Admin

- Lab 2 grades & feedback posted on Moodle

Outline for today

- Evaluation Metrics — the goal is to see how good the model is for a certain task
  - Confusion matrices
  - Precision and recall
- 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

Different metrics for different problems! Ex: checking credit card transactions, be more risk-averse; spam filters, be less risk-averse

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
- Evaluate the model using testing data

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*	

Error $\frac{FN + FP}{everything}$

Accuracy $\frac{TN + TP}{everything}$ $(1 - error)$

Precision $\frac{TP}{P^*}$

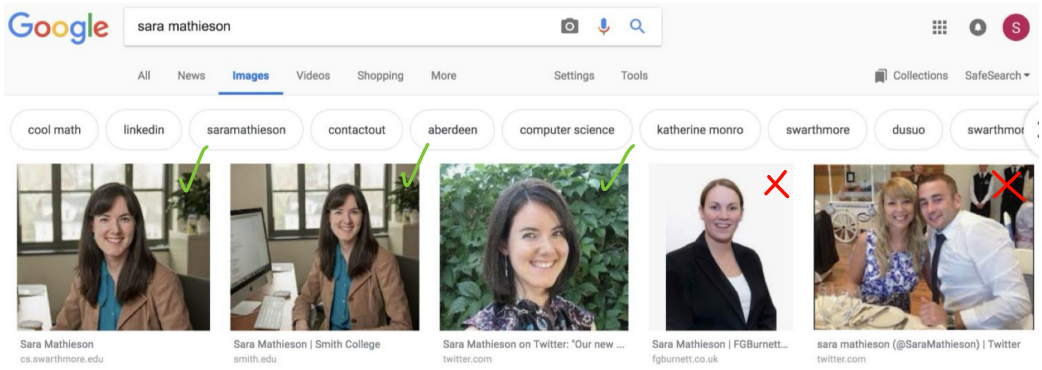
Recall $\frac{TP}{P}$ (true positive rate)

False Positive Rate $\frac{FP}{N}$

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)

$P = 6$ (# of images that are actually Sara)

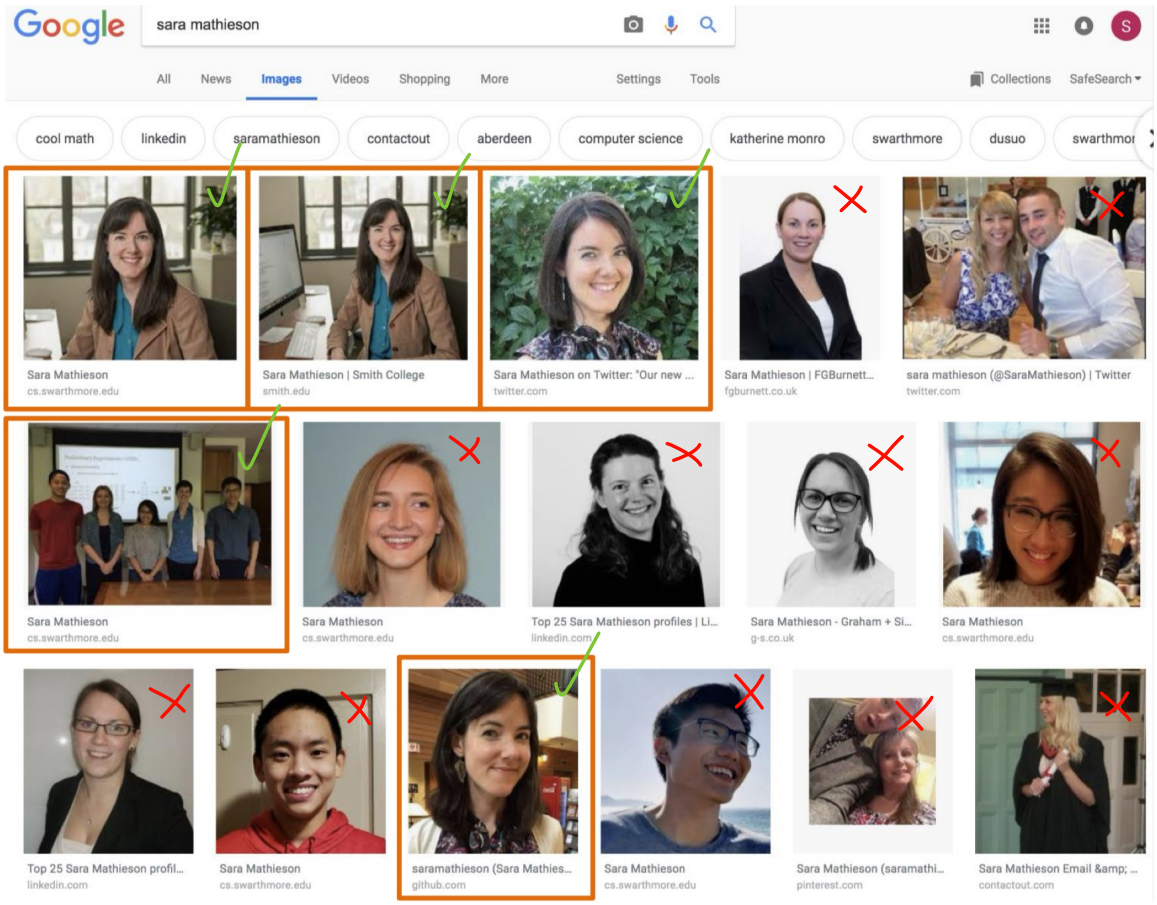


Precision

$$\frac{TP}{P^*} = \frac{3}{5}$$

Recall

$$\frac{TP}{P} = \frac{3}{6}$$



$P = 6$ (# of images that are actually Sara)

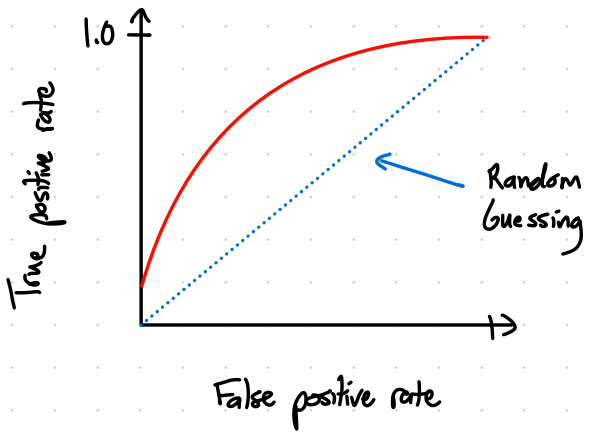
Precision

$$\frac{TP}{P^*} = \frac{5}{16}$$

Recall

$$\frac{TP}{P} = \frac{5}{6}$$

ROC curve (Receiver Operating Characteristic)

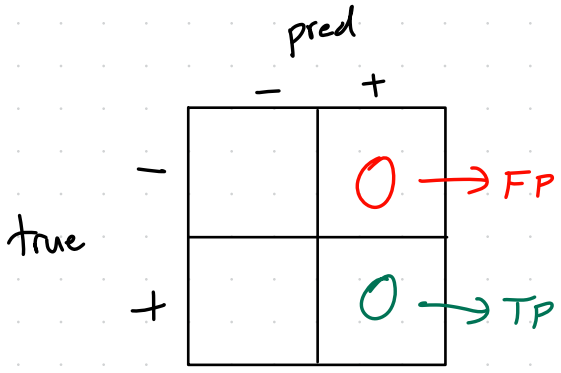


If we classify everything as negative:

All neg

$$TPR = \frac{TP}{N} = \frac{0}{\dots} = 0$$

$$FPR = \frac{FP}{P} = \frac{0}{\dots} = 0$$

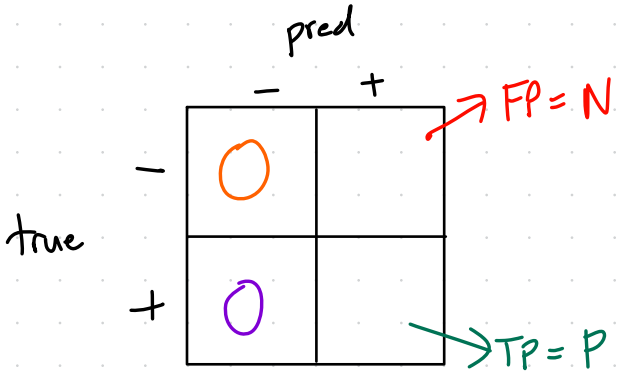


If we classify everything as positive:

All pos

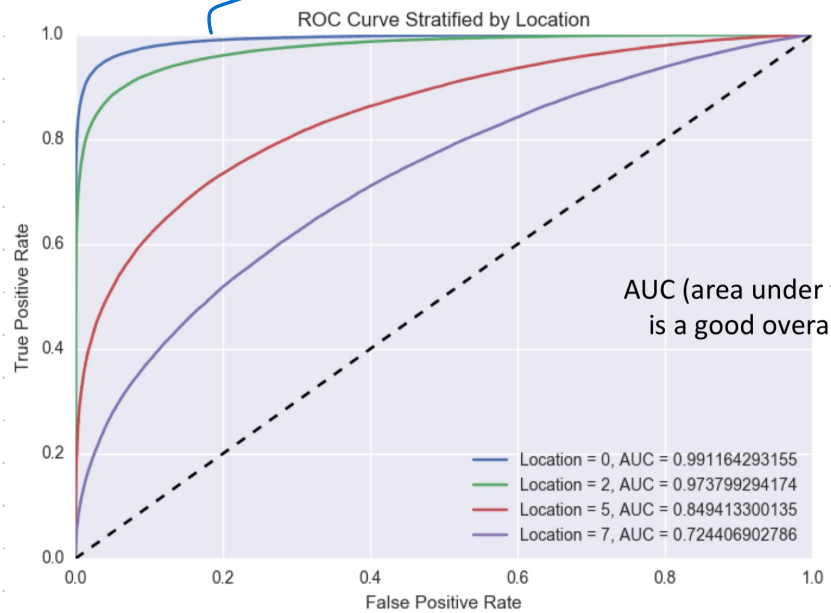
$$TPR = \frac{TP}{N} = 1$$

$$FPR = \frac{FP}{P} = 1$$



ROC curve example: comparing methods

→ top (blue) model is performing the best



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

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, ...)
  - $P(y=1 \mid x) \geq 0.2 \longrightarrow$ classify as 1 (positive)
  - $P(y=1 \mid x) < 0.2 \longrightarrow$ classify as 0 (negative)

HANDOUT 8

①

True \ Pred	−	+	
−	77	3	$n=80$
+	13	7	$p=20$

$n^*=90$ $p^*=10$

②

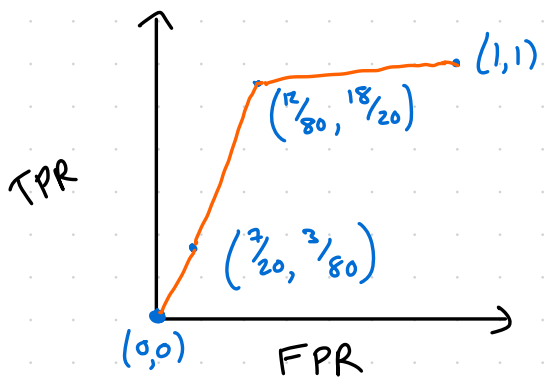
$$\text{precision} = \frac{TP}{p^*} = \frac{7}{10} = 70\%$$

$$\text{recall} = \frac{TP}{P} = \frac{7}{20} = 35\%$$

← recall is TPR! (true positive rate)

$$FPR = \frac{FP}{N} = \frac{3}{80}$$

③



④

True \ Pred	−	+	
−	68	12	$n=80$
+	2	18	$p=20$

$n^*=70$ $p^*=30$

$$TPR = \frac{TP}{P} = \frac{18}{20}$$

$$FPR = \frac{FP}{N} = \frac{12}{80}$$

End of content for Midterm 1 (take-home, details soon)

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, $P(e)$.

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

- 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?

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

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

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