CS 260: Foundations of Data Science

Prof. Sara Mathieson

Fall 2023

HAVERFORD
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

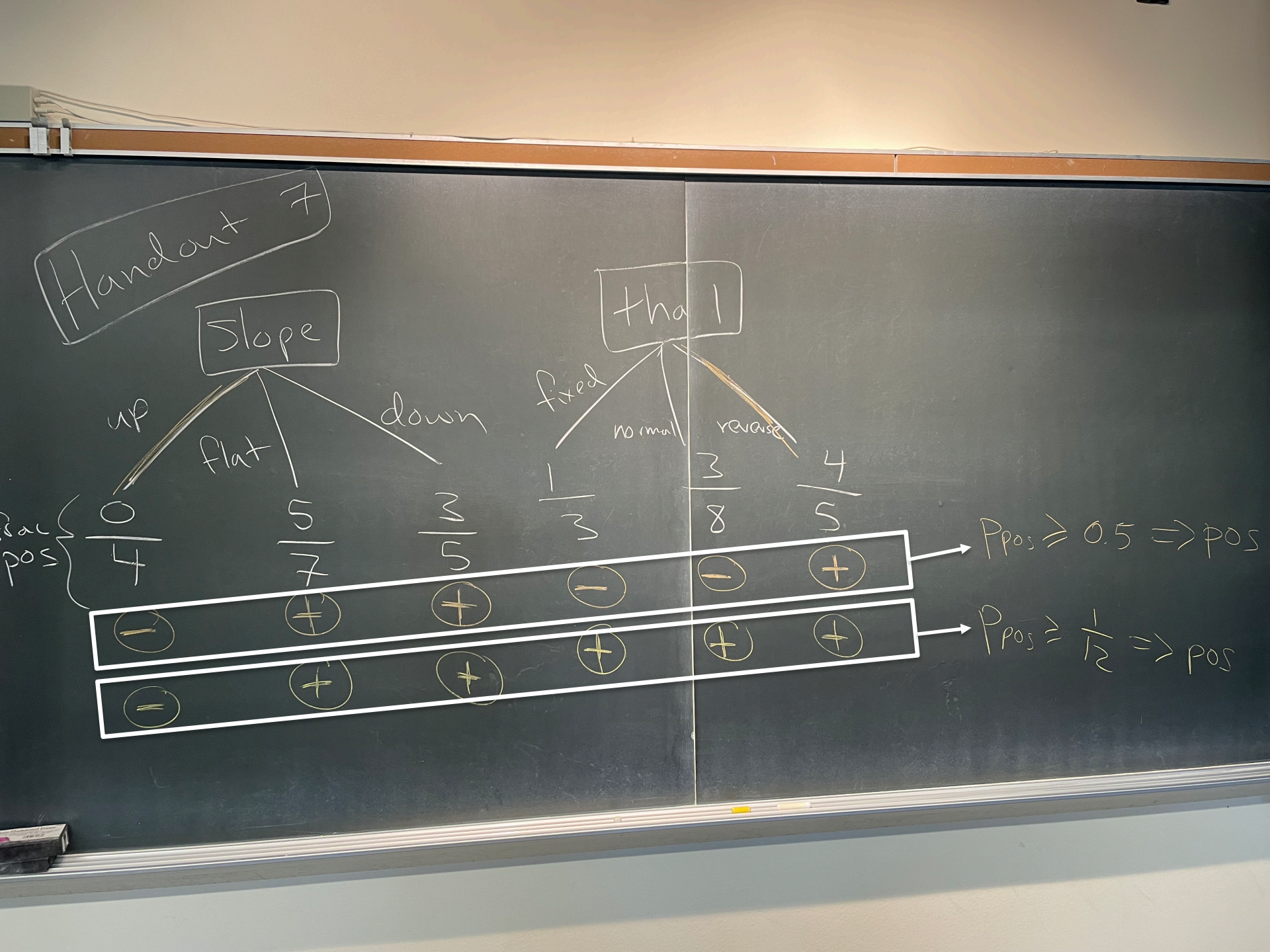
- Begin: Bayesian probability
  - Clinical trials example

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
thal
fixed
normal
reverse
0/4
5/7
3/5
1/3
3/8
4/5
Pac pos
Ppos ≥ 0.5 ⇒ pos
Ppos ≥ 1/2 ⇒ 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

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

truth \ pred	−	+
−	65	15
+	7	13

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

$$\text{accuracy} = \frac{\#\text{correct}}{m}$$

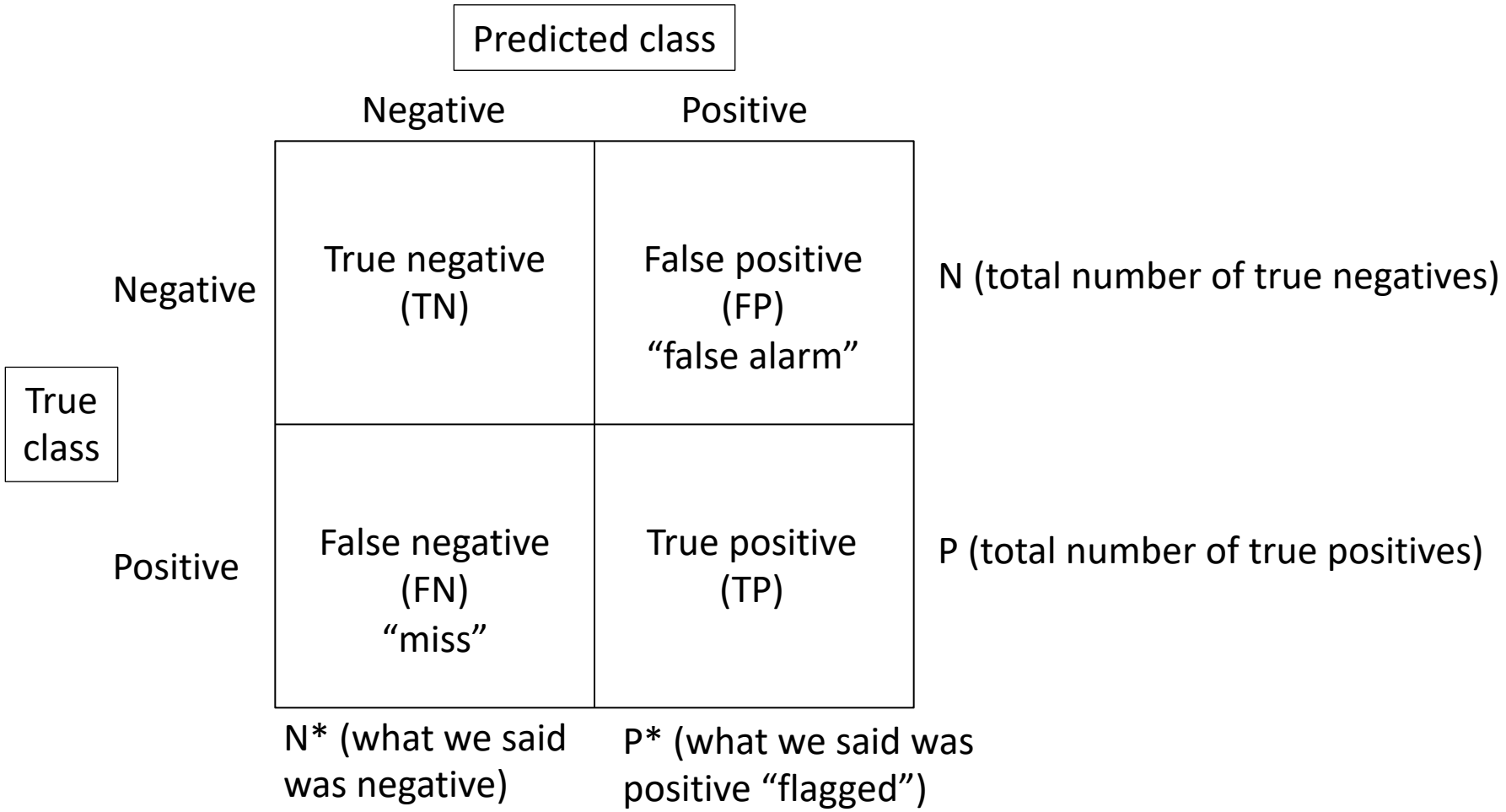
$$= \left[\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(\hat{y}_i == y_i) \right]$$

Note: all the same model, different thresholds!

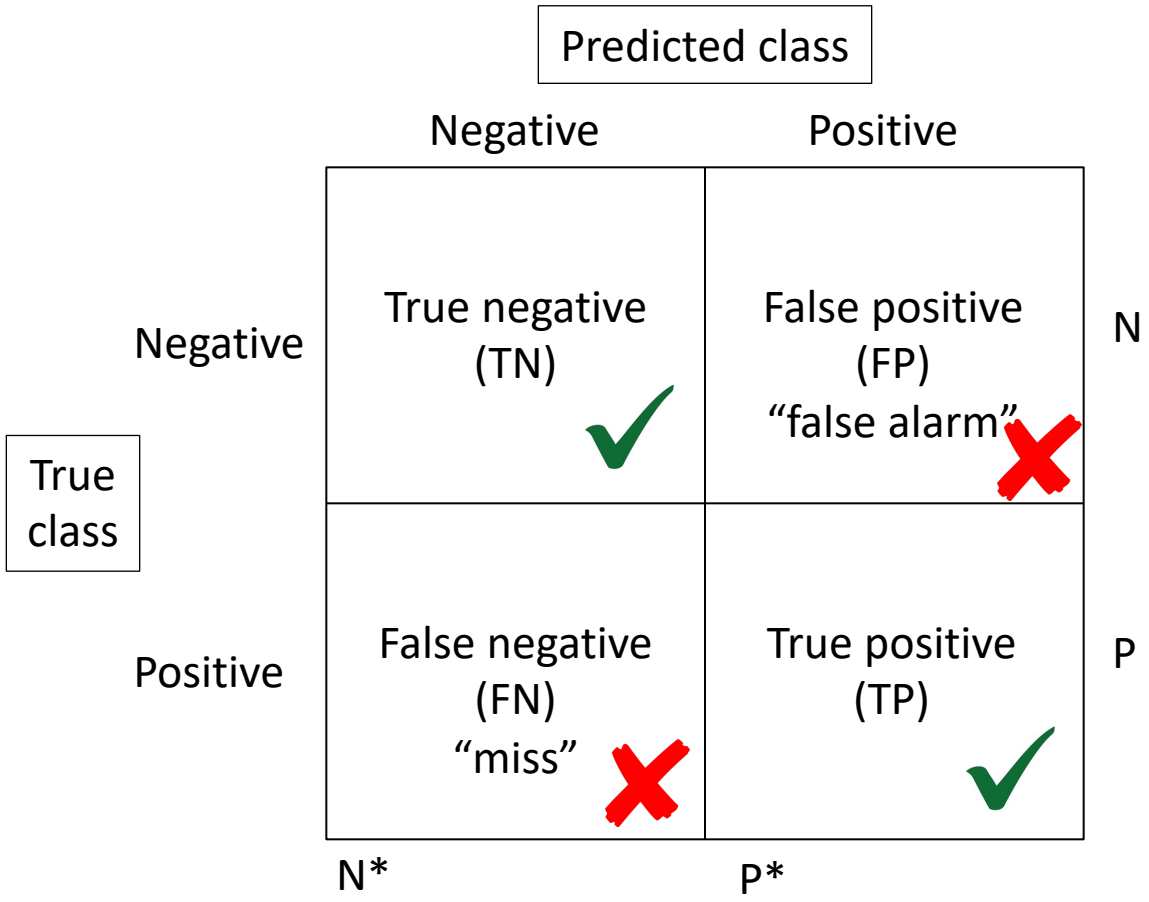
Confusion Matrices

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

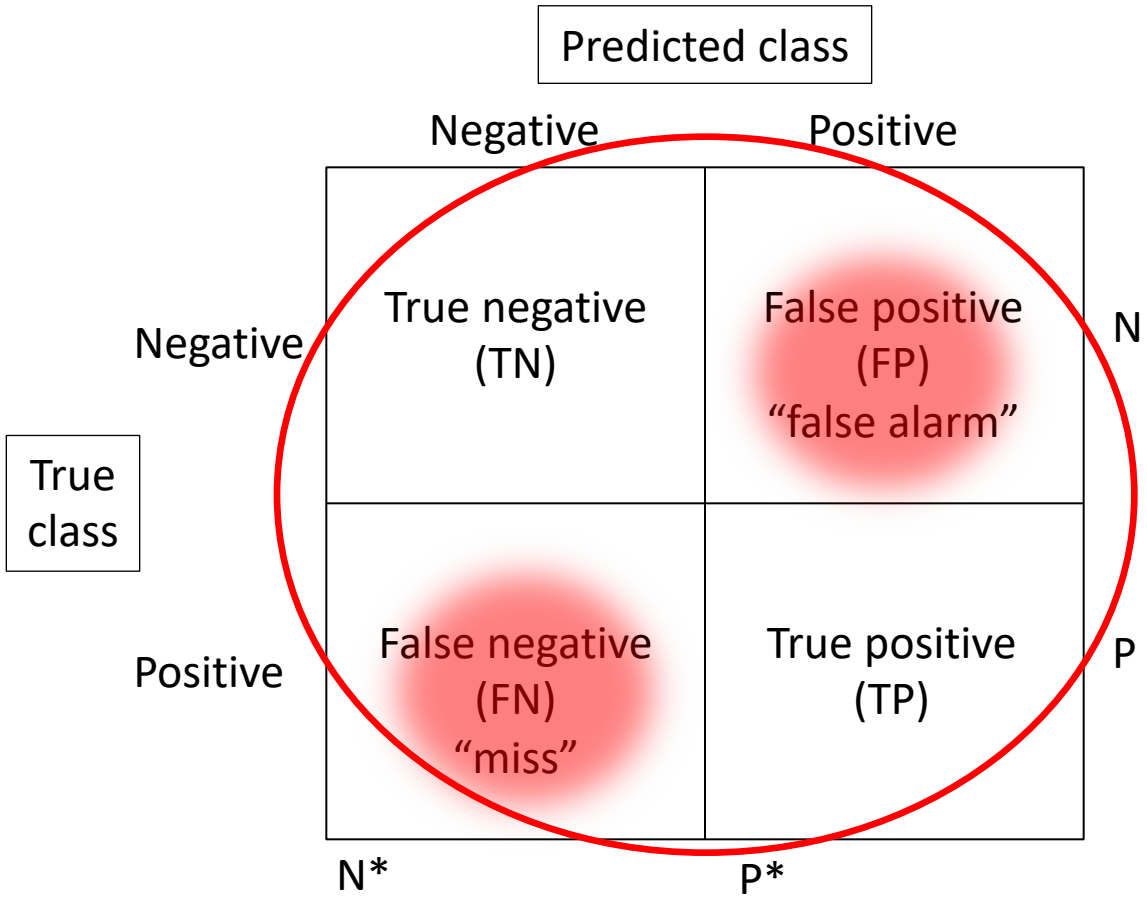
Confusion Matrices



Confusion Matrices



Confusion Matrices

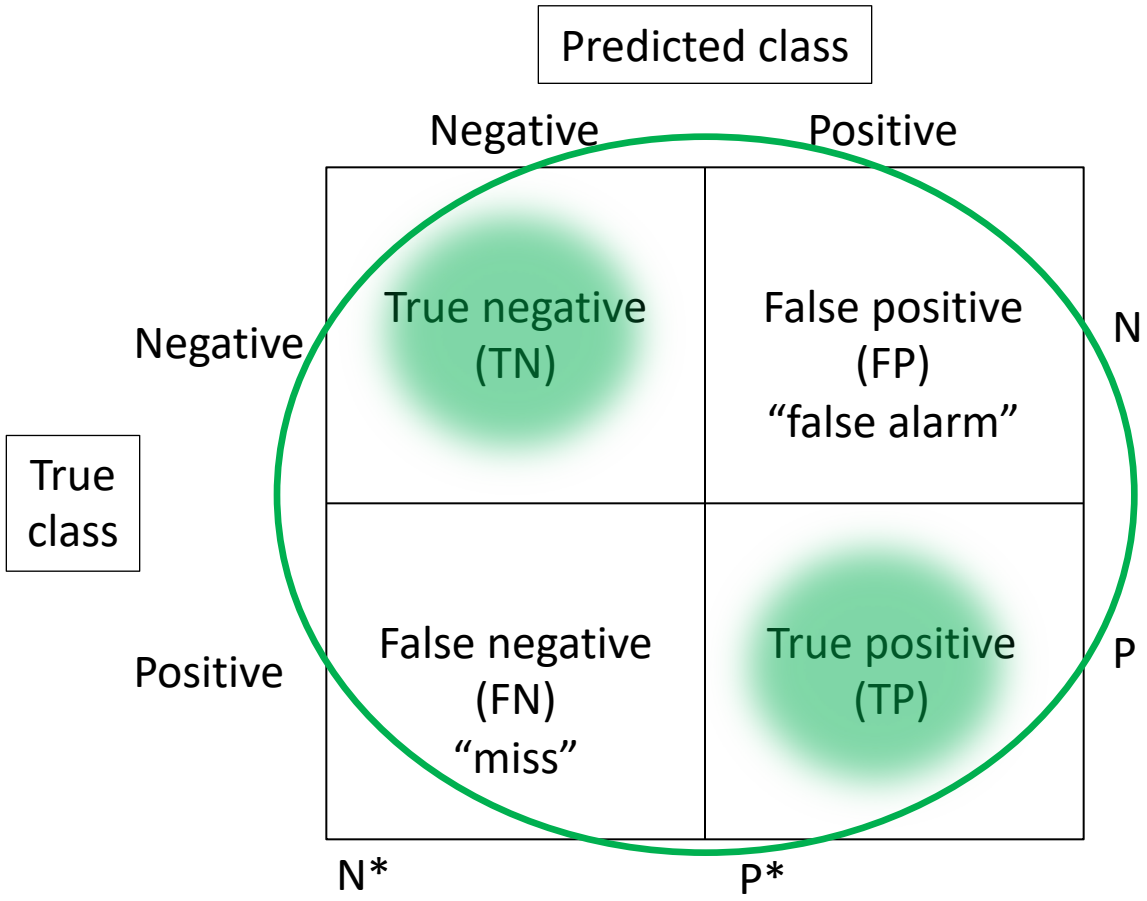


Error:

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

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

Confusion Matrices

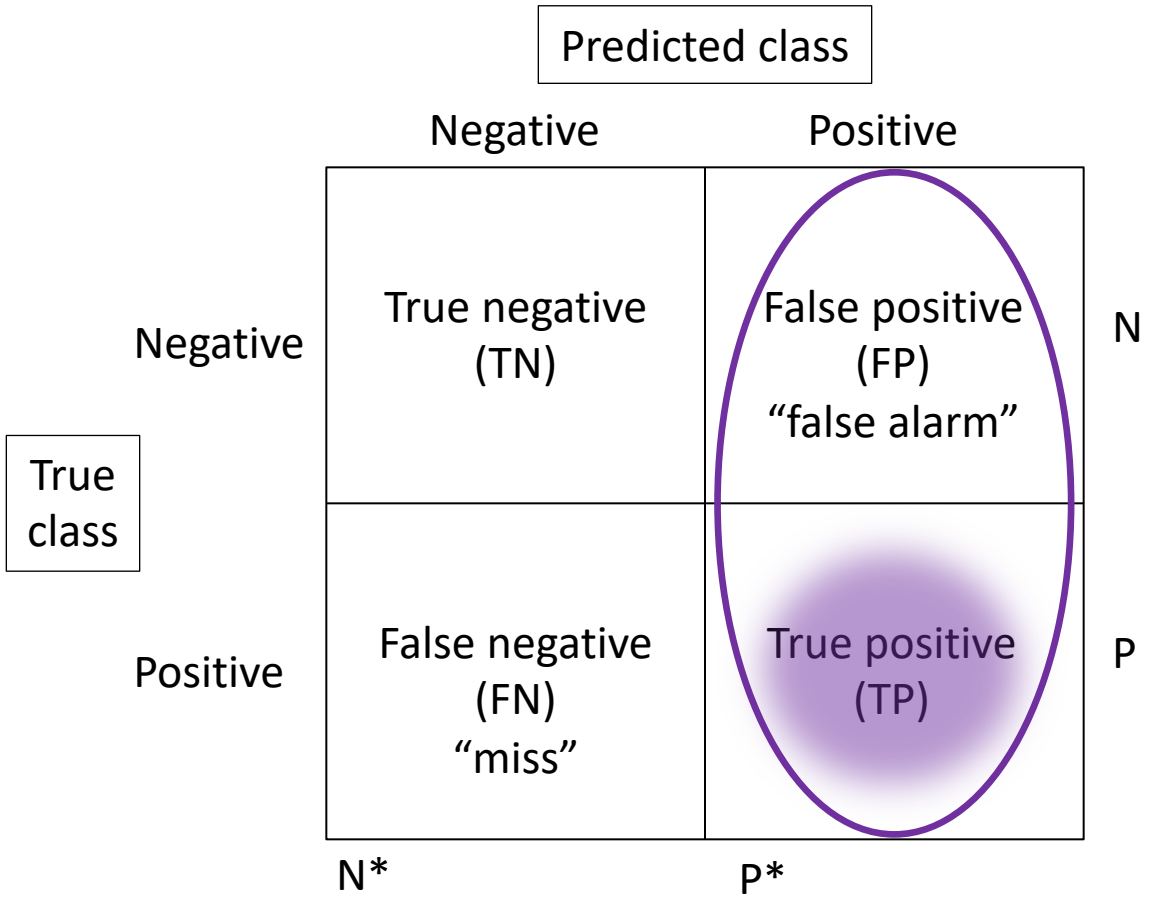


Accuracy = 1-Error:

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

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

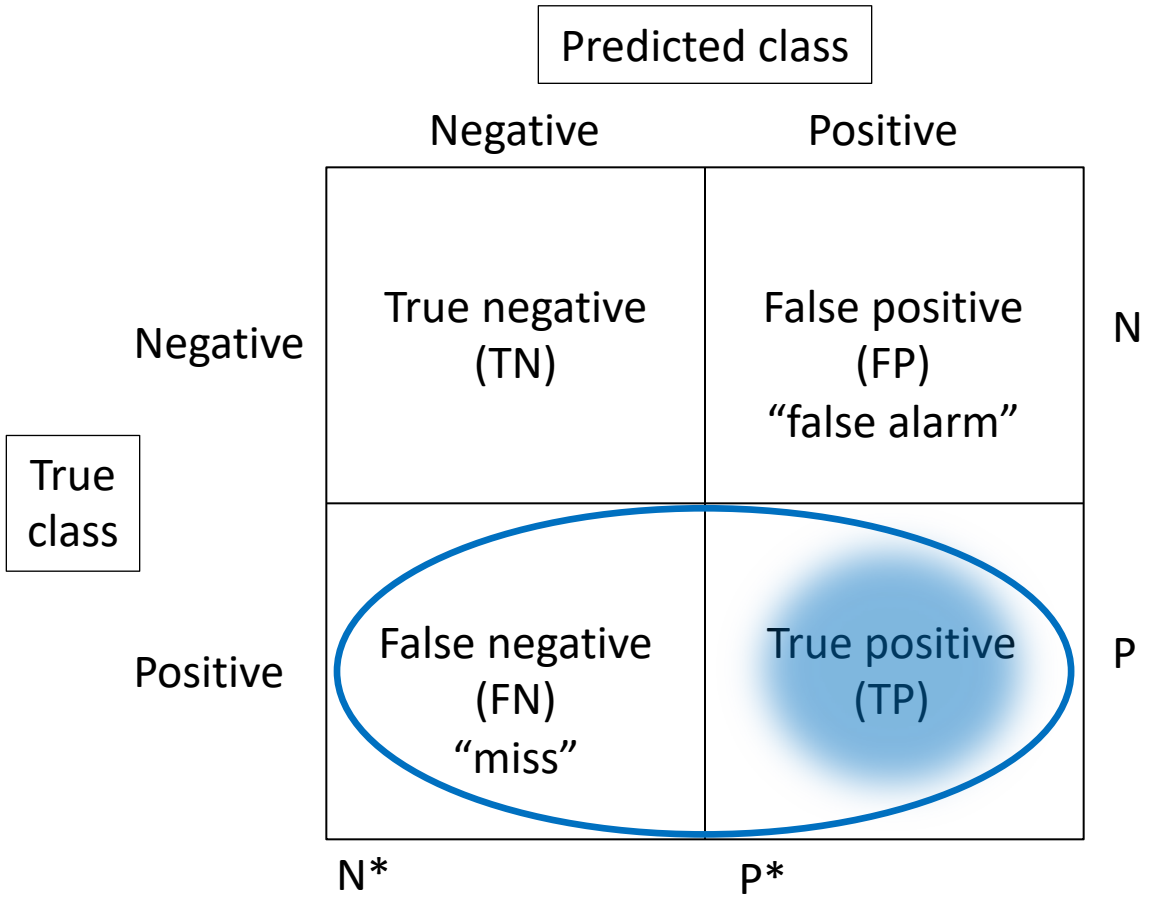
Confusion Matrices



Precision:

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

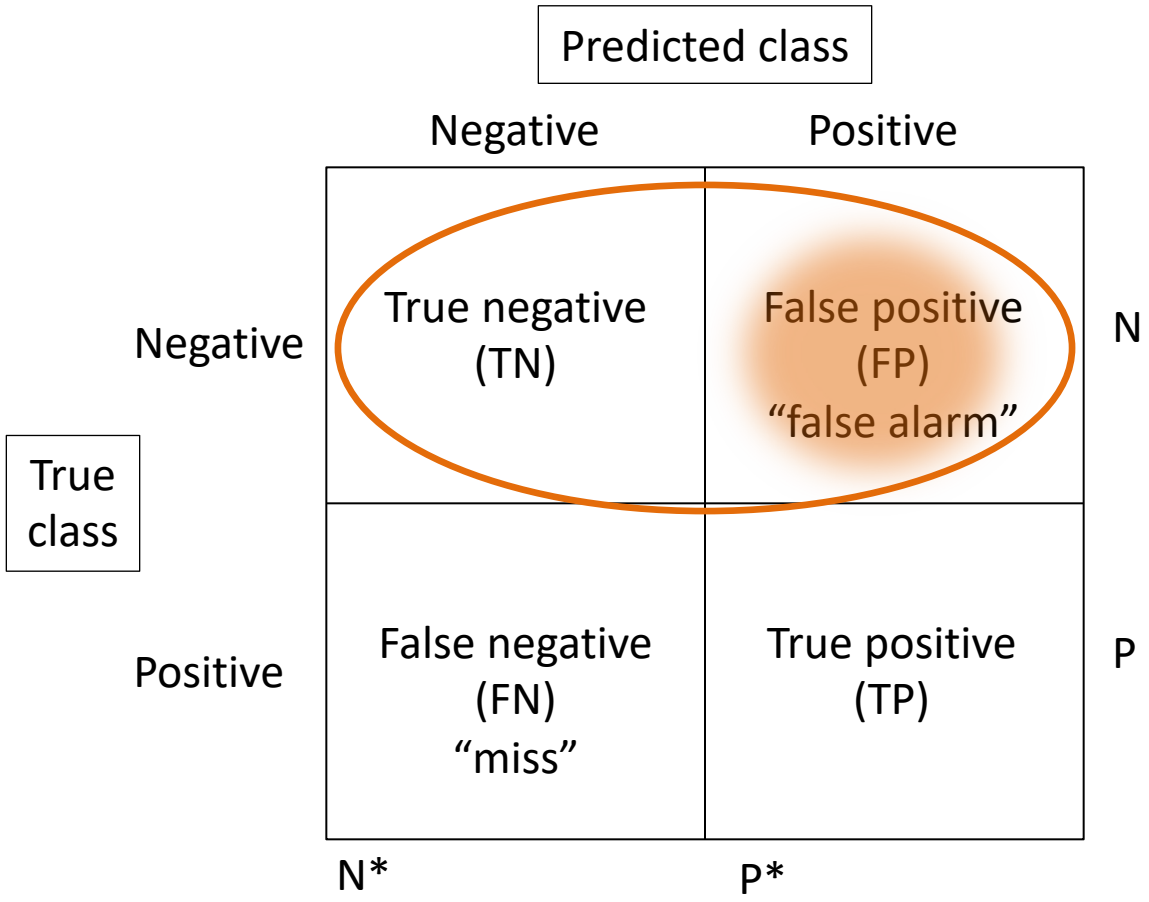
Confusion Matrices



Recall
(True Positive Rate):

TP/(FN+TP) = TP/P

Confusion Matrices



False Positive Rate:

FP/(TN+FP) = FP/N

Precision and Recall

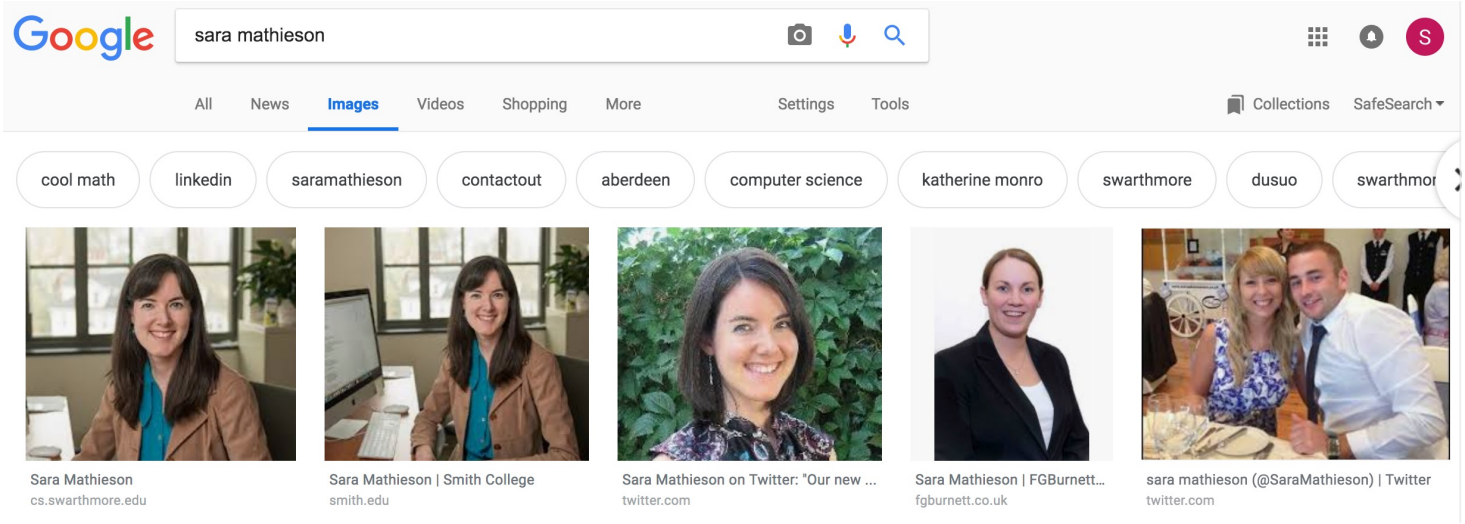
- Precision: of all the "flagged" examples, which ones are actually relevant (i.e. positive)?

  (Purity)

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

  (Completeness)

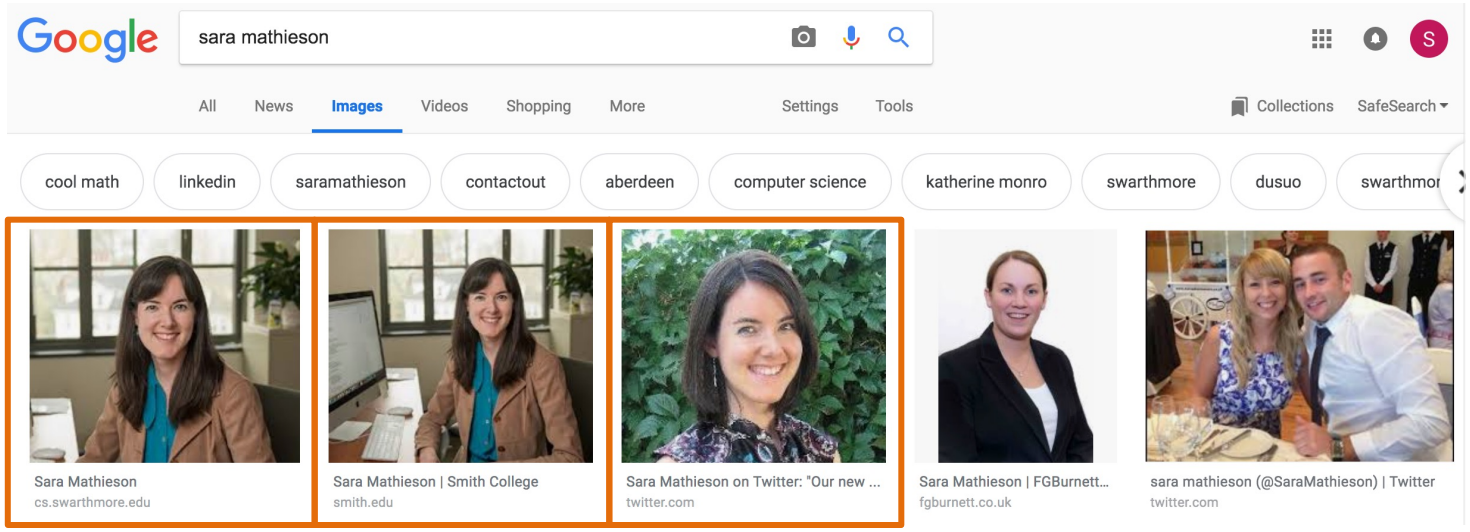
Precision and Recall



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

- Precision?
- Recall?

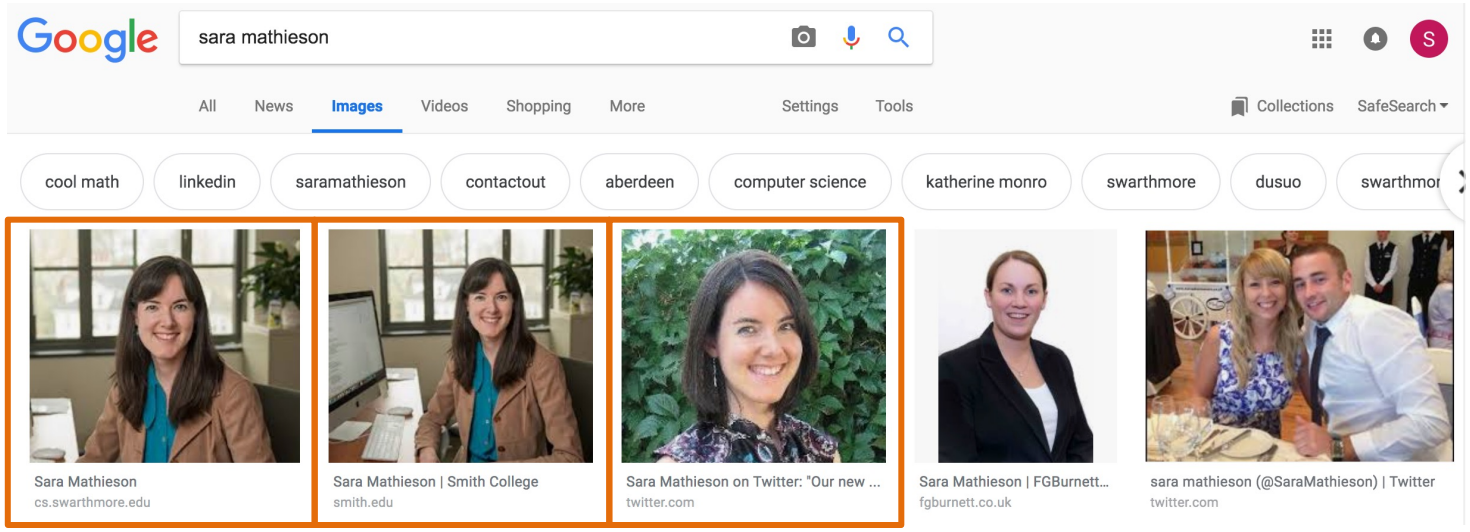
Precision and Recall



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

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

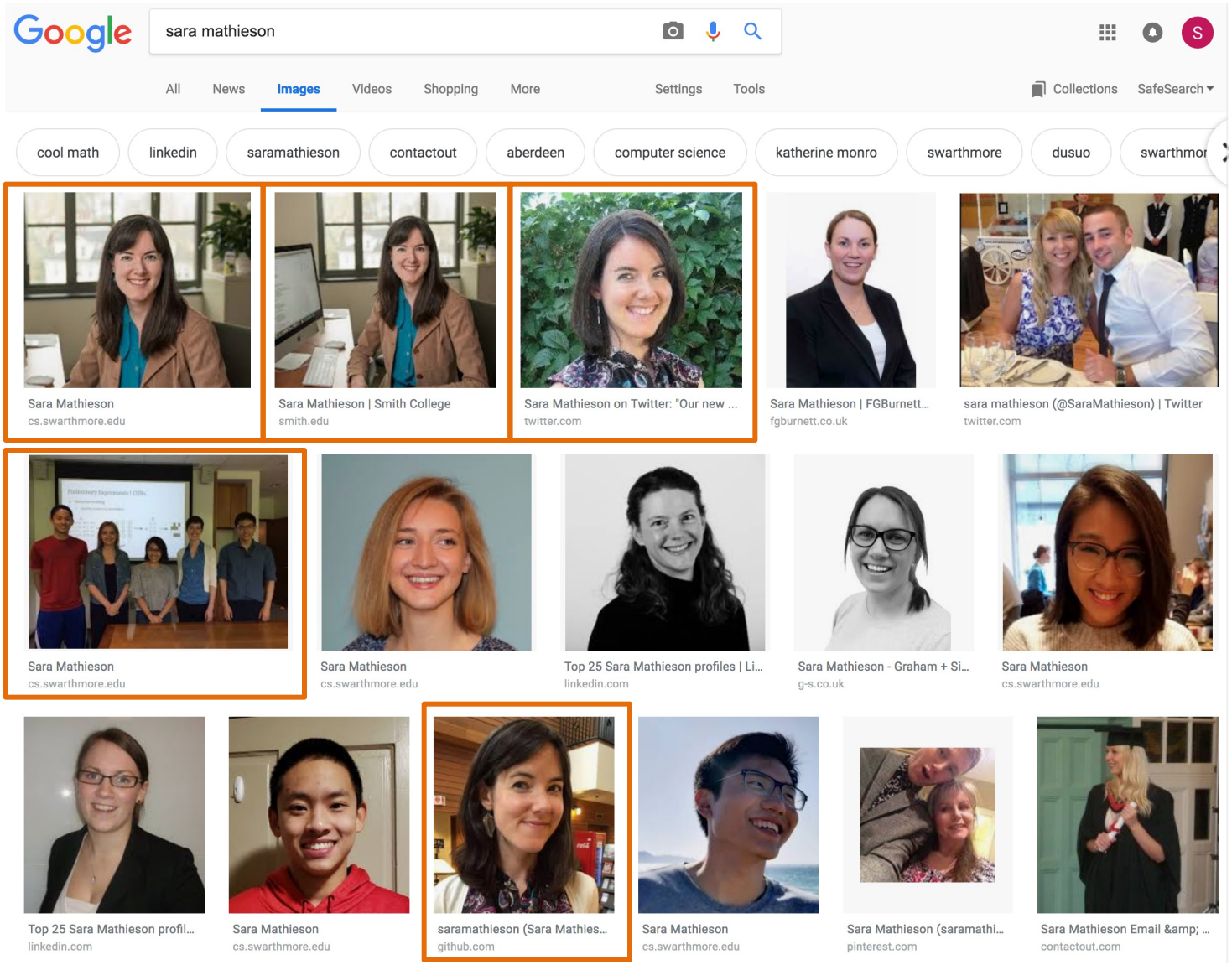
Precision and Recall



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

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

Precision and Recall

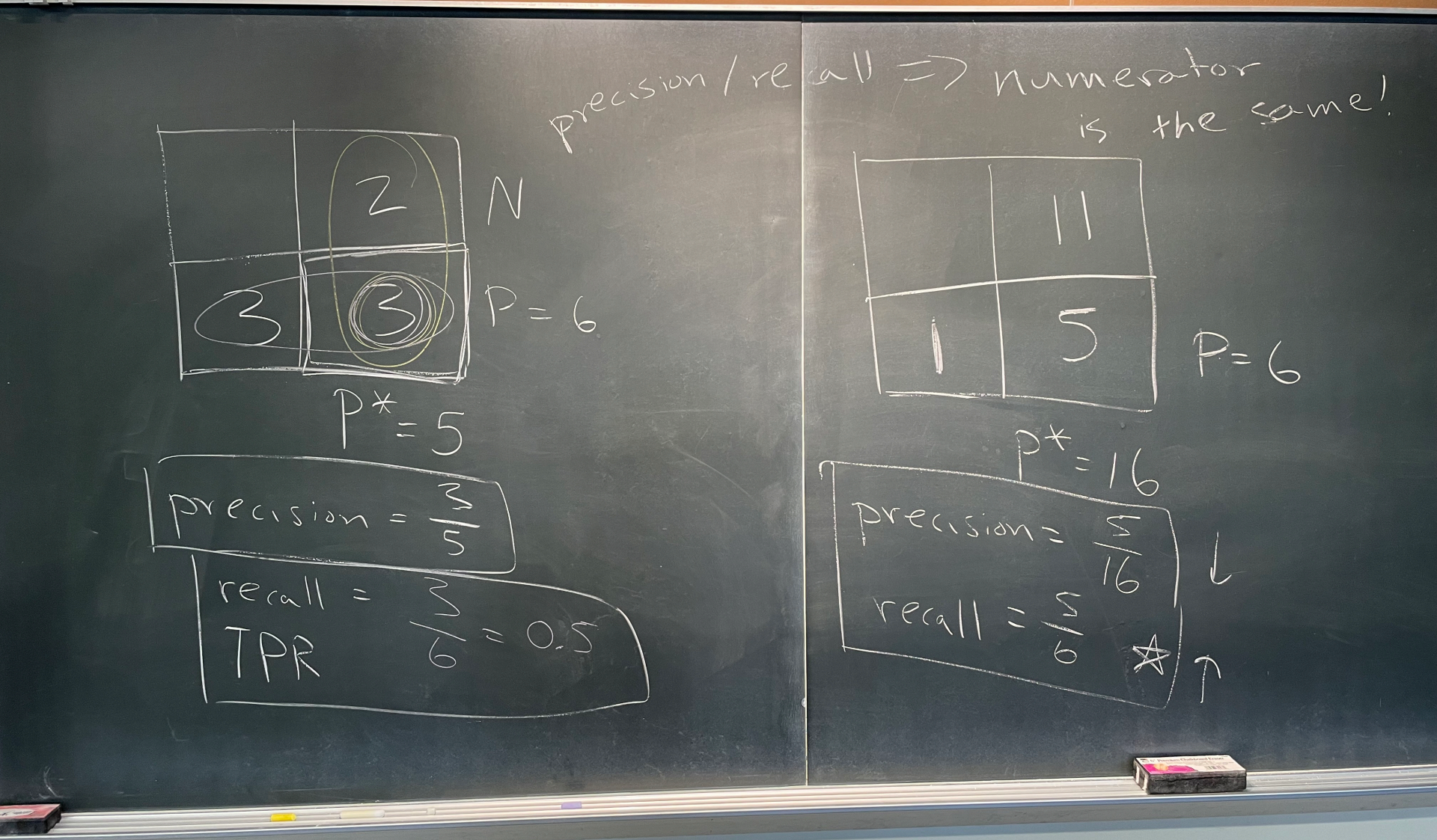


P=6 (number of images that are actually me)

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

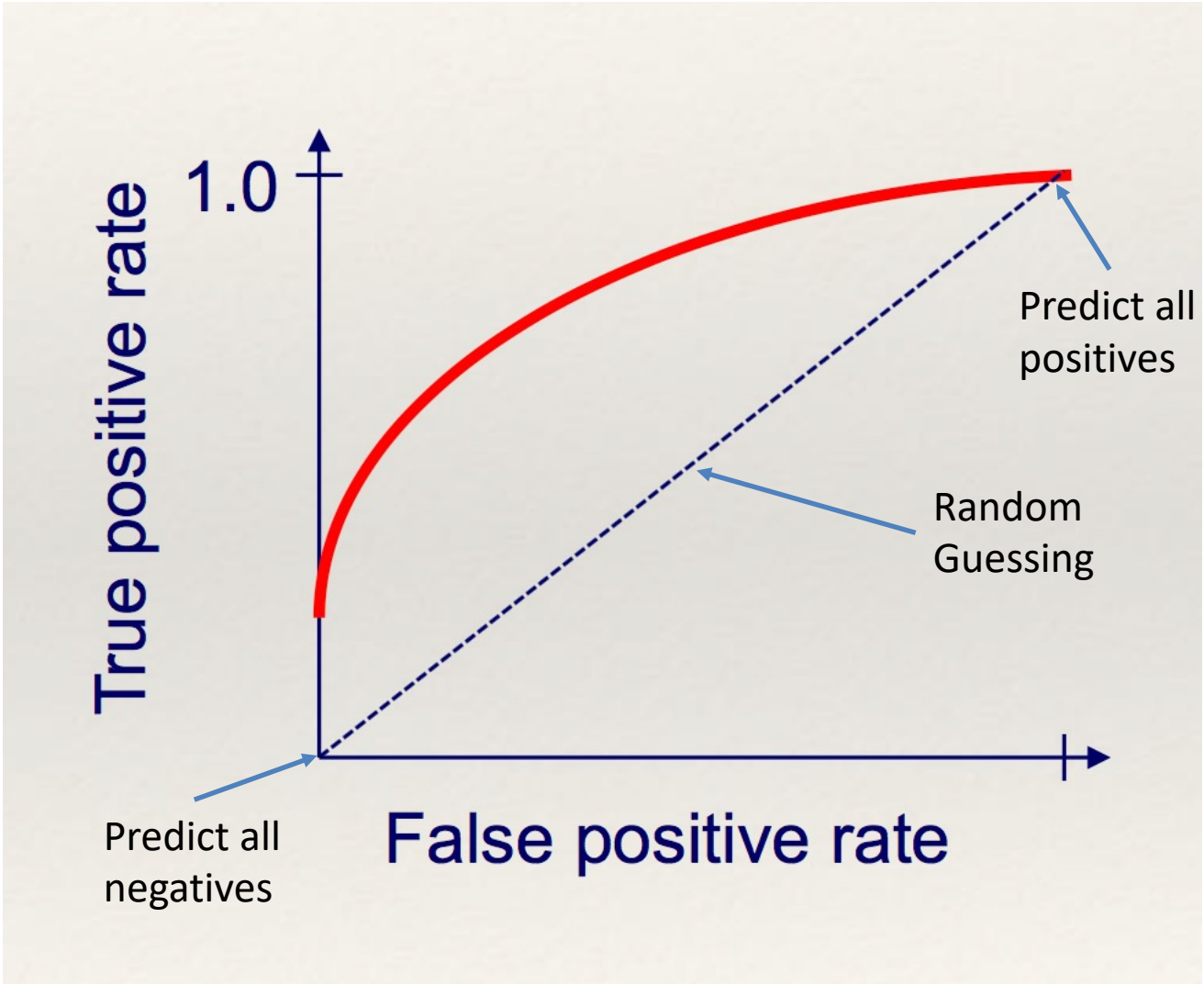
Modified from Ameet Soni

Precision/Recall for google example

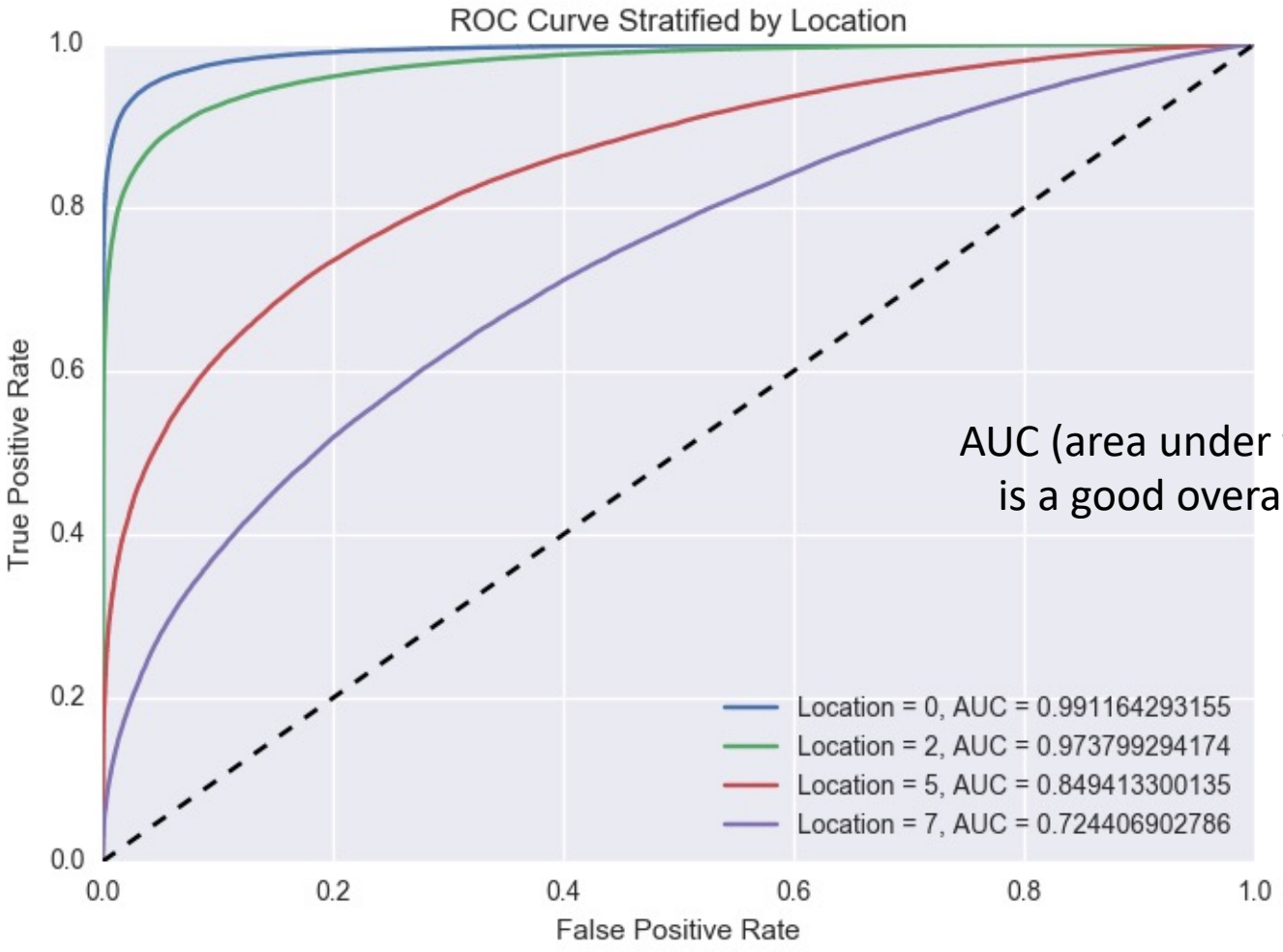


ROC curve (Receiver Operating Characteristic)

More history here! https://en.wikipedia.org/wiki/Receiver_operating_characteristic



ROC curve example: comparing methods



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

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

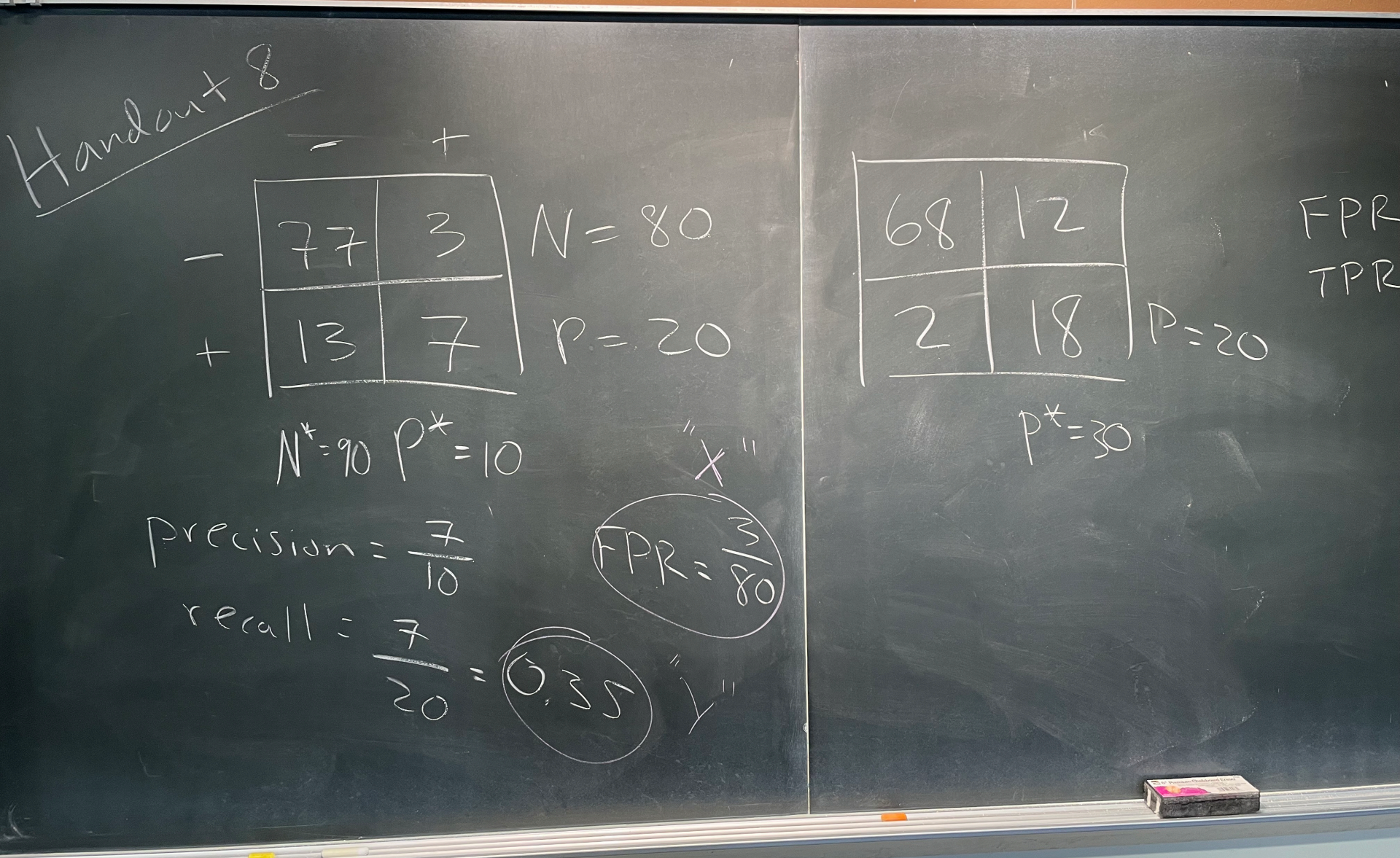
How to get a ROC curve for probabilistic methods?

- Usually we use 0.5 as a threshold for binary classification

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

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

Handout 8


Handout 8
− +
− 77 3
+ 13 7
N = 80
P = 20
N* = 90 P* = 10
precision = 7/10
recall = 7/20 = 0.35
"X"
FPR = 3/80
"Y"
68 12
2 18
P = 20
P* = 30
FPR
TPR

Handout 8