CS 260: Foundations of Data Science

Prof. Thao Nguyen

Fall 2024

HAVERFORD
COLLEGE

Admin

- **Lab 2** grades & feedback posted on Moodle

Outline for today

- Evaluation Metrics
  - Confusion matrices
  - Precision and recall
  - ROC curves

- Introduction to probability

Outline for today

- Evaluation Metrics
  - Confusion matrices
  - Precision and recall
  - ROC curves
- Introduction to probability

Goals of Evaluation

- Think about what metrics are important for the problem at hand
- Compare different methods or models on the same problem
- Common set of tools that other researchers/users can understand

Training and Testing

(high-level idea)

- **Separate** data into "**train**" and "**test**"
  - n = num training examples
  - m = num testing examples

- **Fit** (create) the model using **training data**
  - e.g. sea_ice_1979-2012.csv

- **Evaluate** the model using **testing data**
  - e.g. sea_ice_2013-2020.csv

Confusion Matrices

Predicted class

True class	Negative	Positive	
Negative	True negative (TN)	False positive (FP) “false alarm”	N
Positive	False negative (FN) “miss”	True positive (TP)	P
	N*	P*	

Error:

$(FN+FP)/(TN+FP+FN+TP)$

$= (FN+FP)/(N+P)$

Modified from Jessica Wu

Confusion Matrices

True class \ Predicted class	Negative	Positive	
Negative	True negative (TN)	False positive (FP) “false alarm”	N
Positive	False negative (FN) “miss”	True positive (TP)	P
	N*	P*	

Accuracy = 1-Error:

(TN+TP)/(TN+FP+FN+TP)

= (TN+TP)/(N+P)

Modified from Jessica Wu

Confusion Matrices

		Predicted class		
		Negative	Positive	
True class	Negative	True negative (TN)	False positive (FP) “false alarm”	N
	Positive	False negative (FN) “miss”	True positive (TP)	P
		N*	P*	

Precision:

TP/(FP+TP) = TP/P*

Confusion Matrices

Predicted class

True class	Negative	Positive	
Negative	True negative (TN)	False positive (FP) “false alarm”	N
Positive	False negative (FN) “miss”	True positive (TP)	P
	N*	P*	

Recall
(True Positive Rate):

TP/(FN+TP) = TP/P

Confusion Matrices

		Predicted class	
		Negative	Positive	
True class	Negative	True negative (TN)	False positive (FP) “false alarm”	N
	Positive	False negative (FN) “miss”	True positive (TP)	P
		N*	P*	

False Positive Rate:

$FP/(TN+FP) = FP/N$

Modified from Jessica Wu

Precision and Recall

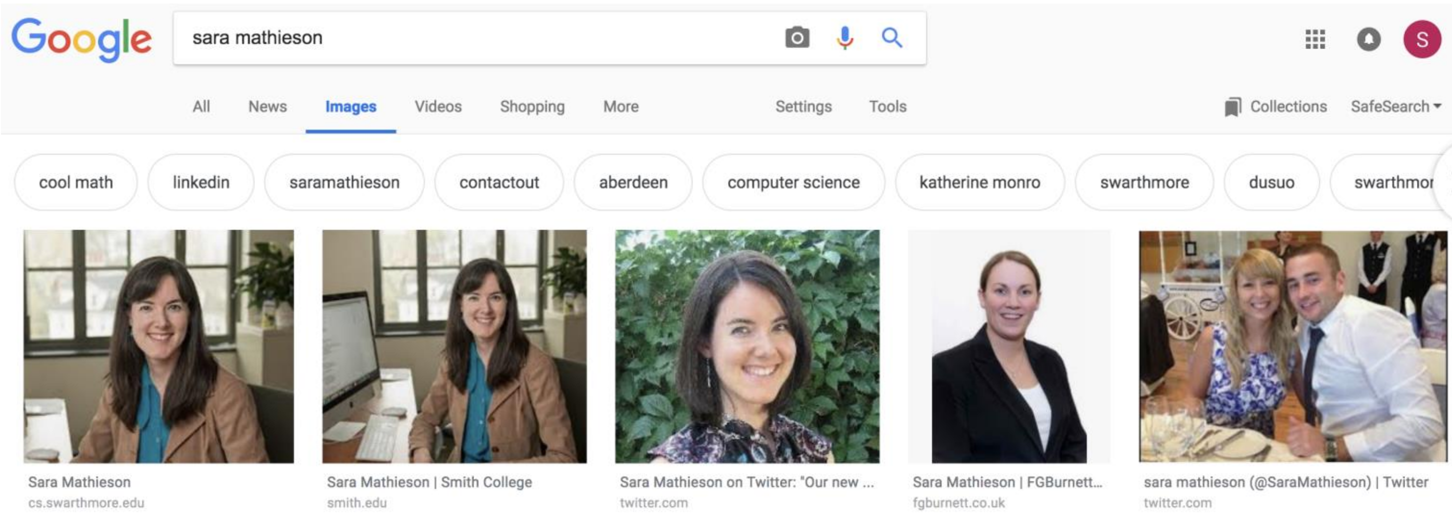
- Precision: of all the “flagged” examples, which ones are actually relevant (i.e. positive)?

(Purity)

- Recall: of all the relevant results, which ones did I actually return?

(Completeness)

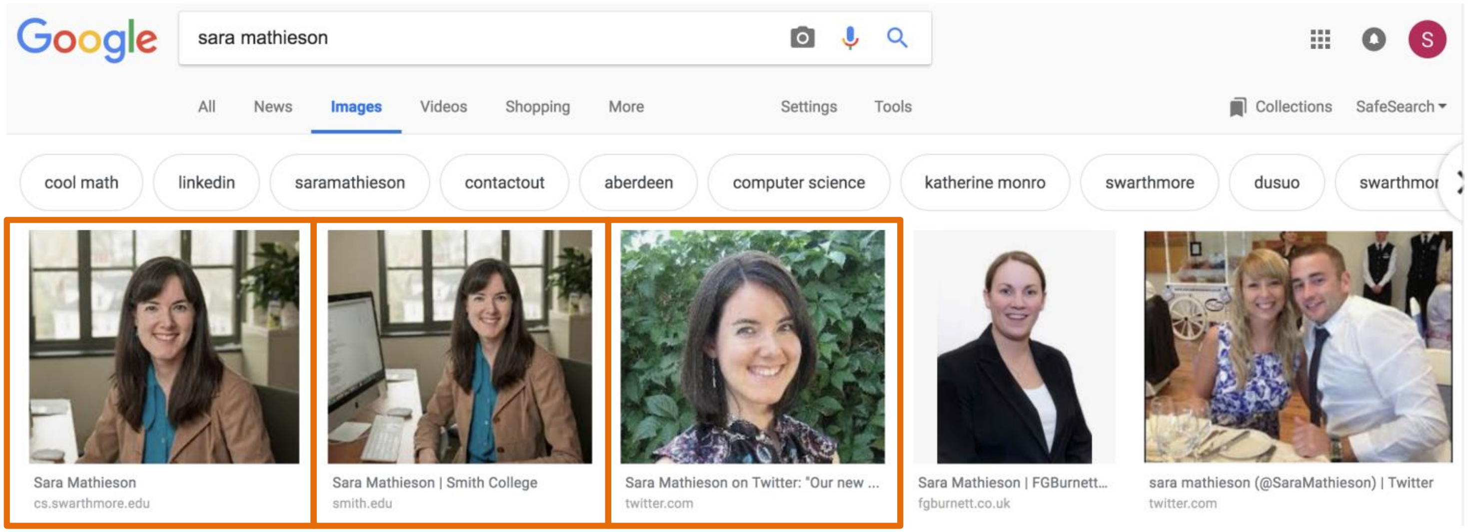
Precision and Recall



P=6 (number of images that are actually Sara)

- Precision?
- Recall?

Precision and Recall

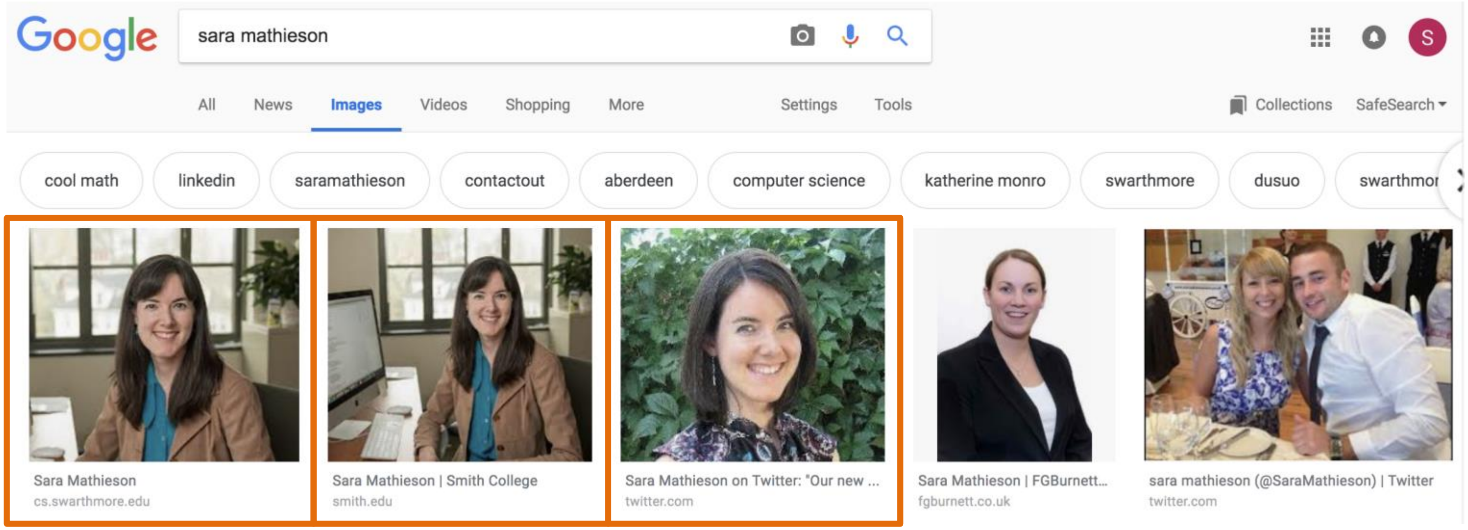


P=6 (number of images that are actually Sara)

- Precision = TP/(FP+TP) = 3/5
- Recall?

Modified from Ameet Soni

Precision and Recall

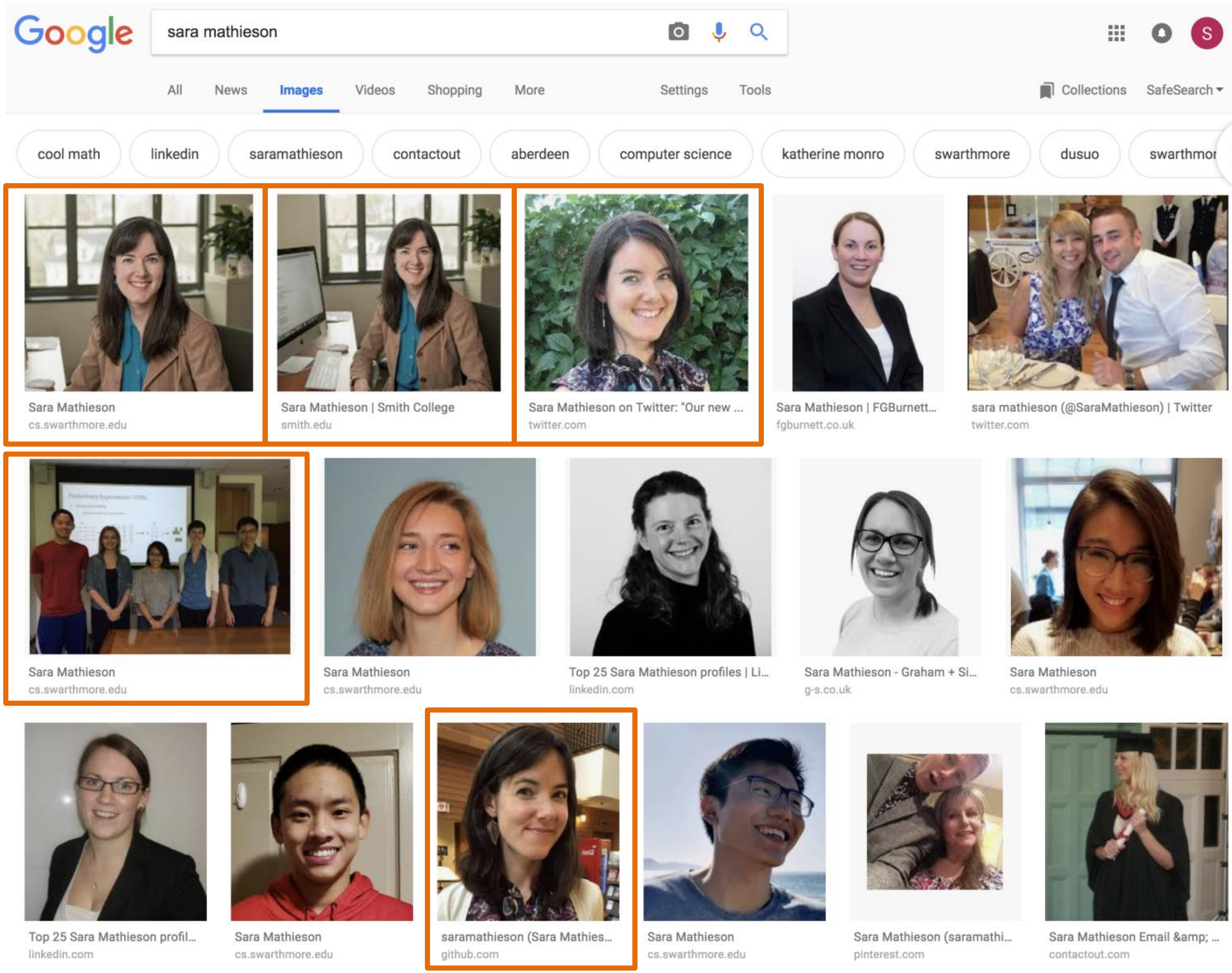


P=6 (number of images that are actually Sara)

- Precision = TP/(FP+TP) = 3/5
- Recall = TP/(FN+TP) = 3/6

Modified from Ameet Soni

Precision and Recall



P=6 (number of images that are actually Sara)

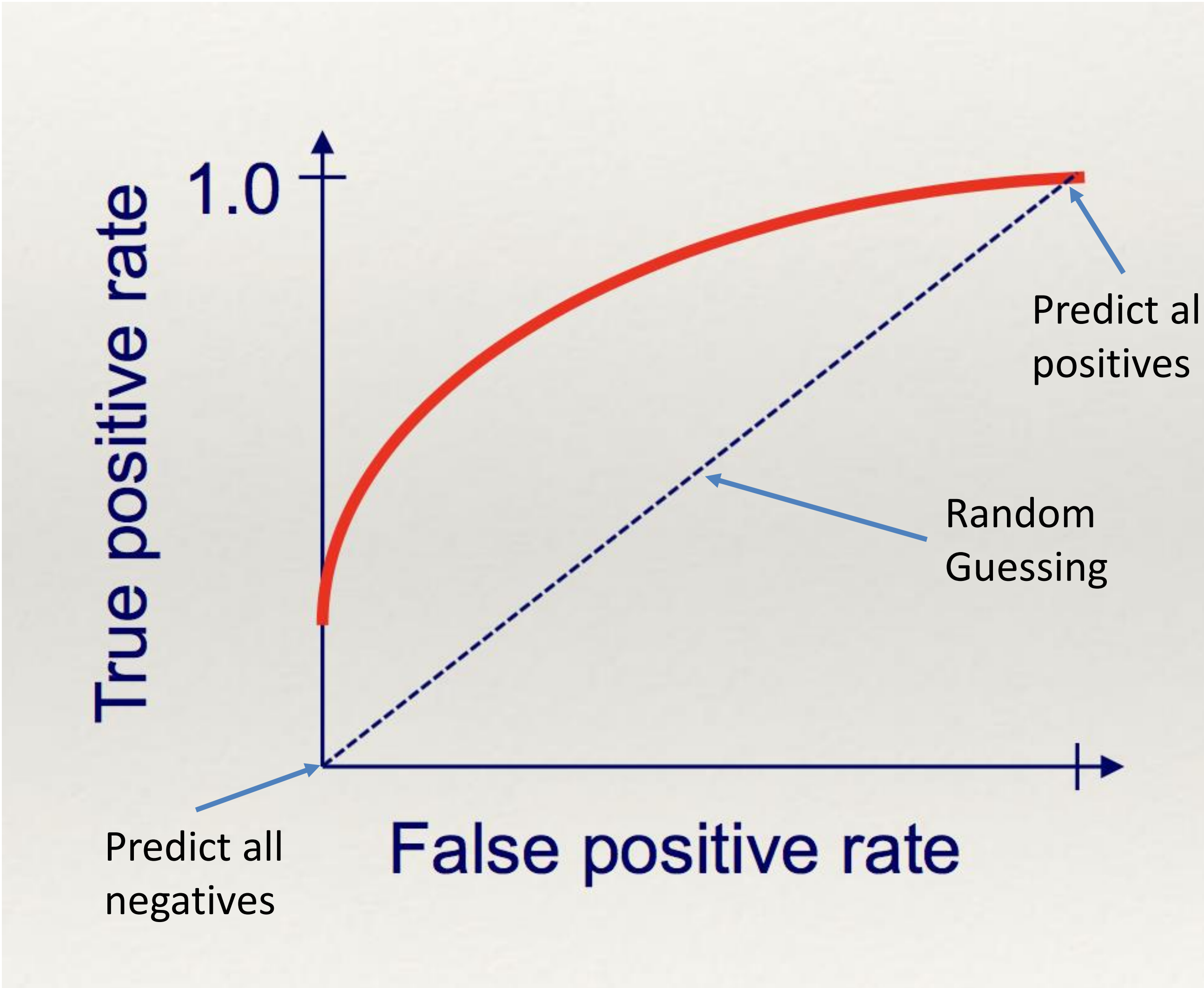
- Precision = 5/16
- Recall = 5/6

Modified from Ameet Soni

ROC curve (Receiver Operating Characteristic)

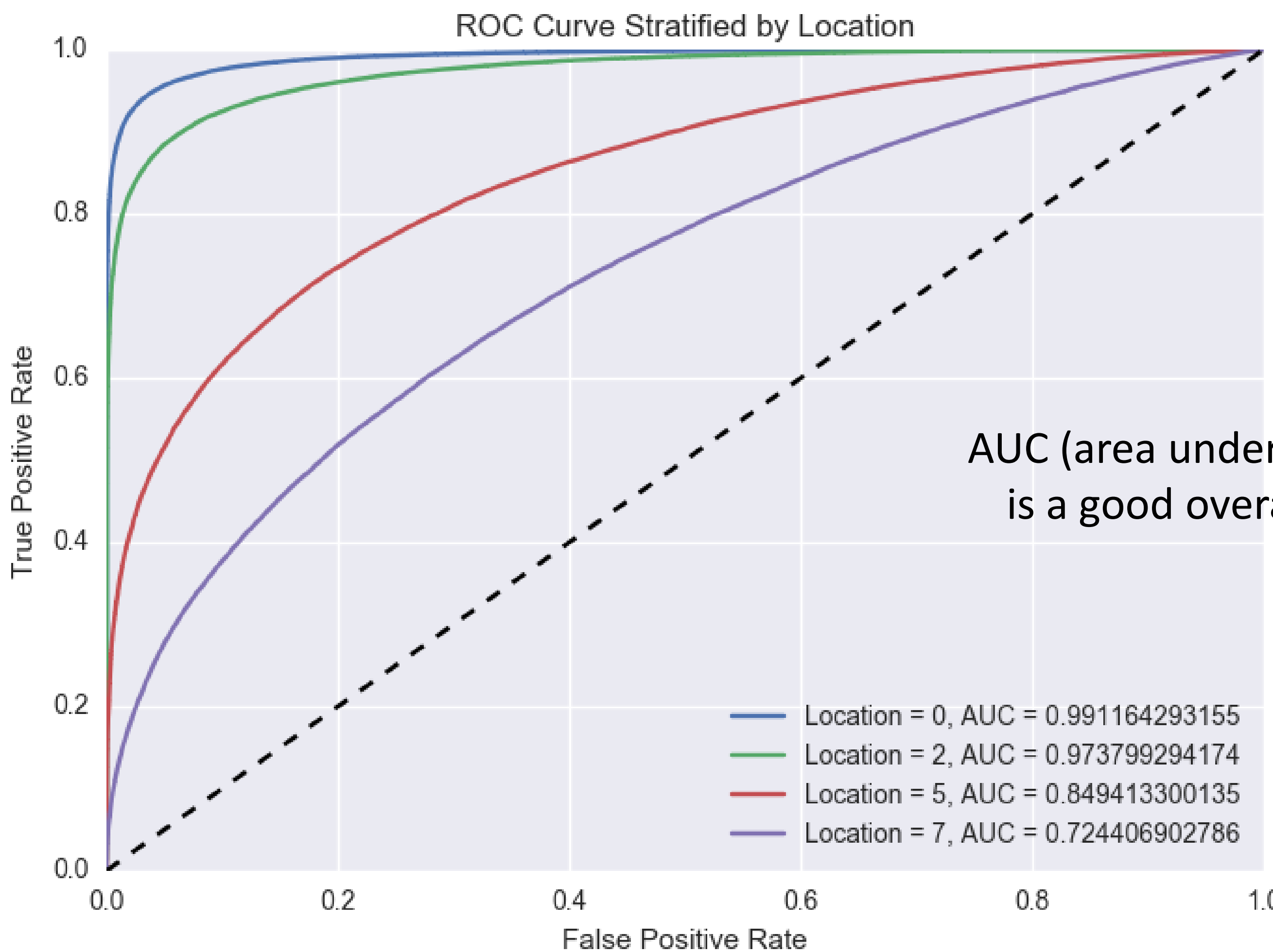
More history here!

https://en.wikipedia.org/wiki/Receiver_operating_characteristic



Modified from Ameet Soni

ROC curve example: comparing methods



AUC (area under the curve)
is a good overall metric

Example of a ROC curve
Chan, Perrone, Spence, Jenkins, Mathieson, Song

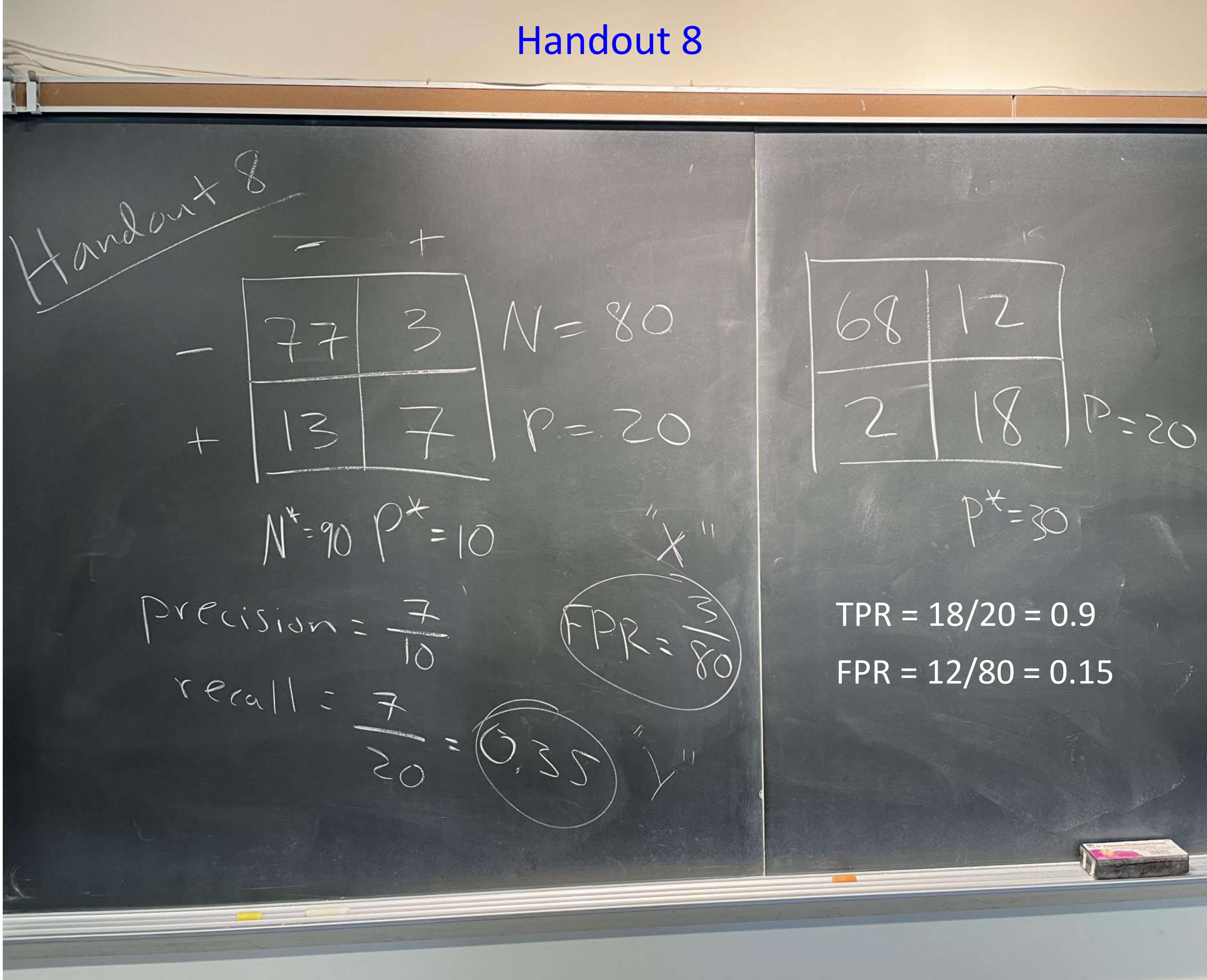
How to get a ROC curve for probabilistic methods?

- Usually we use 0.5 as a threshold for binary classification

- Vary the threshold! (i.e. choose 0, 0.1, 0.2,...)

  – $P(y=1 \mid x) >= 0.2$ => classify as 1 (positive)
  – $P(y=1 \mid x) < 0.2$ => classify as 0 (negative)

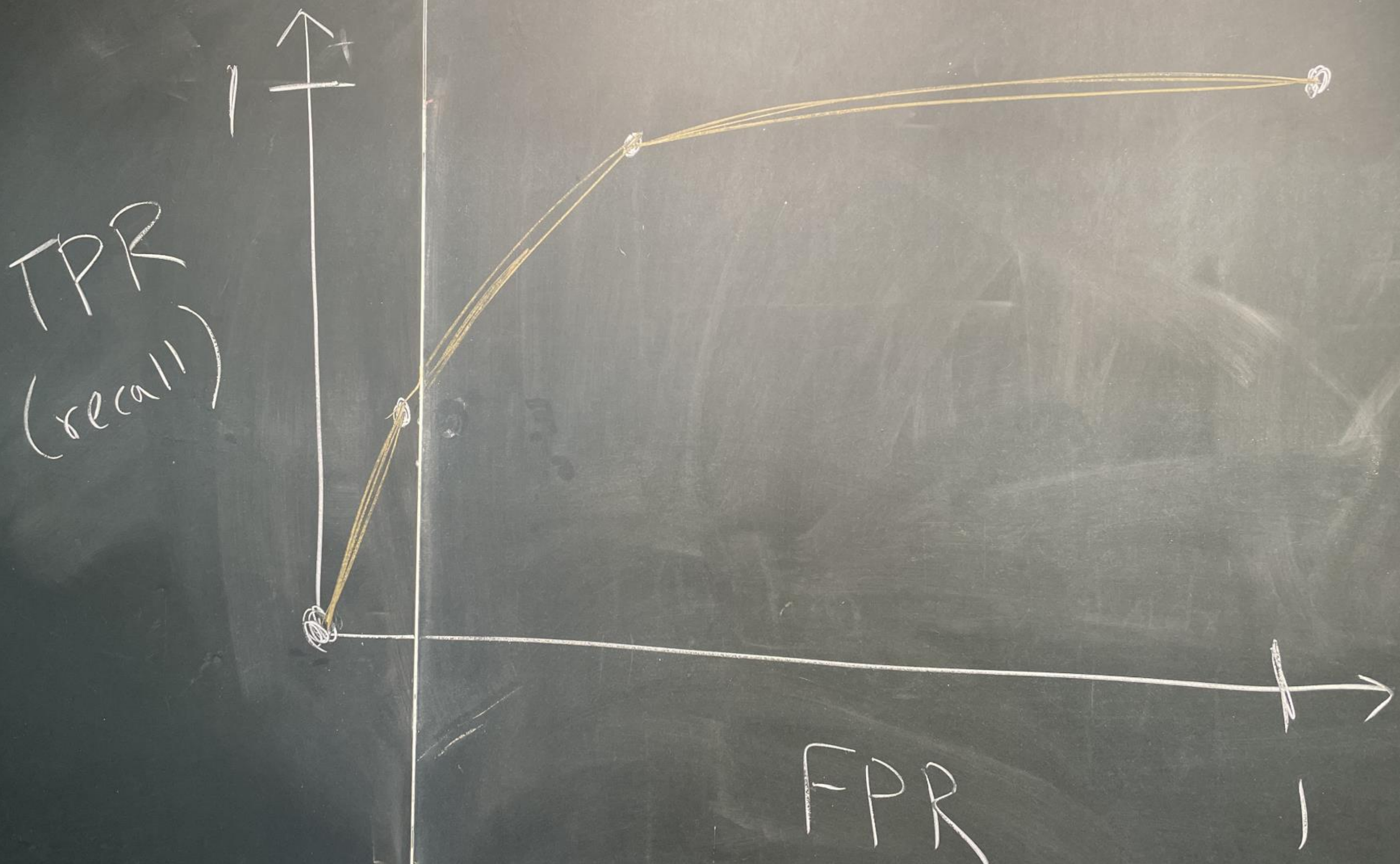
Handout 8



TPR = 18/20 = 0.9

FPR = 12/80 = 0.15

Handout 8



Outline for today

- Evaluation Metrics
  - Confusion matrices
  - Precision and recall
  - ROC curves

- Introduction to probability

Intro to Probability

- The **probability** of an **event** e has a number of epistemological interpretations
- Assuming we have **data**, we can count the number of times e occurs in the dataset to estimate the probability of e, $P(e)$.

$$P(e) = \frac{\mathrm{count}(e)}{\mathrm{count}(\text{all events})}.$$

- If we put all events in a bag, shake it up, and choose one at random (called **sampling**), how likely are we to get e?

Intro to Probability

- Suppose we flip a fair coin
- What is the probability of heads, $P(e = H)$?

Intro to Probability

- Suppose we flip a fair coin
- What is the probability of heads, $P(e = H)$?
- We have "all" of two possibilities, $e \in \{H, T\}$.
- $P(e = H) = \frac{count(H)}{count(H)+count(T)}$

Materials by Alvin Grissom II

Intro to Probability

- Suppose we have a fair 6-sided die.
- What's the probability of getting "1"?

Intro to Probability

- Suppose we have a fair 6-sided die.
- What's the probability of getting "1"?

$$\frac{count(s)}{count(1) + count(2) + count(3) + \cdots + count(6)} = \frac{1}{1+1+1+1+1+1} = \frac{1}{6}$$

Intro to Probability

- What about a die with on ly three numbers $\{1, 2, 3\}$, each of which appears twice?
- What's the probability of getting "1"?

Intro to Probability

- What about a die with on ly three numbers $\{1, 2, 3\}$, each of which appears twice?
- What's the probability of getting "1"?

$$P(e=1) = \frac{count(1)}{count(1) + count(2) + count(3)} = \frac{2}{2+2+2} = \frac{1}{3}.$$

Intro to Probability

- The set of all probabilities for an event e is called a **probability distribution**
- Each coin toss is an independent event (Bernoulli trial).

Intro to Probability

- Which is greater, $P(HHHHH)$ or $P(HHTHH)$?

Intro to Probability

- Which is greater, $P(HHHHH)$ or $P(HHTHH)$?
- Since the events are independent, they're equal

Intro to Probability

Probability Axioms

1. Probabilities of events must be no less than 0. $P(e) \geq 0$ for all e.
2. The sum of all probabilities in a distribution must sum to 1. That is, $P(e_1) + P(e_2) + \ldots + P(e_n) = 1$. Or, more succinctly,

$$\sum_{e \in E} P(e) = 1.$$

Intro to Probability

Joint Probability

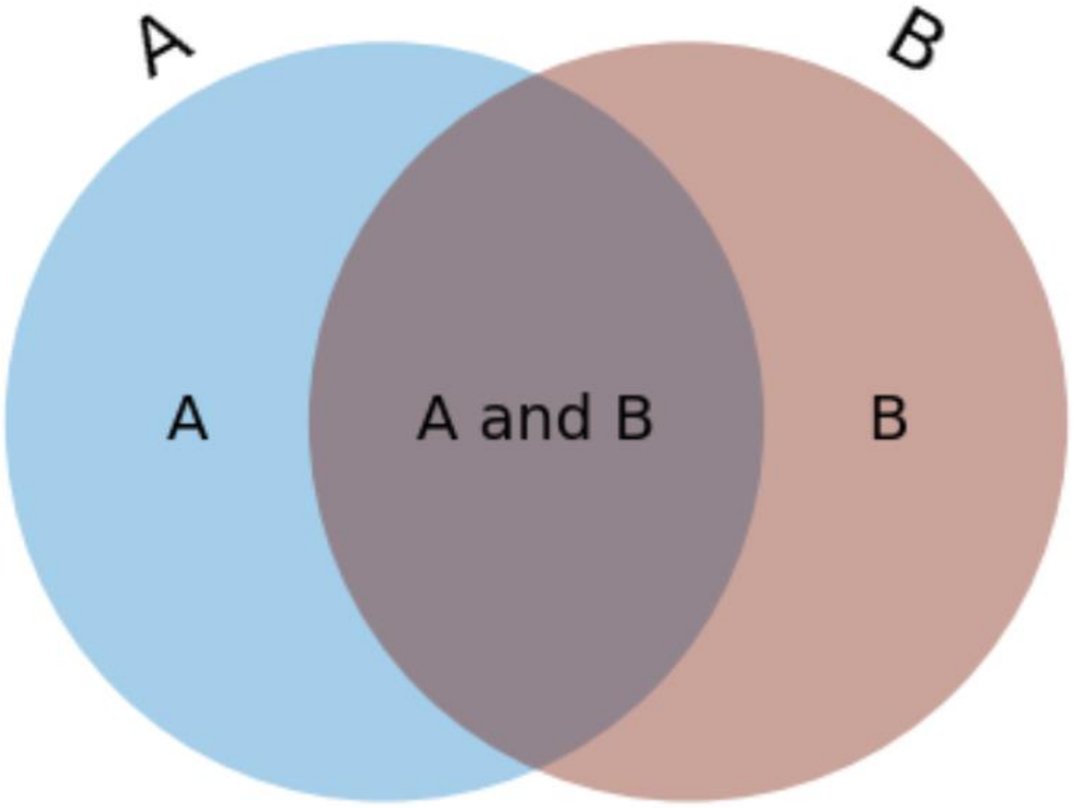
The probability that two independent events e_1 and e_2 *both* occur is given by their product.

$$P(e_1 \wedge e_2) = P(e_1 \cap e_2) = P(e_1)P(e_2) \text{ when } e_1 \cap e_2 = \emptyset$$

- Intuitively, think of every probability as a *scaling factor*.
- You can think of a probability as the fraction of the probability space occupied by an event e_1.
  - $P(e_1 \wedge e_2)$ is the fraction of of e_1's probability space wherein e_2 also occurs.
  - So, if $P(e_1) = \frac{1}{2}$ and $P(e_2) = \frac{1}{3}$, then $P(e_2, e_1)$ is a third of a half of the probability space or $\frac{1}{3} \times \frac{1}{2}$.

Intro to Probability

Joint Probability



Materials by Alvin Grissom II

Intro to Probability

Conditional Probability

- A **conditional probability** is the probability that one event occurs given that we take another for granted.
- The probability of e_2 given e_1 is $P(e_2 \mid e_1)$.
- This is the probability that e_2 will occur given that we take for granted that e_1 occurs.