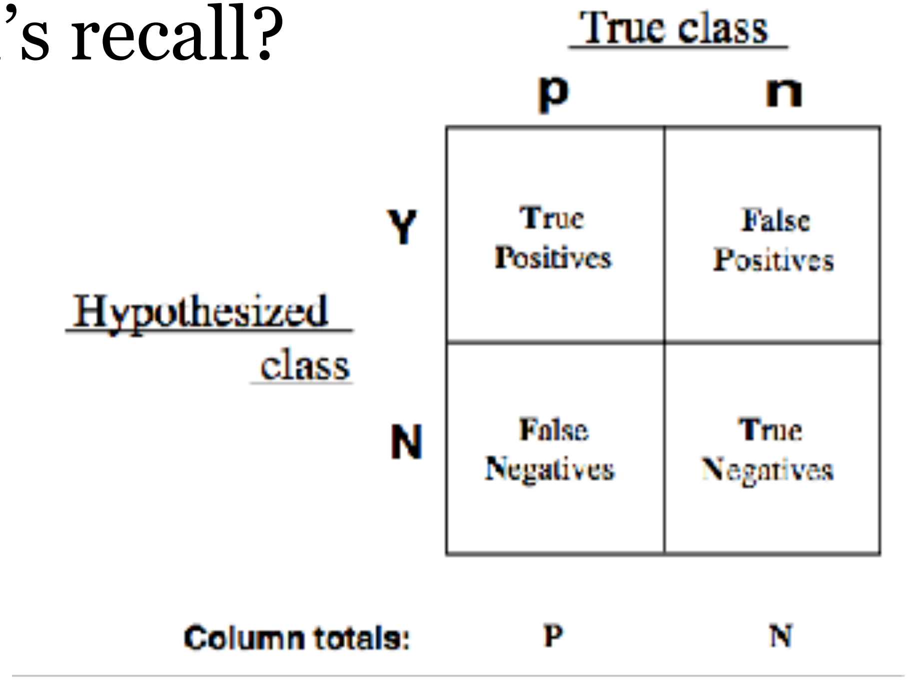


ADVANCED CLASSIFICATION METRICS

- **Recall:** “Of everything that is ***actually positive***, how many did you successfully classify as positive

- A model with high recall doesn’t make false negatives

- What situations would you care about a model’s recall?



ADVANCED CLASSIFICATION METRICS

- A good classifier would have a true positive rate approaching 1 and a false positive rate approaching 0.

- In our smoking problem, this model would accurately predict *all* of the smokers as smokers and not accidentally predict any of the nonsmokers as smokers.

GUIDED PRACTICE

WHICH METRIC SHOULD I USE?

ACTIVITY: WHICH METRIC SHOULD I USE?

EXERCISE

DIRECTIONS (15 minutes)

Examples:

1. A test is developed for determining if a patient has cancer or not.
2. A newspaper company is targeting a marketing campaign for "at risk" users that may stop paying for the product soon.
3. You build a spam classifier for your email system.

DELIVERABLE

Answers for each example

CONCLUSION

TOPIC REVIEW

REVIEW QUESTIONS

- How is logistic regression different than linear regression?
- What kind of machine learning problems does logistic regression address?
- What do the *coefficients* in a logistic regression represent? How does the interpretation differ from ordinary least squares? How is it similar?
- Why are classification metrics more nuances that regression metrics?