

CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019



Outline for March 1

- Evaluation metrics
 - Precision and Recall
 - Confusion matrices revisited
 - ROC curves
 - Relationship to probabilistic methods
 - Cross-validation
- **Lab 4 due March 8** (week from today)
- Office hours **TODAY** 1-3pm

Outline for March 1

- Evaluation metrics
 - Precision and Recall
 - Confusion matrices revisited
 - ROC curves
 - Relationship to probabilistic methods
 - Cross-validation

Thanks for all your work on the exam!



Goals of Evaluation

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

Precision and Recall

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

Precision and Recall

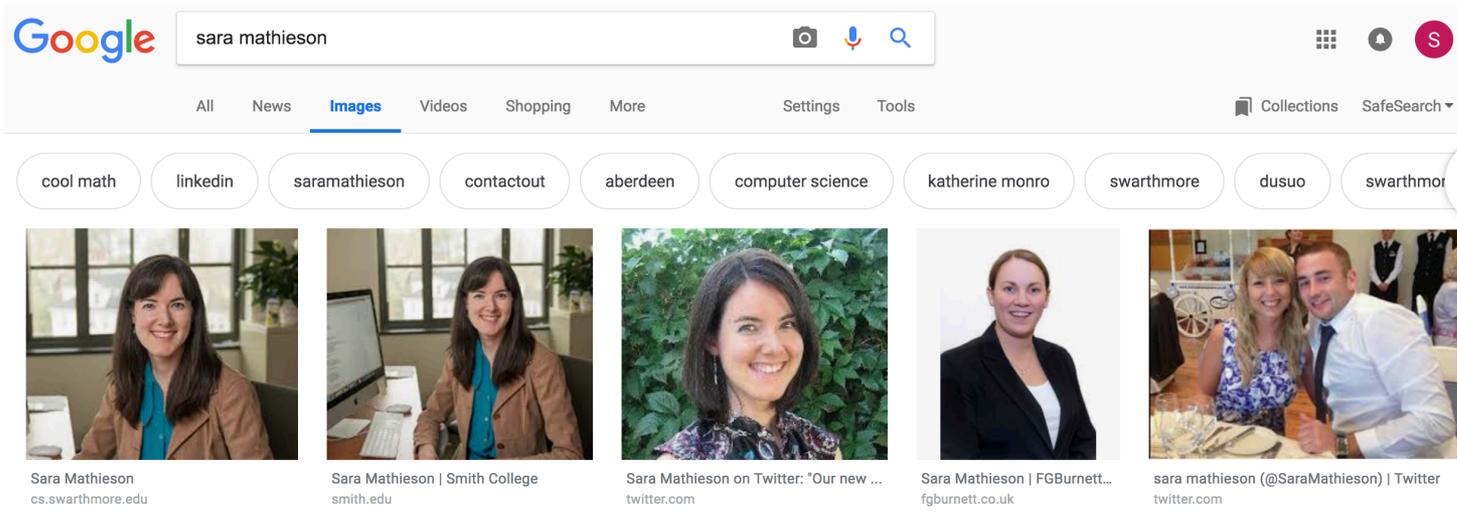
- Precision: of all the “flagged” examples, which ones are actually relevant (i.e. positive)?

(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 tabs 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 with a caption below it:

- 

Sara Mathieson
cs.swarthmore.edu
- 

Sara Mathieson | Smith College
smith.edu
- 

Sara Mathieson on Twitter: "Our new ...
twitter.com
- 

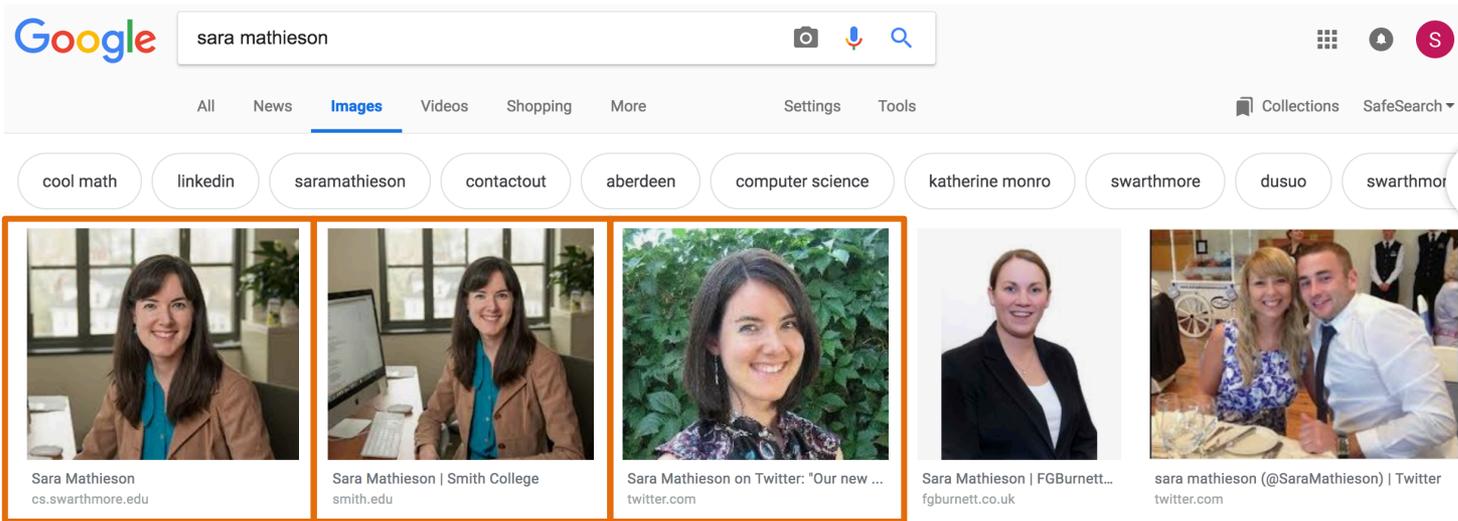
Sara Mathieson | FGBurnett...
fgburnett.co.uk
- 

sara mathieson (@SaraMathieson) | Twitter
twitter.com

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

- Precision?
- Recall?

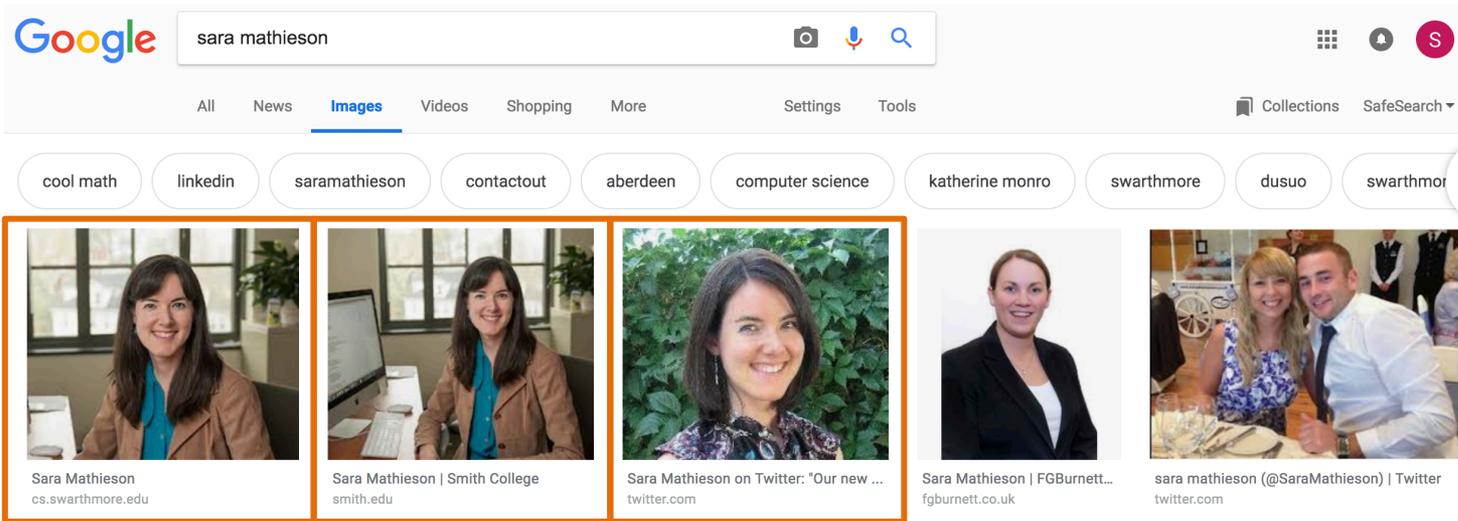
Precision and Recall



$P=6$ (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 search bar contains "sara mathieson" and the search filters are set to "Images". The results are displayed in a grid. Six images are highlighted with an orange border:

- Row 1, Column 1: Sara Mathieson, cs.swarthmore.edu
- Row 1, Column 2: Sara Mathieson | Smith College, smith.edu
- Row 1, Column 3: Sara Mathieson on Twitter: "Our new ...", twitter.com
- Row 2, Column 1: Sara Mathieson - Graham + Si..., g-s.co.uk
- Row 2, Column 2: Sara Mathieson, cs.swarthmore.edu
- Row 3, Column 3: saramathieson (Sara Mathieson...), github.com

The other 10 images are not highlighted. The search interface includes the Google logo, search bar, filters (All, News, Images, Videos, Shopping, More), and navigation options (Settings, Tools, Collections, SafeSearch).

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

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

Outline for March 1

- Evaluation metrics
 - Precision and Recall
 - **Confusion matrices revisited**
 - ROC curves
 - Relationship to probabilistic methods
 - Cross-validation

Recap Confusion Matrices

Predicted class

Negative

Positive

Negative

True negative
(TN)

False positive
(FP)

True
class

Positive

False negative
(FN)

True positive
(TP)

Recap 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)

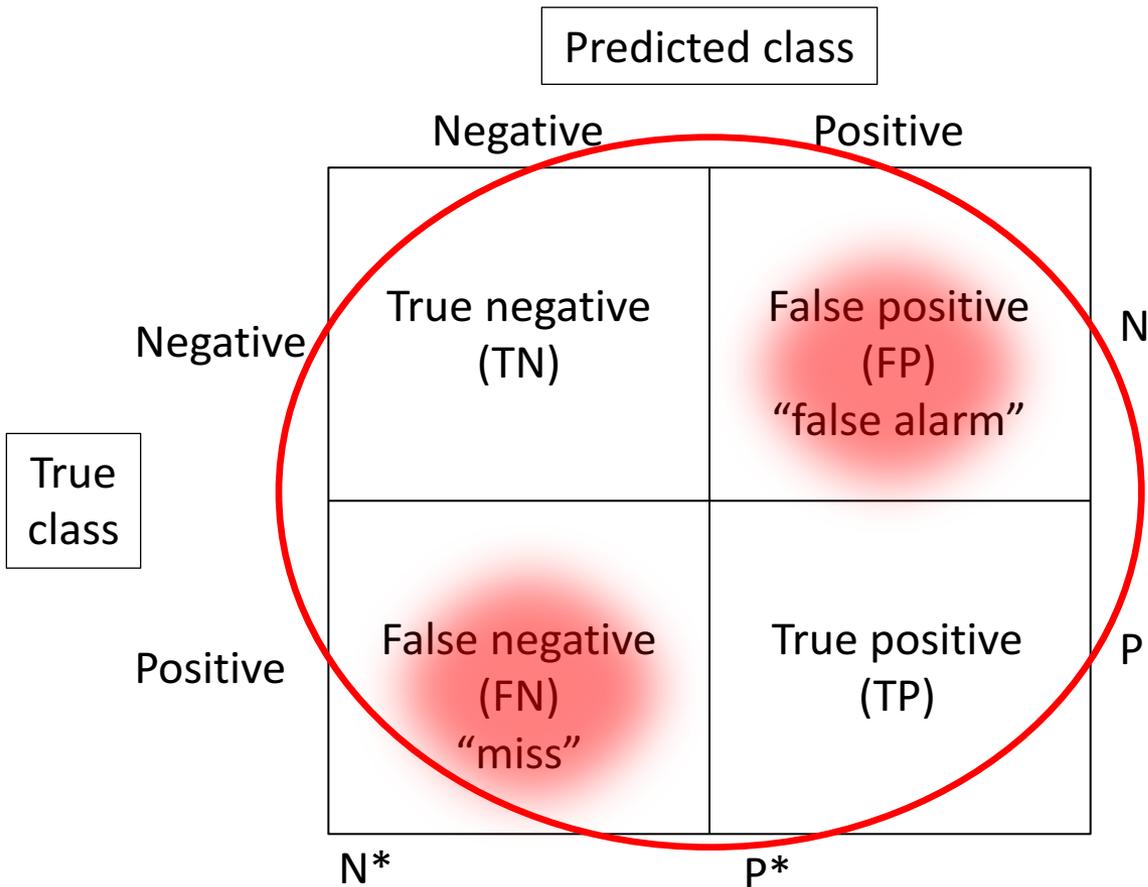
N* (what we said was negative)

P* (what we said was positive "flagged")

Recap 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*

Recap Confusion Matrices

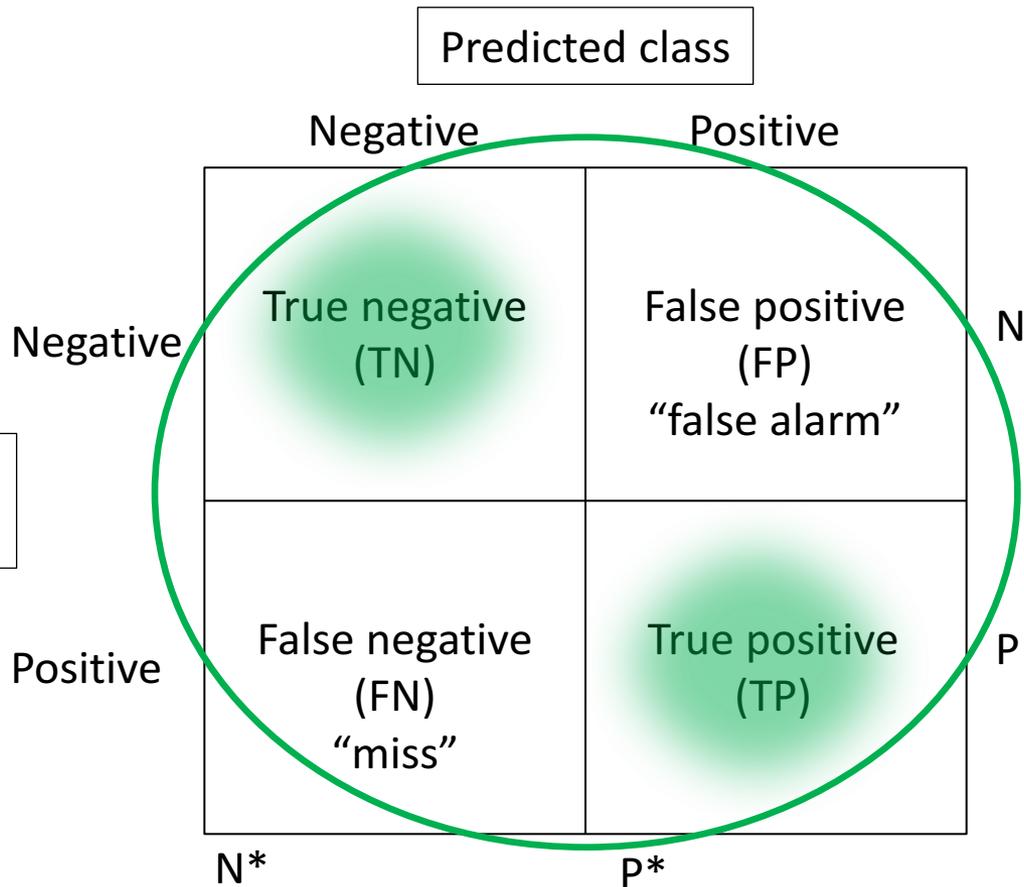


Error:

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

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

Recap Confusion Matrices



Accuracy = 1-Error:

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

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

Recap 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 cells contain: True negative (TN), False positive (FP) "false alarm", False negative (FN) "miss", and True positive (TP). A purple oval highlights the FP and TP cells. Marginal counts N* and p* are shown at the bottom, and N and P are shown on the right side.

Precision:

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

Recap 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 bottom row is circled in blue. Marginal counts N* and p* are shown at the bottom, and N and P are shown on the right.

Recall
(True Positive Rate):

$$TP/(FN+TP) = TP/P$$

Recap 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 'Positive' under the 'True class' label. The left column is labeled 'Negative' and the right column 'Positive' under the 'Predicted class' label. The cells contain: True negative (TN), False positive (FP) "false alarm", False negative (FN) "miss", and True positive (TP). Marginal counts are N* for the bottom row and p* for the right column. The top-right cell (FP) is highlighted with a brown oval and a gradient background.

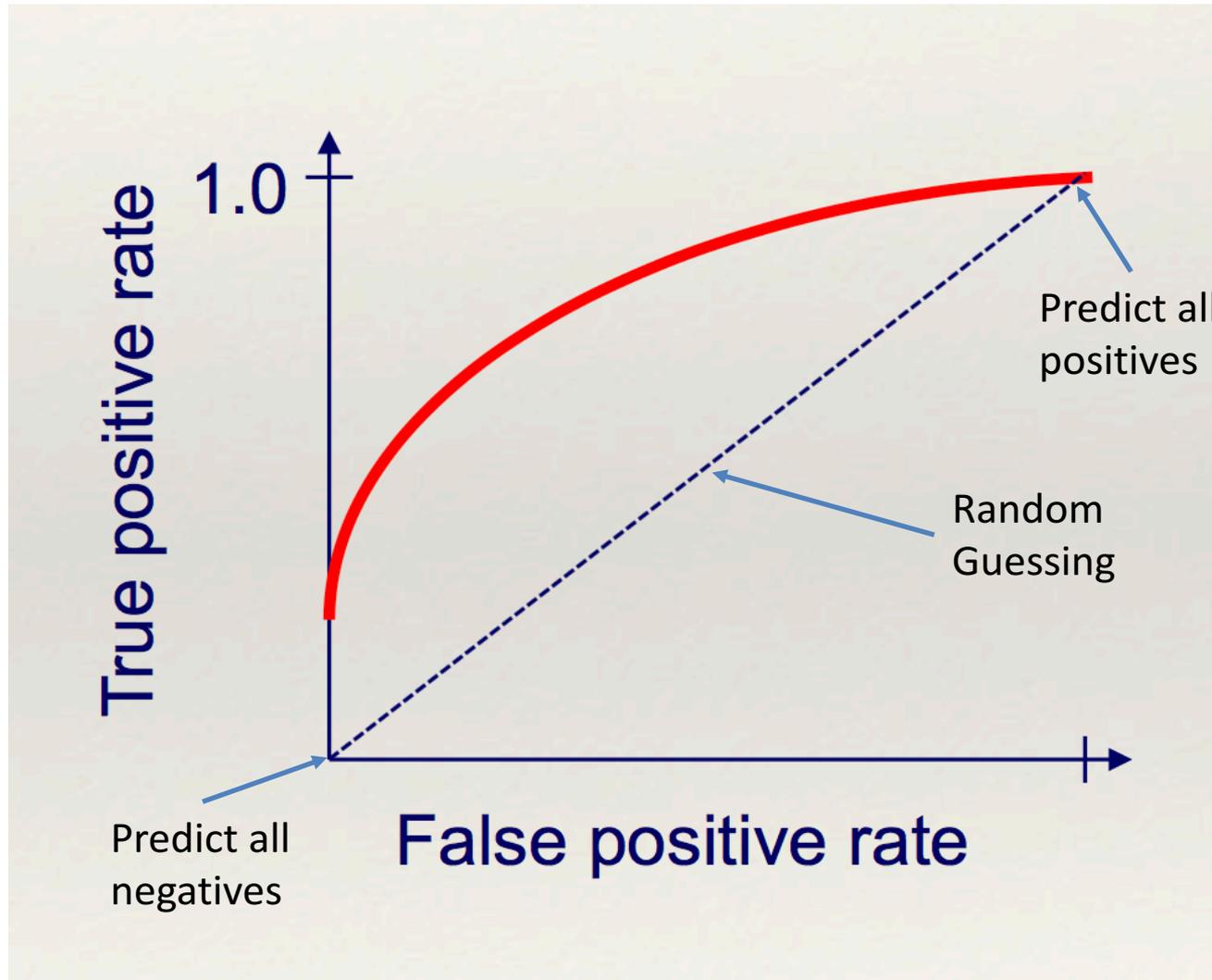
False Positive Rate:

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

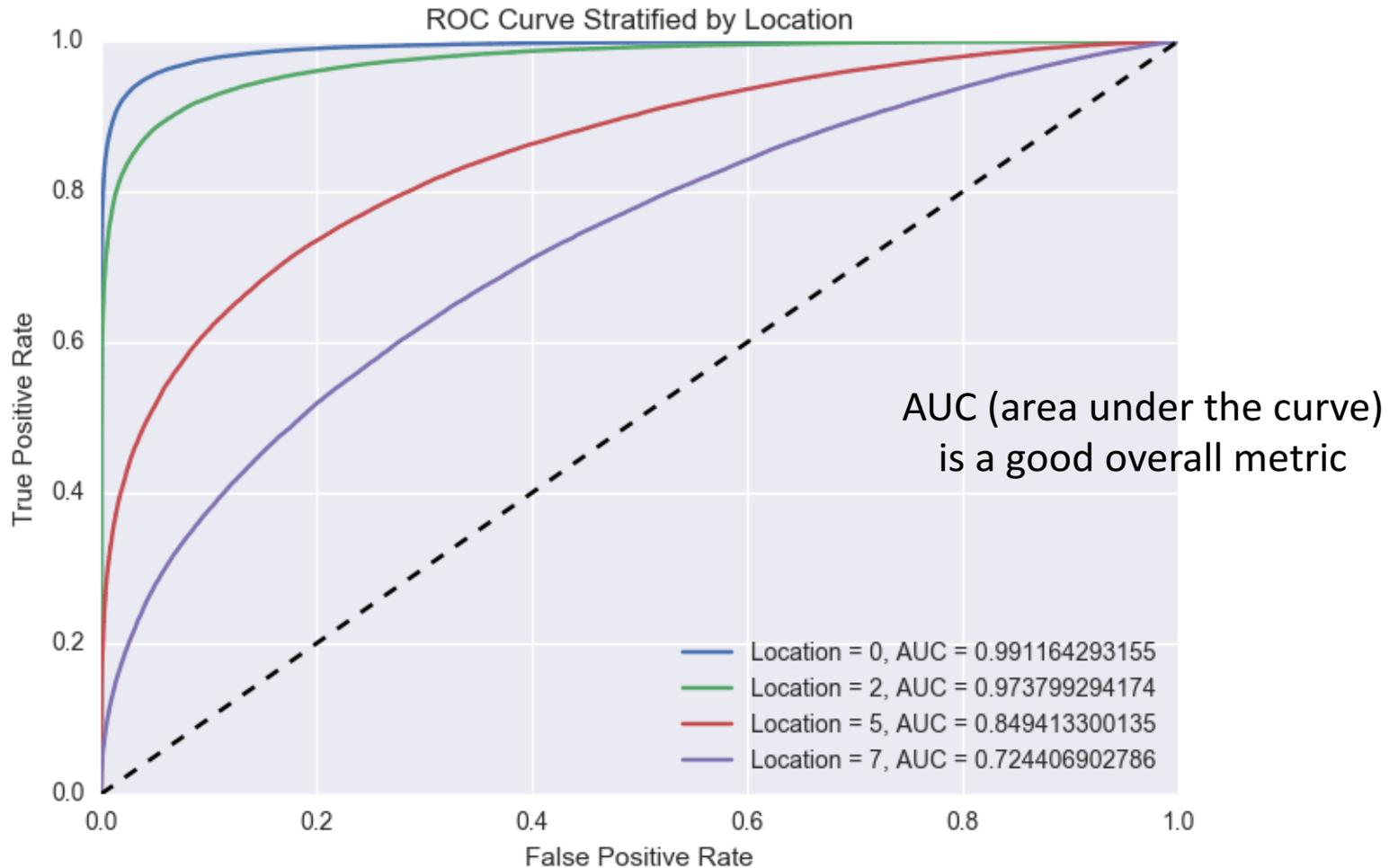
Outline for March 1

- Evaluation metrics
 - Precision and Recall
 - Confusion matrices revisited
 - ROC curves
 - Relationship to probabilistic methods
 - Cross-validation

ROC curve (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 0.25)
 - $P(y=1 \mid x) > 0.25 \quad \Rightarrow$ classify as 1 (positive)
 - $P(y=1 \mid x) \leq 0.25 \quad \Rightarrow$ classify as 0 (negative)

✱

55	35
2	21

$$FPR = \frac{35}{90}$$

$$TPR = \frac{21}{23}$$

true

	<u>pred</u>		
	neg	pos	
neg	70	(20)	N = 90
pos	5	(18)	
			N* = 75 P* = 38

$$\text{accuracy} = \frac{88}{113}$$

$$FPR = \frac{20}{70+20} \approx .22$$

$$TPR = \frac{18}{5+18} \approx .78$$

	n	r	p
n	0	90	
p	0	23	

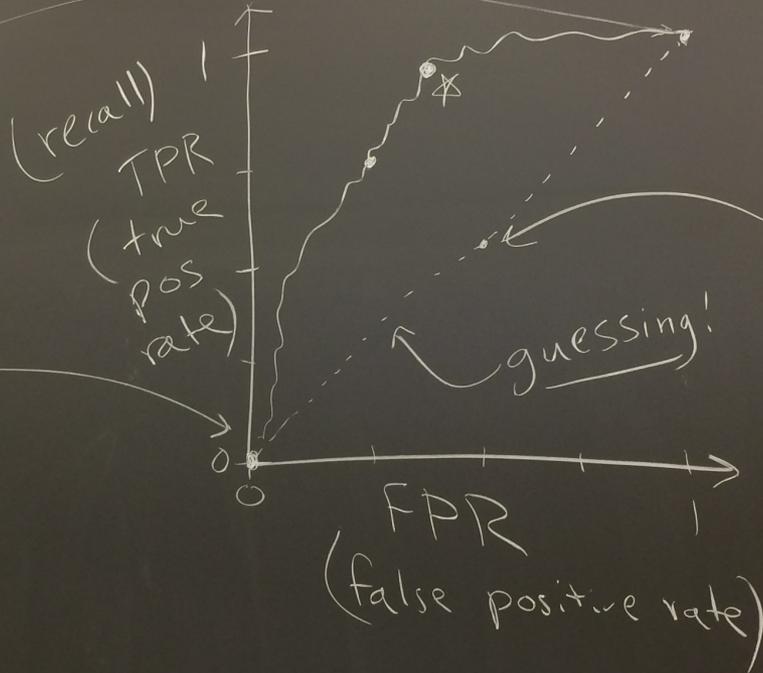
all positive

	n	p
n	90	0
p	23	0

all negative

(could be best accuracy)

$$\frac{90}{113}$$



guessing

45	45
11	12

Handout 7

(a)

	n	p	
n	77	3	80
p	13	7	20
	90	10	

$$\text{FPR} = \frac{3}{77+3} = .04$$

precision: $\frac{7}{10} = .7$

recall: $\frac{7}{20} = .35$

(d)

	80	
68	12	80
2	18	20
70	30	

$$\text{FPR} = \frac{12}{80}$$

$$\text{TPR} = \frac{18}{20} = .9$$

precision: $\frac{18}{30}$

