

CS 260: Foundations of Data Science

Prof. Sara Mathieson

Fall 2023



HVERFORD
COLLEGE

Admin

- **Lab 4** due Tuesday Oct 3
- **Lab 2** grades posted on Moodle
 - If there was an issue with your figures being blank please send me an email!

Outline for September 28

- Finish Handout 7
- Evaluation Metrics
 - Confusion matrices
 - Precision and recall
 - ROC curves
- Begin: Bayesian probability
 - Clinical trials example

Outline for September 28

- Finish Handout 7
- Evaluation Metrics
 - Confusion matrices
 - Precision and recall
 - ROC curves
- Begin: Bayesian probability
 - Clinical trials example

Handout 7

Slope

up
flat
down

bal
pos

$$\frac{0}{4}$$

$$\frac{5}{7}$$

$$\frac{3}{5}$$

$$\frac{1}{3}$$

$$\frac{3}{8}$$

$$\frac{4}{5}$$



thal

fixed
normal
reverse

$$P_{pos} \geq 0.5 \Rightarrow pos$$

$$P_{pos} \geq \frac{1}{2} \Rightarrow pos$$

Outline for September 28

- Finish Handout 7
- **Evaluation Metrics**
 - Confusion matrices
 - Precision and recall
 - ROC curves
- Begin: Bayesian probability
 - Clinical trials example

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

$$\frac{65+13}{100} = 78\% \text{ Pred}$$

	-	+
truth -	65	15
truth +	7	13

Accuracy =

Thresh = 0.5

test data

m = 100

50	30
1	19

Thresh 0.25

acc = 69%

80 negatives
20 positives

76	4
11	9

Thresh 0.75

correct

$$\frac{1}{m} \sum_{i=1}^m \mathbb{1}(\hat{y}_i = y_i)$$

Note: all the same model, different thresholds!



Confusion Matrices

Predicted class

Negative

Positive

Negative

	Negative	Positive
Negative	True negative (TN)	False positive (FP)
Positive	False negative (FN)	True positive (TP)

True class

Positive

Confusion Matrices

Predicted class

Negative

Positive

Negative

True negative (TN)	False positive (FP) “false alarm”
False negative (FN) “miss”	True positive (TP)

N (total number of true negatives)

True
class

Positive

P (total number of true positives)

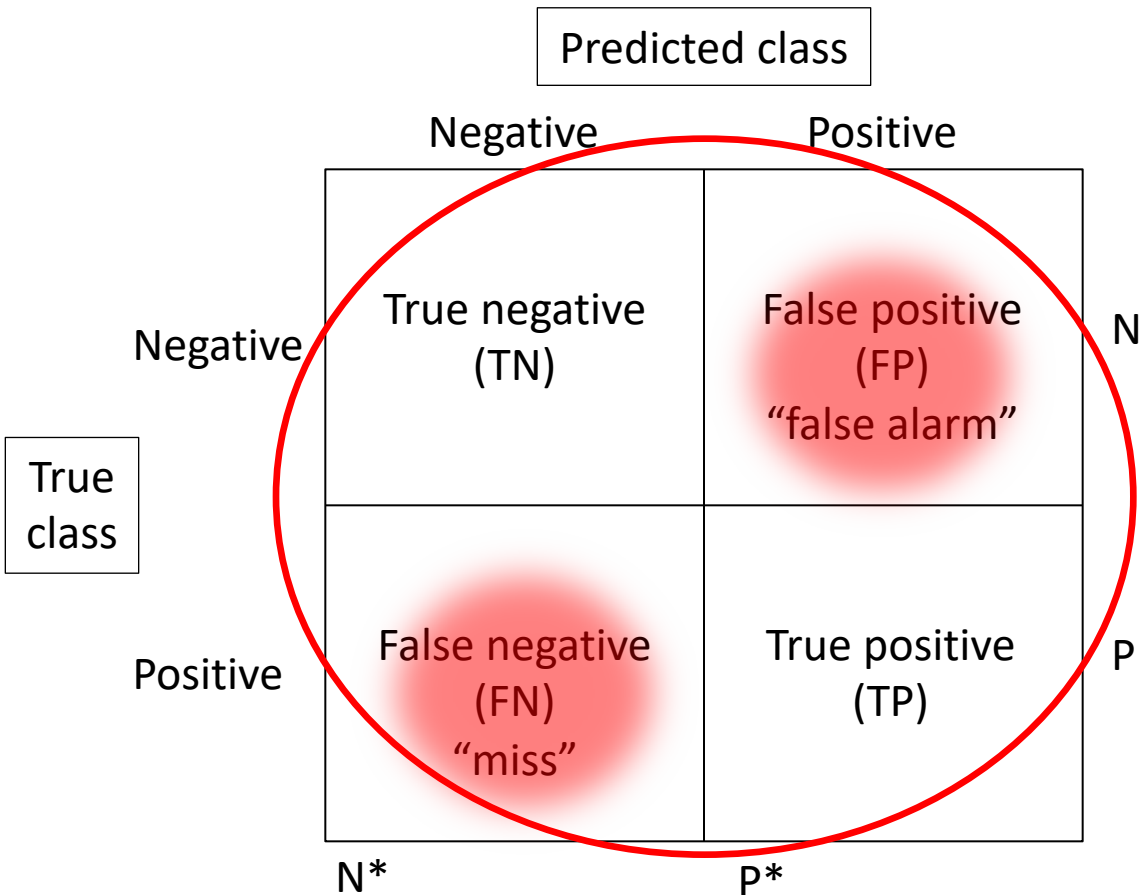
N* (what we said
was negative)

P* (what we said was
positive “flagged”)

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*	

Confusion Matrices

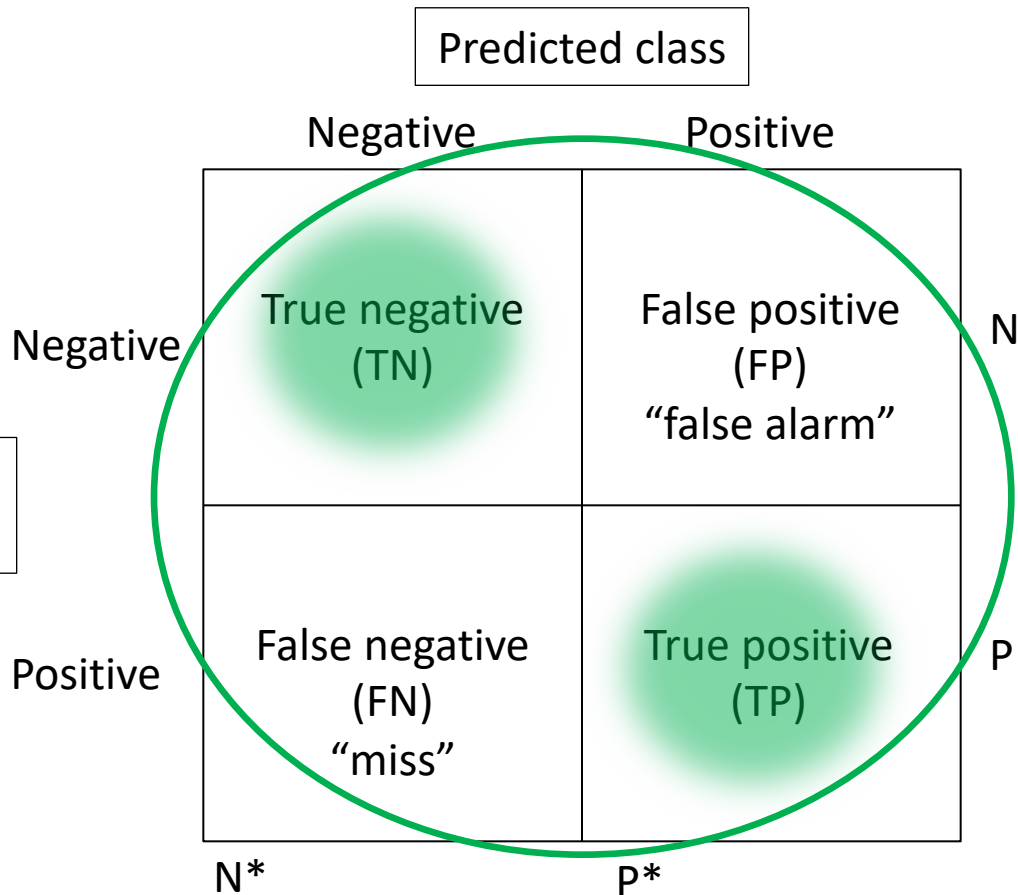


Error:

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

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

Confusion Matrices



Accuracy = 1-Error:

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

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

Confusion Matrices

		Predicted class	
		Negative	Positive
True class	Negative	True negative (TN)	False positive (FP) "false alarm"
	Positive	False negative (FN) "miss"	True positive (TP)
		N*	p*

The diagram shows a 2x2 confusion matrix. The columns are labeled 'Negative' and 'Positive' under the heading 'Predicted class'. The rows are labeled 'Negative' and 'Positive' under the heading 'True class'. The four quadrants are: Top-Left: True negative (TN); Top-Right: False positive (FP) with the note "false alarm"; Bottom-Left: False negative (FN) with the note "miss"; Bottom-Right: True positive (TP). A purple oval highlights the right column (FP and TP), and a purple gradient circle highlights the TP cell. Marginal counts are N* for the bottom row and p* for the right column. On the right side of the matrix, 'N' is next to the top row and 'P' is next to the bottom row.

Precision:

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

Confusion Matrices

		Predicted class	
		Negative	Positive
True class	Negative	True negative (TN)	False positive (FP) "false alarm"
	Positive	False negative (FN) "miss"	True positive (TP)
		N*	p*

The diagram shows a 2x2 confusion matrix. The top row is labeled 'Negative' and the bottom row is labeled 'Positive' under the 'True class' header. The left column is labeled 'Negative' and the right column is labeled 'Positive' under the 'Predicted class' header. The cells contain: True negative (TN), False positive (FP) "false alarm", False negative (FN) "miss", and True positive (TP). The TP cell is highlighted with a blue oval. Marginal counts N* and p* are shown at the bottom, and N and P are shown on the right side of the matrix.

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"
	Positive	False negative (FN) "miss"	True positive (TP)
		N*	p*

The table is a 2x2 matrix. The top row is labeled 'Predicted class' and has columns for 'Negative' and 'Positive'. The left column is labeled 'True class' and has rows for 'Negative' and 'Positive'. The top-right cell (FP) is highlighted with a brown oval and a gradient. The bottom-left cell (FN) is labeled 'miss'. Marginal counts are N* for the bottom row and p* for the right column. The total number of negative true class samples is N, and the total number of positive predicted samples is p.

False Positive Rate:

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

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

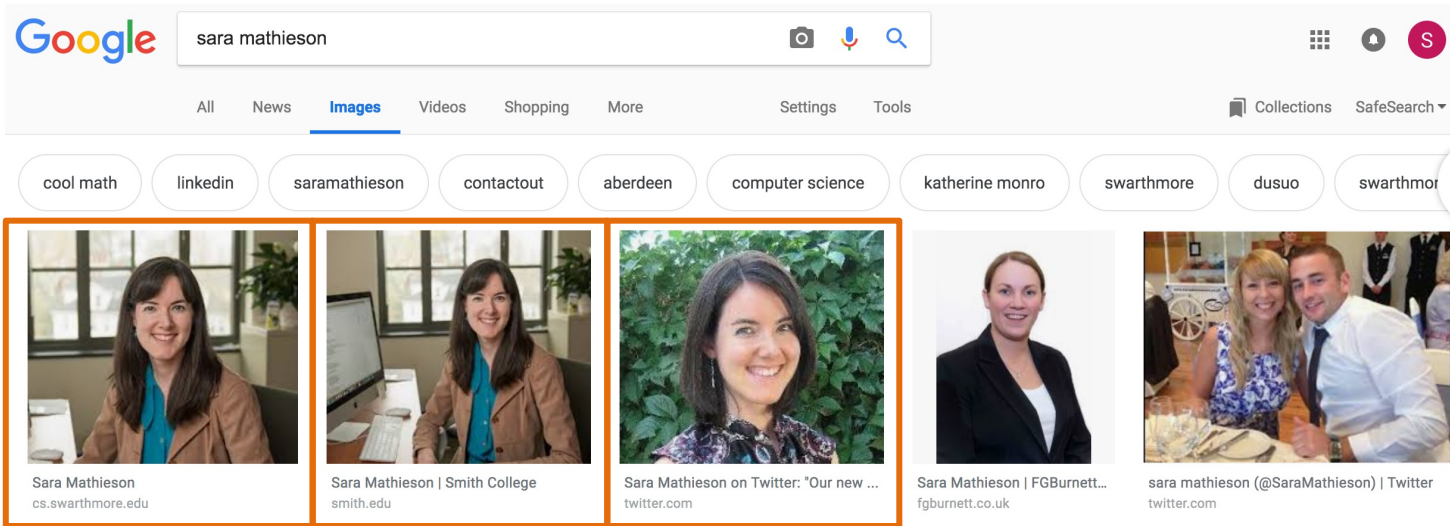
The screenshot shows a Google search for "sara mathieson" with the "Images" tab selected. The search bar contains "sara mathieson" and the Google logo is on the left. Below the search bar are navigation tabs for "All", "News", "Images", "Videos", "Shopping", and "More". To the right are "Settings" and "Tools". Further right are "Collections" and "SafeSearch". Below the navigation are several filter buttons: "cool math", "linkedin", "saramathieson", "contactout", "aberdeen", "computer science", "katherine monro", "swarthmore", "dusuo", and "swarthmor". The search results are displayed in a grid of five images. Each image has a caption below it:

- Image 1: Sara Mathieson, cs.swarthmore.edu
- Image 2: Sara Mathieson | Smith College, smith.edu
- Image 3: Sara Mathieson on Twitter: "Our new ...", twitter.com
- Image 4: Sara Mathieson | FGBurnett..., fgburnett.co.uk
- Image 5: sara mathieson (@SaraMathieson) | Twitter, twitter.com

$P=6$ (number of images that are actually me)

- Precision?
- Recall?

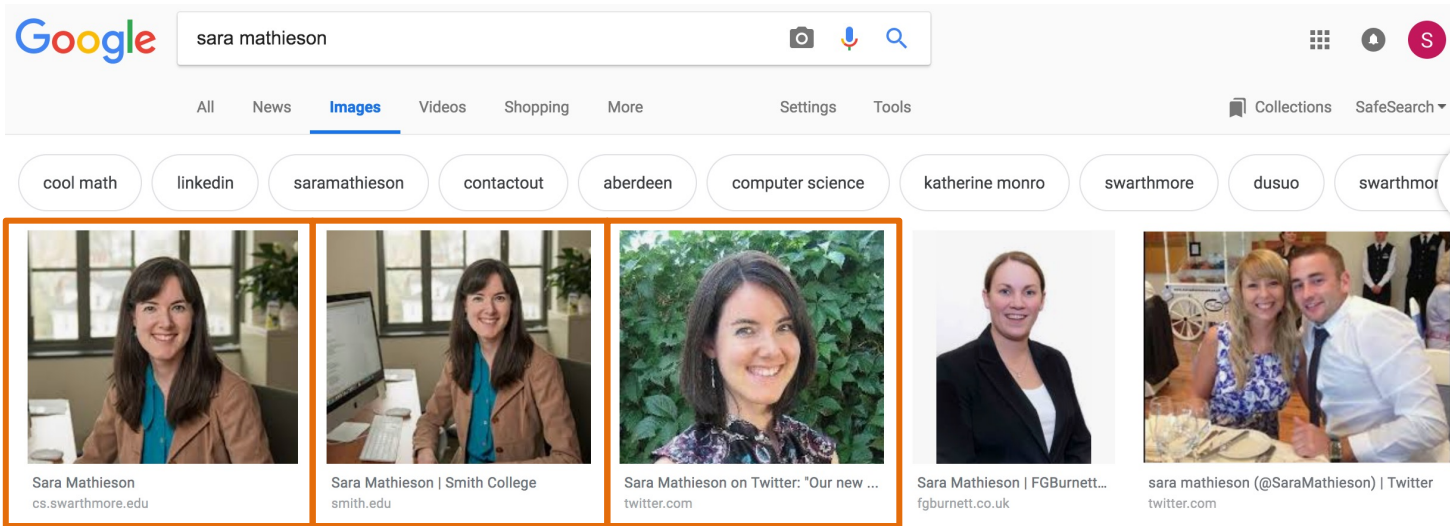
Precision and Recall



$P=6$ (number of images that are actually me)

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

Precision and Recall



$P=6$ (number of images that are actually me)

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

Precision and Recall

The screenshot shows a Google search for 'sara mathieson' with 16 image results. The results are arranged in a grid. Six results are highlighted with orange borders, indicating they are true positives (actual images of Sara Mathieson). The other ten results are false positives (images of other people).

True Positives (6):

- Sara Mathieson, cs.swarthmore.edu
- Sara Mathieson | Smith College, smith.edu
- Sara Mathieson on Twitter: "Our new ...", twitter.com
- Sara Mathieson - Graham + Si..., g-s.co.uk
- Sara Mathieson (Saramathi...), github.com
- Sara Mathieson Email &..., contactout.com

False Positives (10):

- Sara Mathieson | FGBurnett..., fgburnett.co.uk
- sara mathieson (@SaraMathieson) | Twitter, twitter.com
- Sara Mathieson, cs.swarthmore.edu
- Top 25 Sara Mathieson profiles | Li..., linkedin.com
- Sara Mathieson - Graham + Si..., g-s.co.uk
- Sara Mathieson, cs.swarthmore.edu
- Top 25 Sara Mathieson profil..., linkedin.com
- Sara Mathieson, cs.swarthmore.edu
- Sara Mathieson (saramathi...), pinterest.com
- Sara Mathieson Email &..., contactout.com

$P=6$ (number of images that are actually me)

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

Precision/Recall for google example

precision / recall \Rightarrow numerator is the same!

	2	N
3	3	P=6

$$P^* = 5$$

$$\text{precision} = \frac{2}{5}$$

$$\text{recall} = \frac{2}{6} = 0.33$$
$$\text{TPR} = \frac{2}{6} = 0.33$$

	11	
1	5	P=6

$$P^* = 16$$

$$\text{precision} = \frac{5}{16}$$

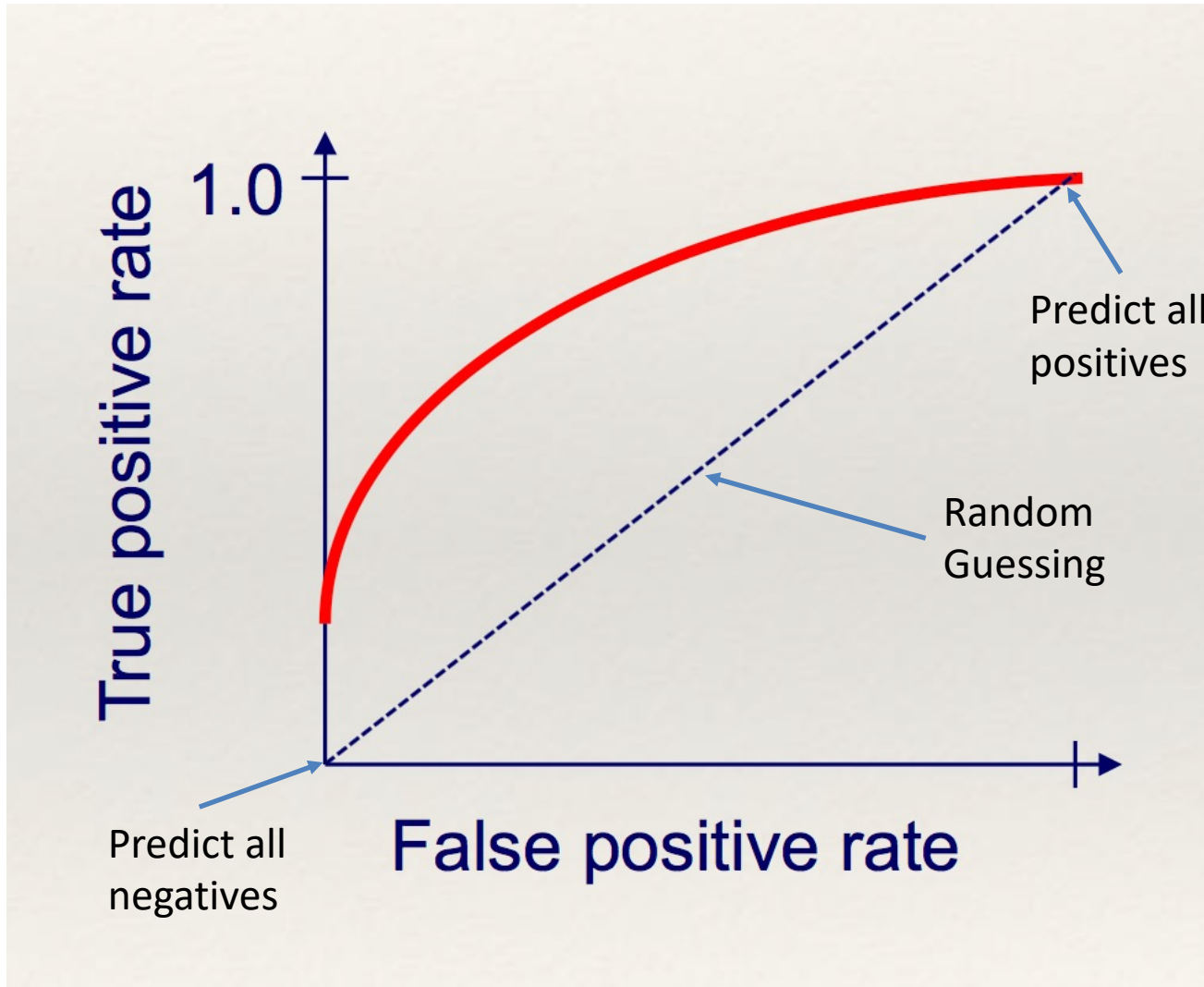
$$\text{recall} = \frac{5}{6}$$

☆ \uparrow

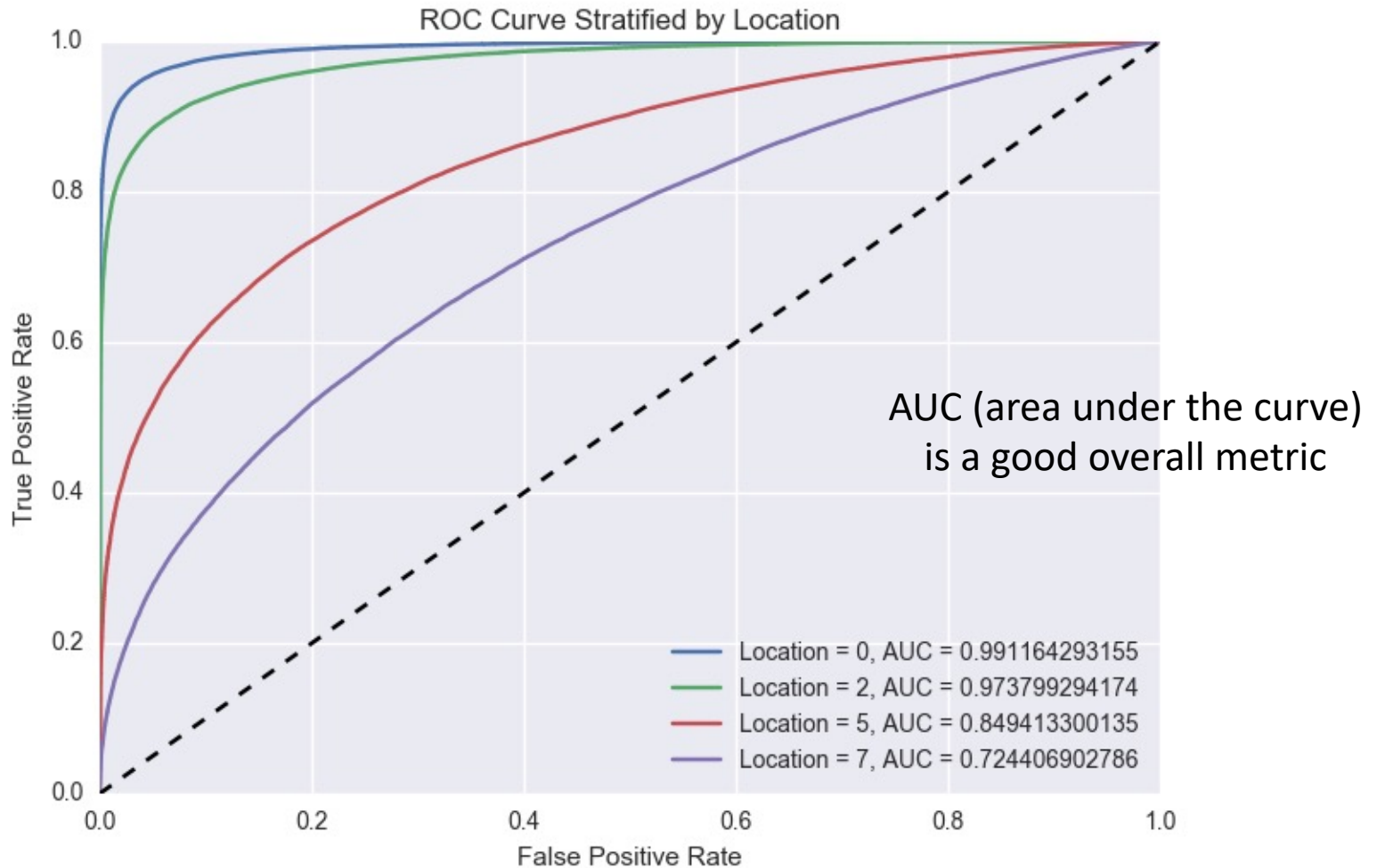
ROC curve (Receiver Operating Characteristic)

More history here!

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



ROC curve example: comparing methods



Example of a ROC curve from my research
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,...)
 - $P(y=1 \mid x) \geq 0.2$ \Rightarrow classify as 1 (positive)
 - $P(y=1 \mid x) < 0.2$ \Rightarrow classify as 0 (negative)

Handout 8

Handout 8

	-	+	
-	77	3	$N = 80$
+	13	7	$P = 20$

$$N^* = 90 \quad P^* = 10$$

$$\text{precision} = \frac{7}{10}$$

$$\text{recall} = \frac{7}{20}$$

$$0.35$$

$$\text{FPR} = \frac{3}{80}$$

"Y"

	-	+	
-	68	12	
+	2	18	$P = 20$

$$P^* = 30$$

FPR
TPR



Handout 8

