Evaluating classifiers:

Precision & Recall

Emily Fox & Carlos Guestrin

Machine Learning Specialization

University of Washington

©2015-2016 Emily Fox & Carlos Guestrin

Using reviews to promote my restaurant

Goal: increase # guests by 30%

Reviews

Need an automated, "authentic" marketing campaign

Great quotes

"Easily best sushi in Seattle."

Great spokespeople

©2015-2016 Emily Fox & Carlos Guestrin

How do I find sentences with positive sentiment?

All reviews
for my restaurant

©2015-2016 Emily Fox & Carlos Guestrin

Intelligent restaurant review system

All reviews for restaurant

Break all reviews into sentences

The seaweed salad was just OK, vegetable salad was just ordinary.

I like the interior decoration and the blackboard menu on the wall.

All the sushi was delicious.

My wife tried their ramen and it was pretty forgettable.

The sushi was amazing, and the rice is just outstanding.

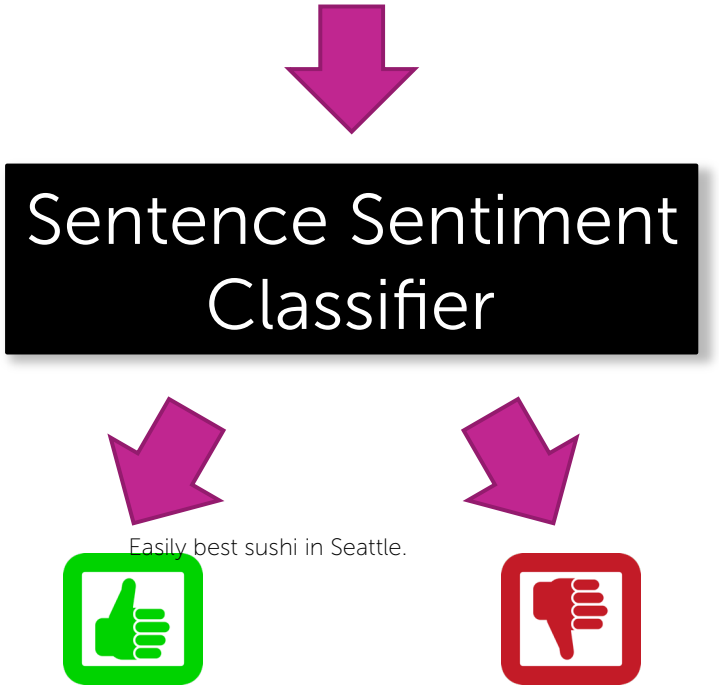
The service is somewhat hectic.

Easily best sushi in Seattle.

©2015-2016 Emily Fox & Carlos Guestrin

Sentiment classifier

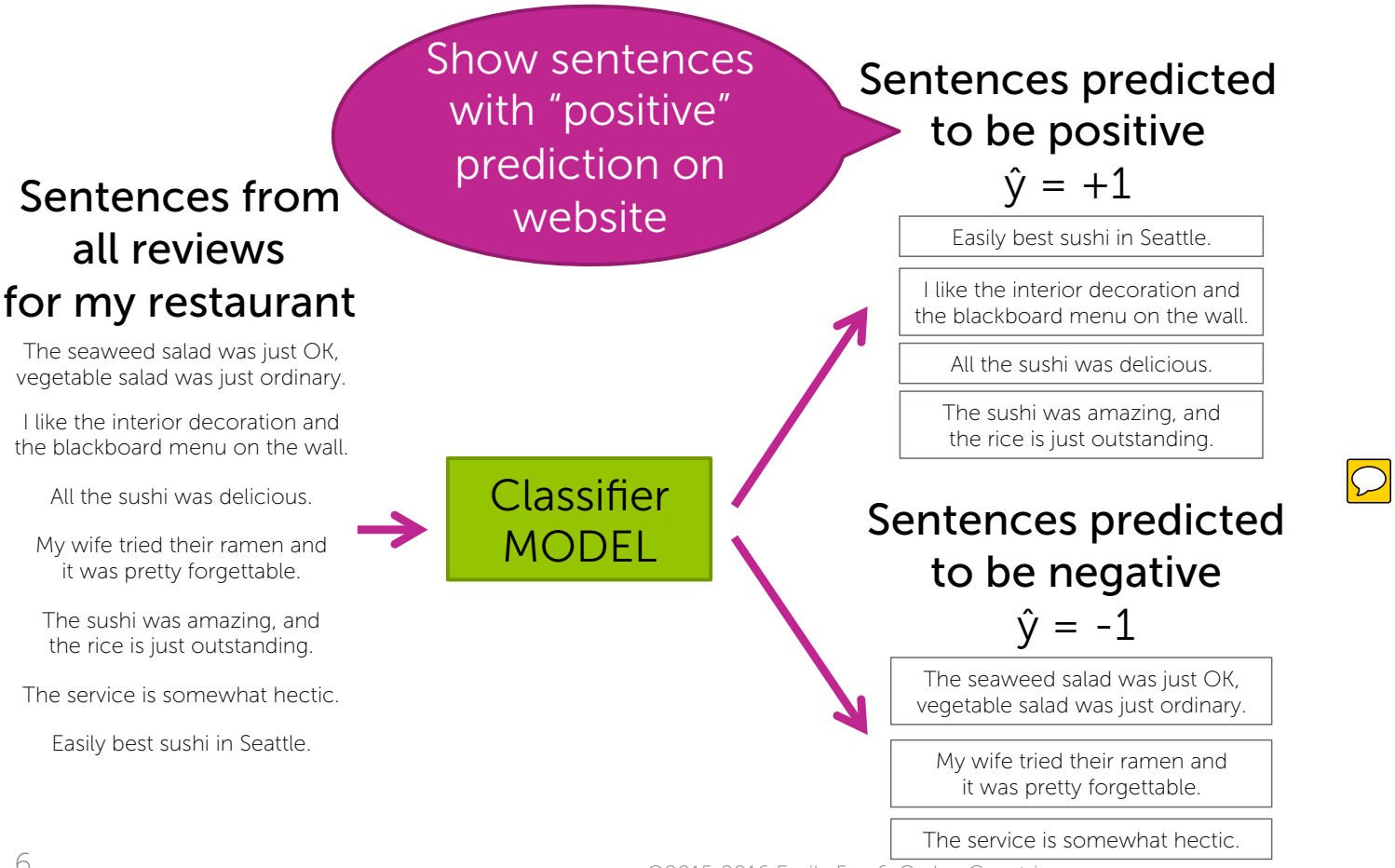
Input $\mathbf{x}_i$: Easily best sushi in Seattle.



Output: $\hat{y}_i$
Predicted sentiment

©2015-2016 Emily Fox & Carlos Guestrin

Use the sentiment classifier model!



©2015-2016 Emily Fox & Carlos Guestrin

What does it mean for a classifier to be good?

©2015-2016 Emily Fox & Carlos Guestrin

Previously, we asked the question:

"What is good accuracy?"

©2015-2016 Emily Fox & Carlos Guestrin

We explored accuracy of random classifier as baseline

- For binary classification:
  - Half the time, you'll get it right! (on average)
    - ➔ classification error = 0.5
- For k classes, error= 1 - 1/k
  - error = 0.666 for 3 classes, 0.75 for 4 classes,…

At the very, very, very least,
you should healthily beat random…
Otherwise, it's (usually) pointless…

©2015-2016 Emily Fox & Carlos Guestrin

We explored the pitfalls of imbalanced problems:

Is 90% accuracy good? Depends …

90% of sentences are negative!

90% accuracy by predicting every sentence is negative!!!

Amazing "performance" but not useful for me right now!

©2015-2016 Emily Fox & Carlos Guestrin

Automated marketing campaign cares about something else…

Website shows 10 sentences from recent reviews

PRECISION

Did I (mistakenly) show a negative sentence???

RECALL

Did I not show a (great) positive sentence???

Accuracy doesn't capture these issues well…

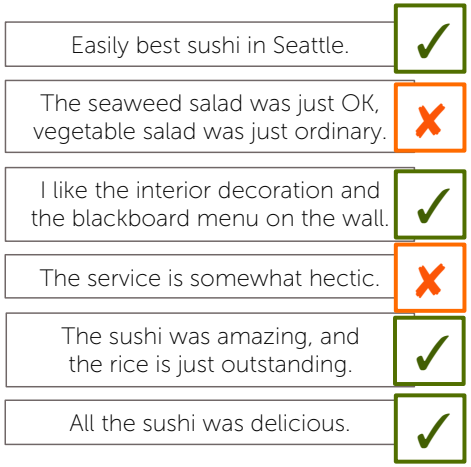
©2015-2016 Emily Fox & Carlos Guestrin

Precision:
Fraction of positive predictions that are actually positive

©2015-2016 Emily Fox & Carlos Guestrin

What fraction of the positive predictions are correct?

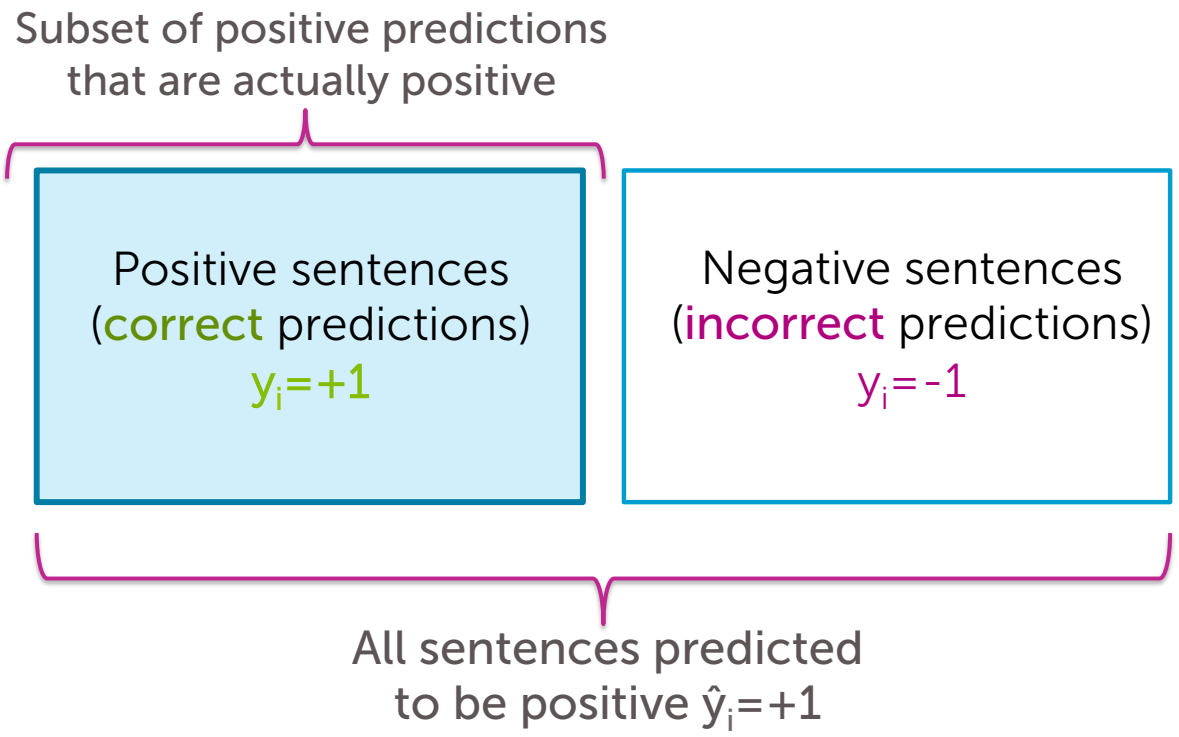
Sentences predicted to be positive: $\hat{y}_i = +1$



Only 4 out of 6 sentences predicted to be positive are actually positive

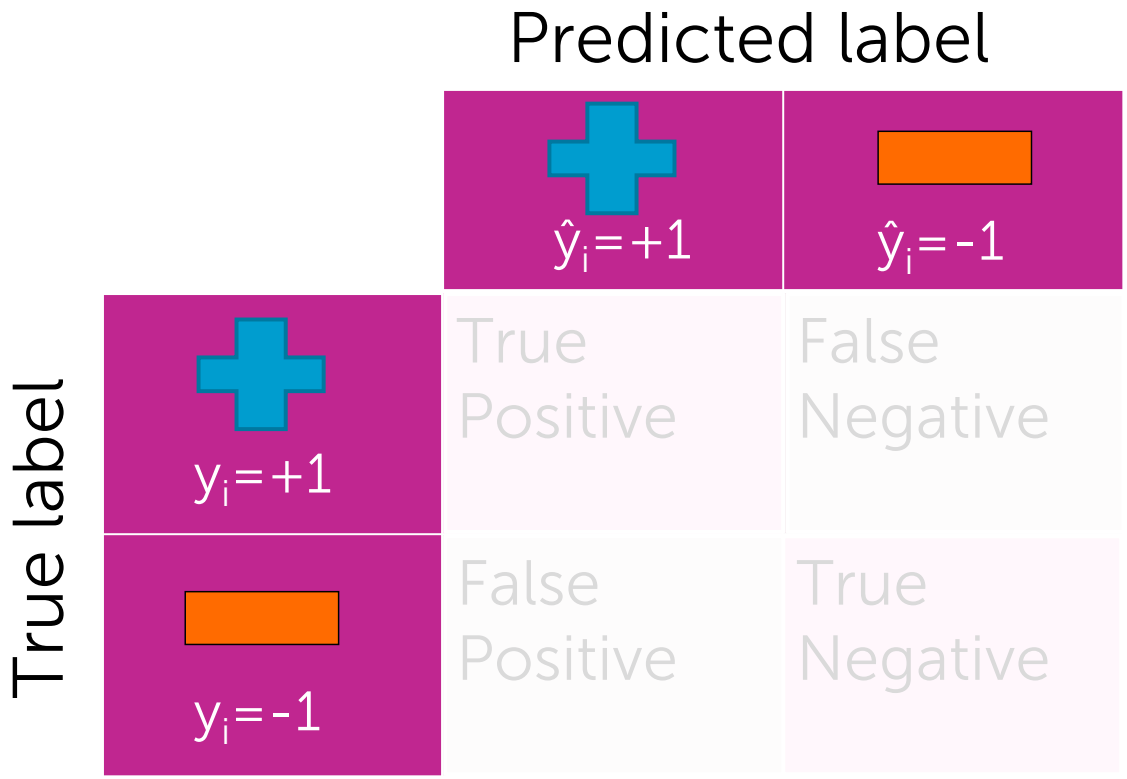
©2015-2016 Emily Fox & Carlos Guestrin

Precision: Fraction of positive predictions that are actually positive



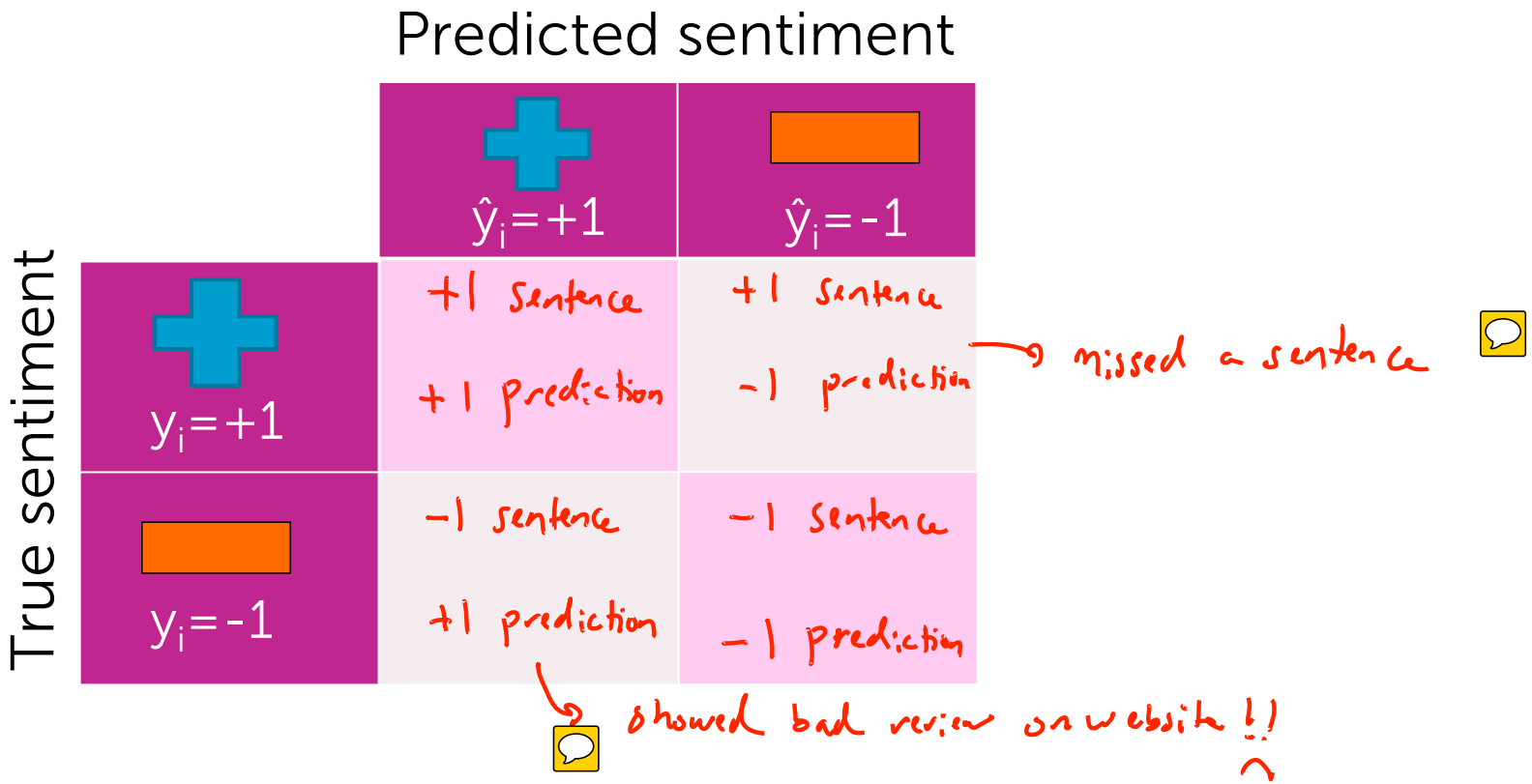
©2015-2016 Emily Fox & Carlos Guestrin

Types of error: *Review*



©2015-2016 Emily Fox & Carlos Guestrin

Confusion matrix for sentiment analysis



©2015-2016 Emily Fox & Carlos Guestrin

Precision - Formula

- Fraction of positive predictions that are correct

$$\text{precision} = \frac{\text{\# true positives}}{\text{\# true positives + \# false positives}}$$

  - Best possible value : 1.0
  - Worst possible value : 0.0

©2015-2016 Emily Fox & Carlos Guestrin

Example: Calculating precision

Sentences predicted to be positive: $\hat{y}_i = +1$

4 correct

2 mistakes

$$\text{precision} = \frac{4}{4+2} = \frac{2}{3}$$

©2015-2016 Emily Fox & Carlos Guestrin

Why precision is important

Shown on website

Sentences predicted to be positive: $\hat{y}_i = +1$

2 negative sentences shown to potential customers… ☹

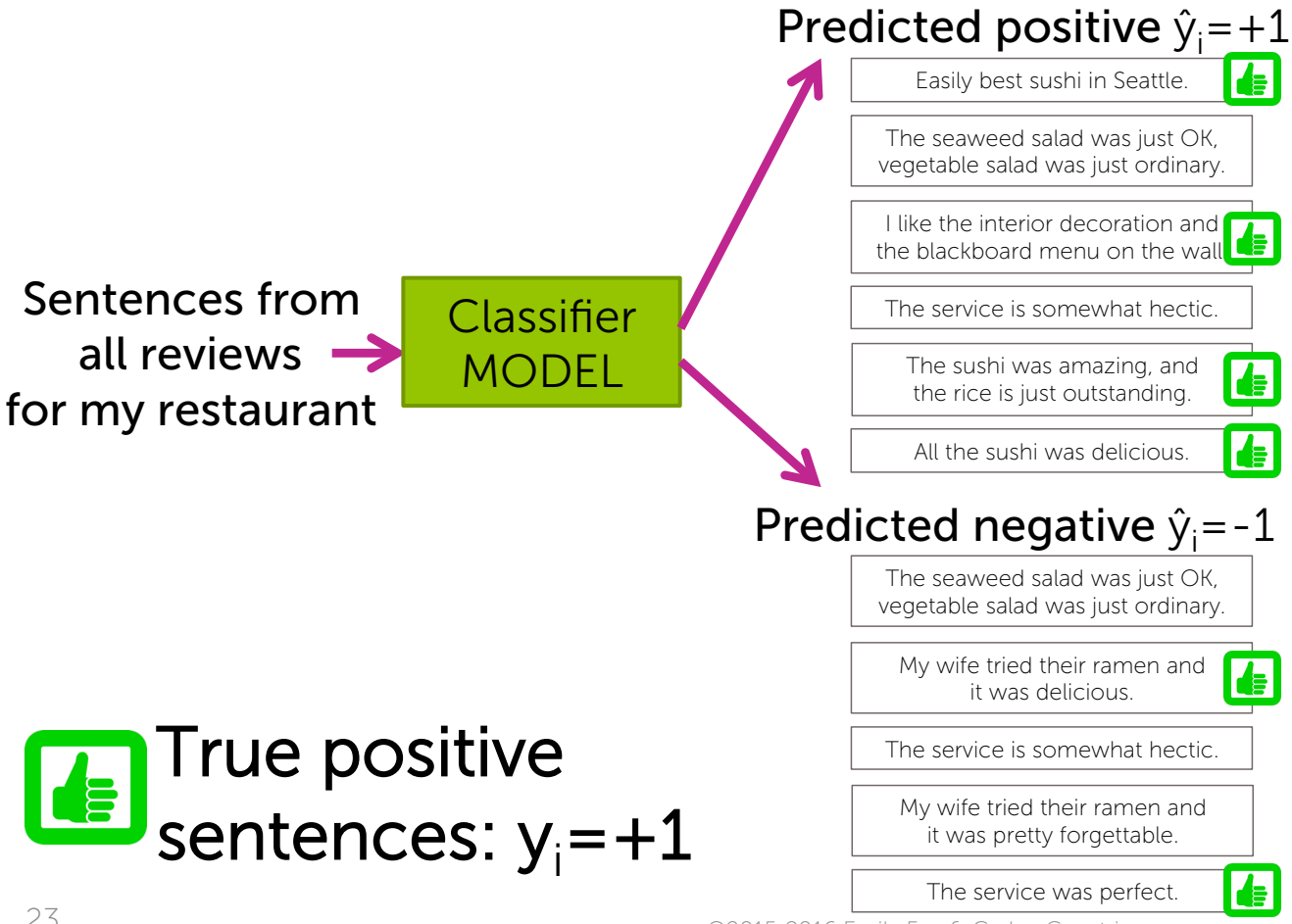
High precision means positive predictions actually likely to be positive!

©2015-2016 Emily Fox & Carlos Guestrin

Recall:
Fraction of positive data predicted to be positive

©2015-2016 Emily Fox & Carlos Guestrin

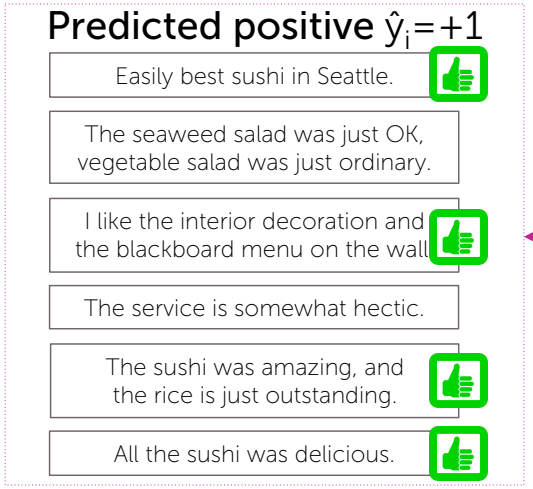
Did I find all the positive sentences?



True positive sentences: $y_i=+1$

©2015-2016 Emily Fox & Carlos Guestrin

What fraction of positive sentences were missed out?



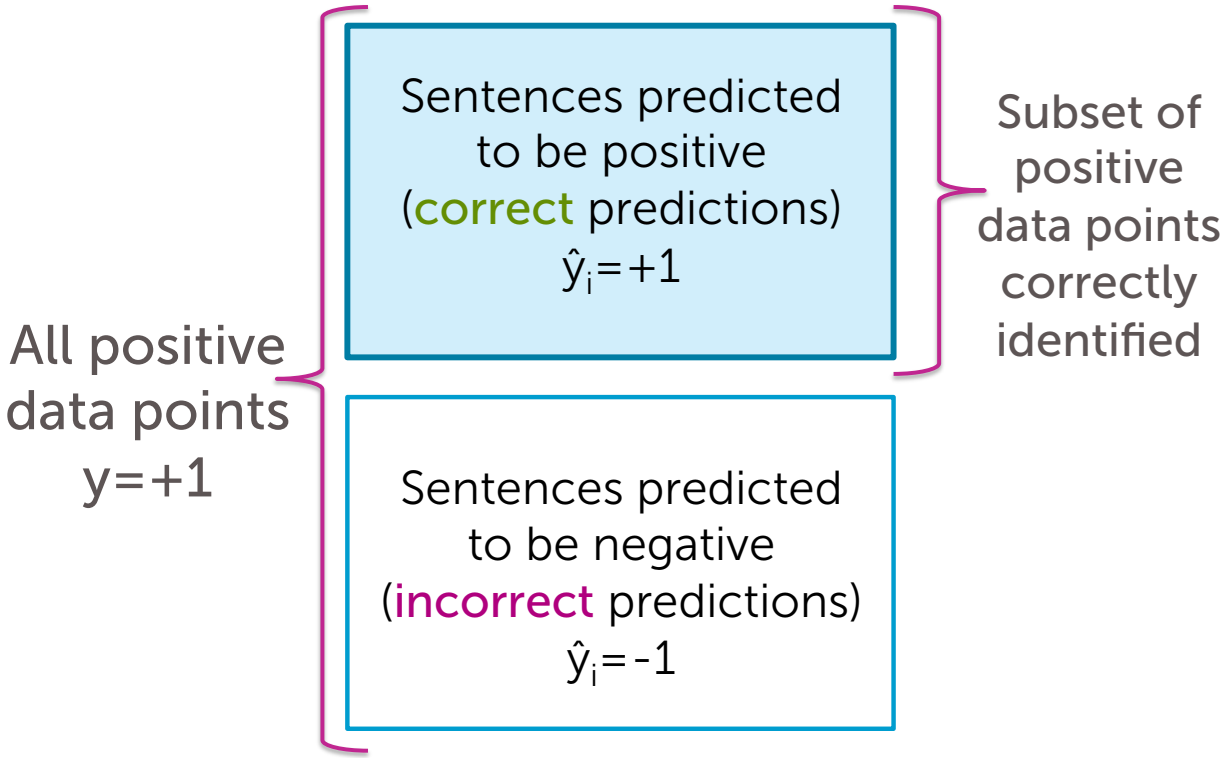
Found 4 positive sentences

Model could not find 2 sentences that were actually positive

Missed 2 positive sentences

©2015-2016 Emily Fox & Carlos Guestrin

Recall: Fraction of positive data predicted to be positive



©2015-2016 Emily Fox & Carlos Guestrin

Recall - Formula

- Fraction of positive data points correctly classified

$$\text{Recall} = \frac{\text{\# true positives}}{\text{\# true positives + \# false negatives}}$$

  - Best possible value : 1.0
  - Worst possible value : 0.0

©2015-2016 Emily Fox & Carlos Guestrin

Why is recall important?

Predicted positive $\hat{y}_i = +1$

- Easily best sushi in Seattle.
- The seaweed salad was just OK, vegetable salad was just ordinary.
- I like the interior decoration and the blackboard menu on the wall
- The service is somewhat hectic.
- The sushi was amazing, and the rice is just outstanding.
- All the sushi was delicious.

Predicted negative $\hat{y}_i = -1$

- The seaweed salad was just OK, vegetable salad was just ordinary.
- My wife tried their ramen and it was delicious.
- The service is somewhat hectic.
- My wife tried their ramen and it was pretty forgettable.
- The service was perfect.

Want to show positive sentences on website

2 positive sentences not shown to potential customers... ☹

High recall means positive data points are very likely to be discovered!

©2015-2016 Emily Fox & Carlos Guestrin

Resumen:

Precision-recall extremes

©2015-2016 Emily Fox & Carlos Guestrin

Optimistic model:

High recall, low precision

Sentences from all reviews for my restaurant

OPTIMISTIC MODEL

Predict almost everything as positive

Predicted positive $\hat{y}_i = +1$

- Easily best sushi in Seattle.
- The seaweed salad was just OK, vegetable salad was just ordinary.
- I like the interior decoration and the blackboard menu on the wall
- The service is somewhat hectic.
- The sushi was amazing, and the rice is just outstanding.
- All the sushi was delicious.
- The seaweed salad was just OK, vegetable salad was just ordinary.
- My wife tried their ramen and it was delicious.
- The service was perfect.
- My wife tried their ramen and it was pretty forgettable.

Predicted negative $\hat{y}_i = -1$

- The service is somewhat hectic.

True positive sentences: $y_i = +1$

©2015-2016 Emily Fox & Carlos Guestrin

Pessimistic model:

High precision, low recall

Sentences from all reviews for my restaurant

PESSIMISTIC MODEL

Predict positive only when very sure

Predicted positive $\hat{y}_i = +1$

- Easily best sushi in Seattle.
- The sushi was amazing, and the rice is just outstanding.

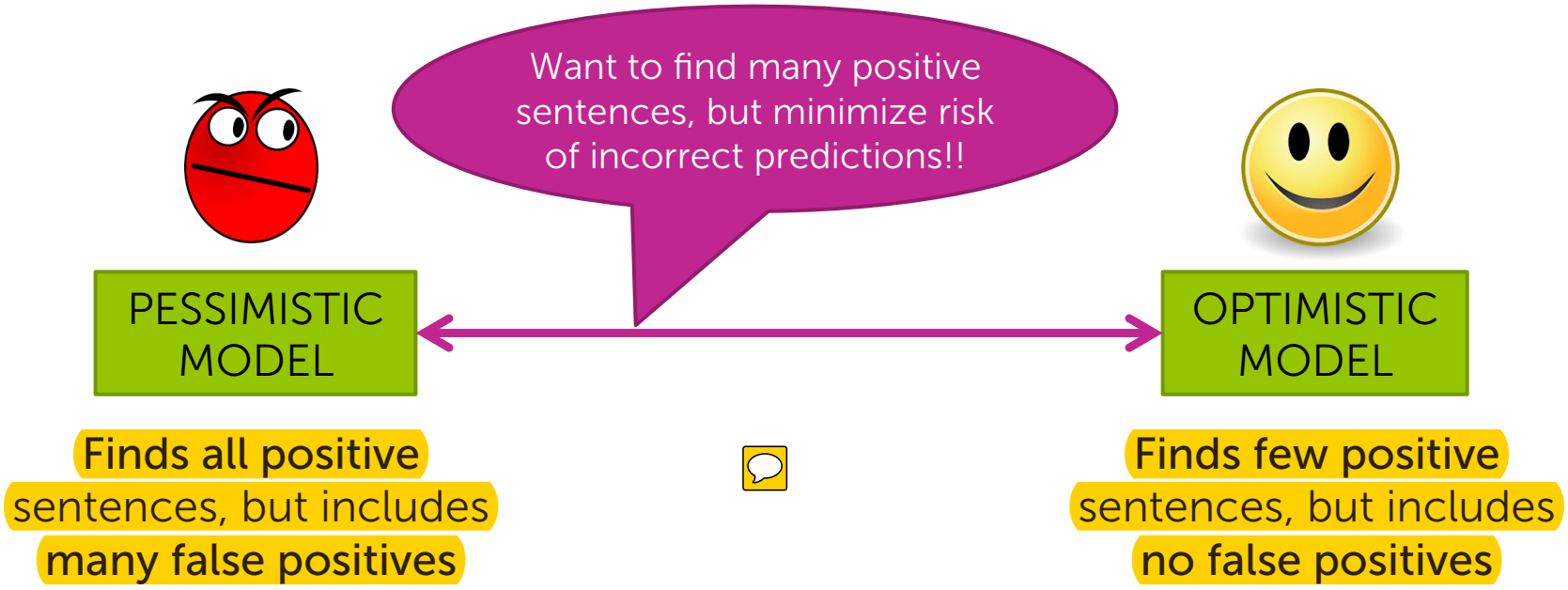
Predicted negative $\hat{y}_i = -1$

- I like the interior decoration and the blackboard menu on the wall.
- The service is somewhat hectic.
- The seaweed salad was just OK, vegetable salad was just ordinary.
- All the sushi was delicious.
- The seaweed salad was just OK, vegetable salad was just ordinary.
- My wife tried their ramen and it was delicious.
- The service was perfect.
- My wife tried their ramen and it was pretty forgettable.
- The service is somewhat hectic.

True positive sentences: $y_i = +1$

©2015-2016 Emily Fox & Carlos Guestrin

Balancing precision & recall



©2015-2016 Emily Fox & Carlos Guestrin

Tradeoff precision and recall

©2015-2016 Emily Fox & Carlos Guestrin

Can we tradeoff precision & recall?

Low precision, high recall

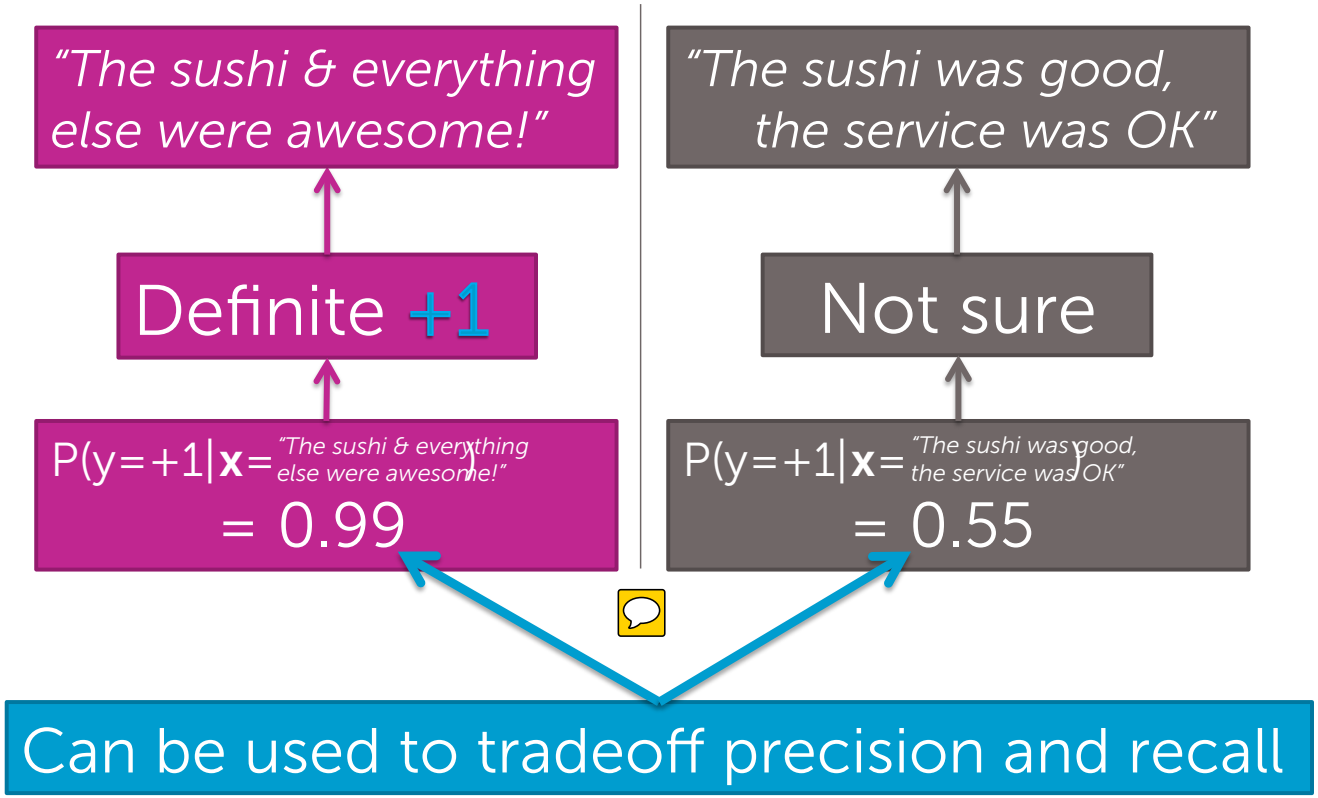
High precision, low recall

Optimistic Model
Predict almost everything as positive

Pessimistic Model
Predict positive only when very sure

©2015-2016 Emily Fox & Carlos Guestrin

How confident is your prediction?



©2015-2016 Emily Fox & Carlos Guestrin

Basic classifier

Sentence from review

Input: $\mathbf{x}_i$

Predict most likely class

If $\hat{P}(y=+1|x_i) > 0.5$:

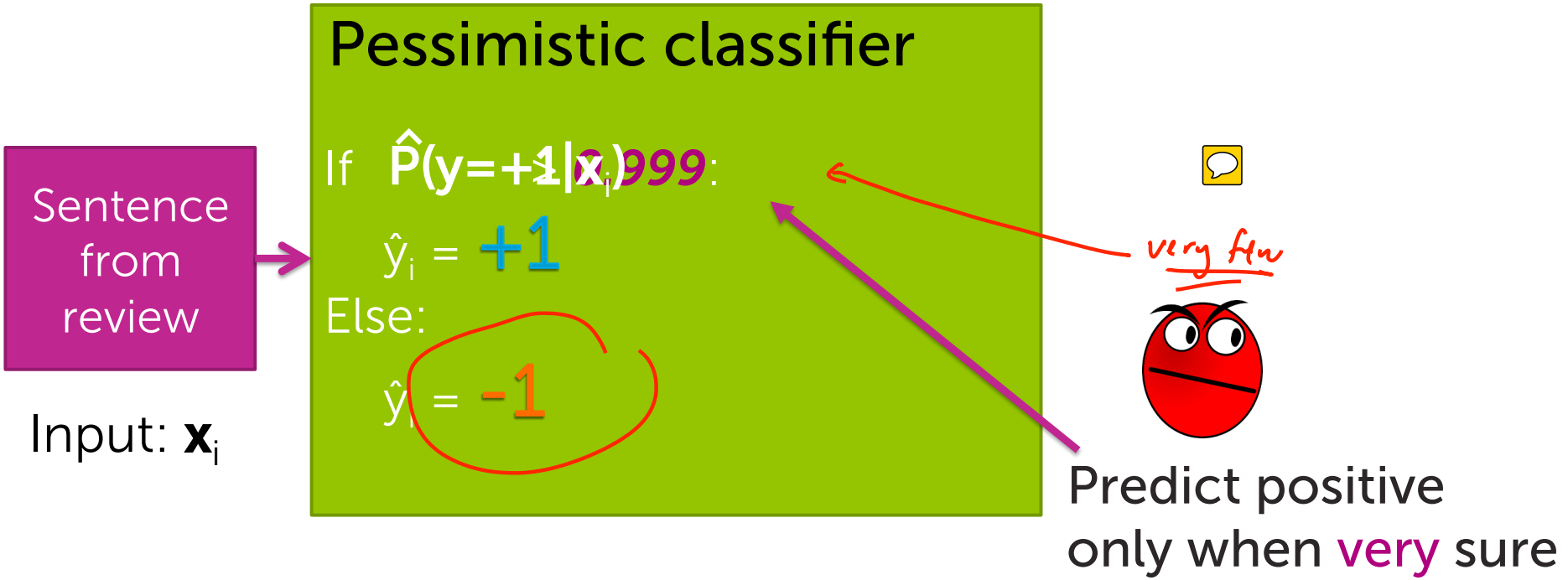
$\hat{y}_i = +1$

Else:

$\hat{y}_i = -1$

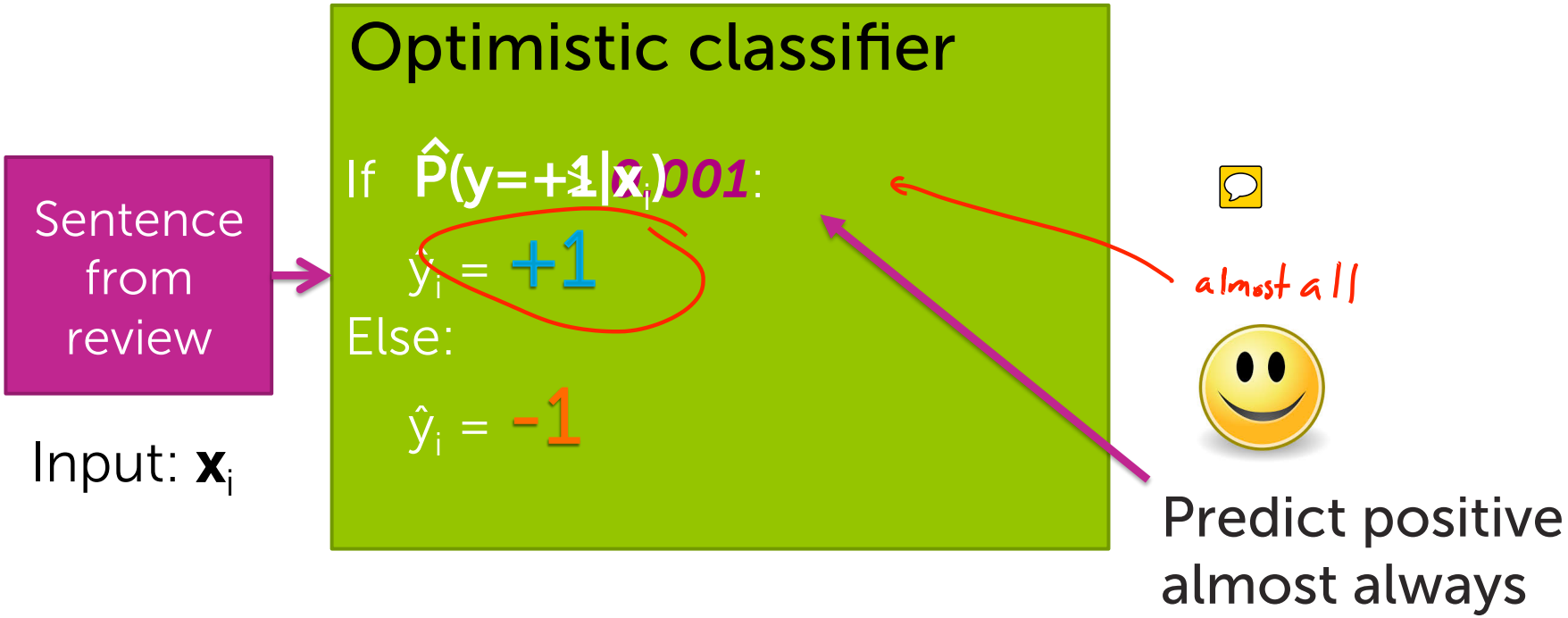
©2015-2016 Emily Fox & Carlos Guestrin

Pessimistic: High precision, low recall



©2015-2016 Emily Fox & Carlos Guestrin

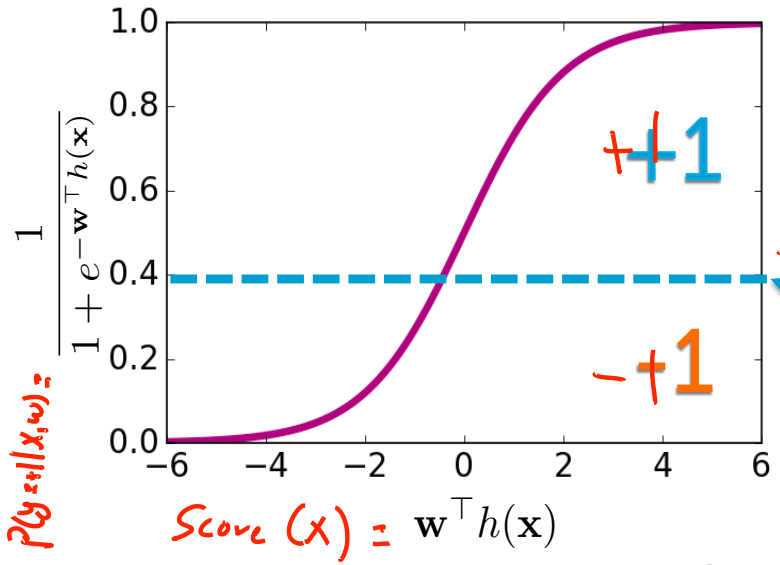
Optimistic: Low precision, high recall



©2015-2016 Emily Fox & Carlos Guestrin

Prediction probability threshold

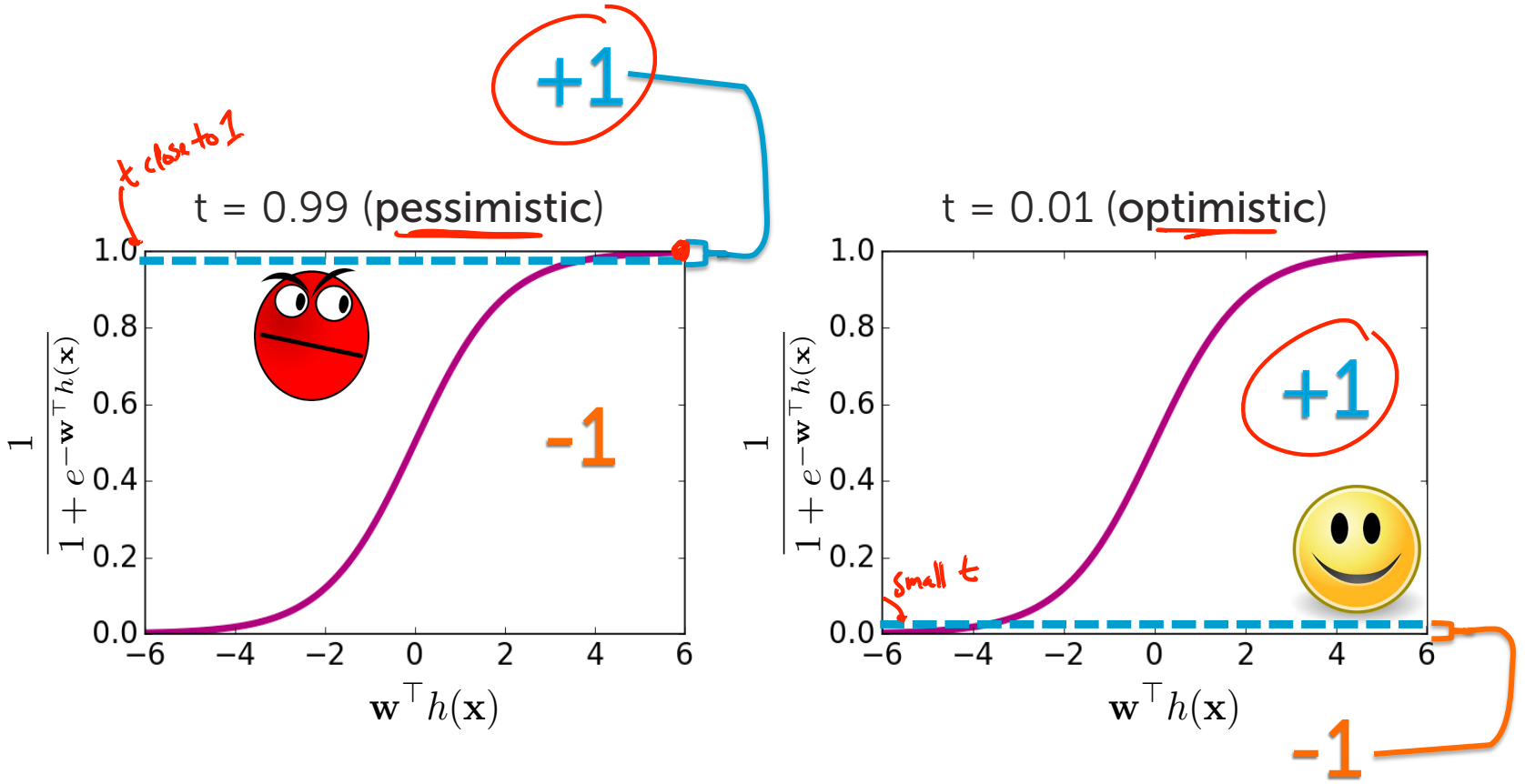
Probability t above which model predicts true



Set $\hat{y} = +1$ if $\hat{P}(y|x) >= t$

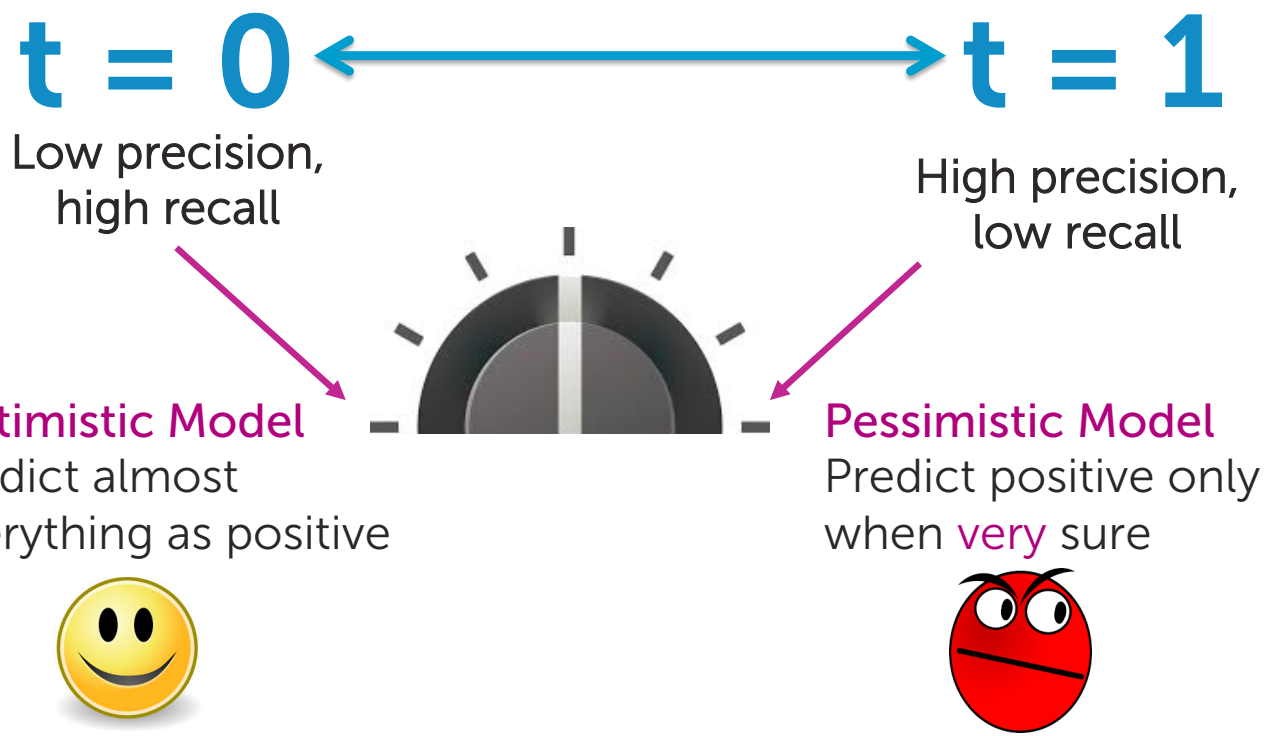
©2015-2016 Emily Fox & Carlos Guestrin

Example threshold values



©2015-2016 Emily Fox & Carlos Guestrin

Tradeoff precision & recall with threshold

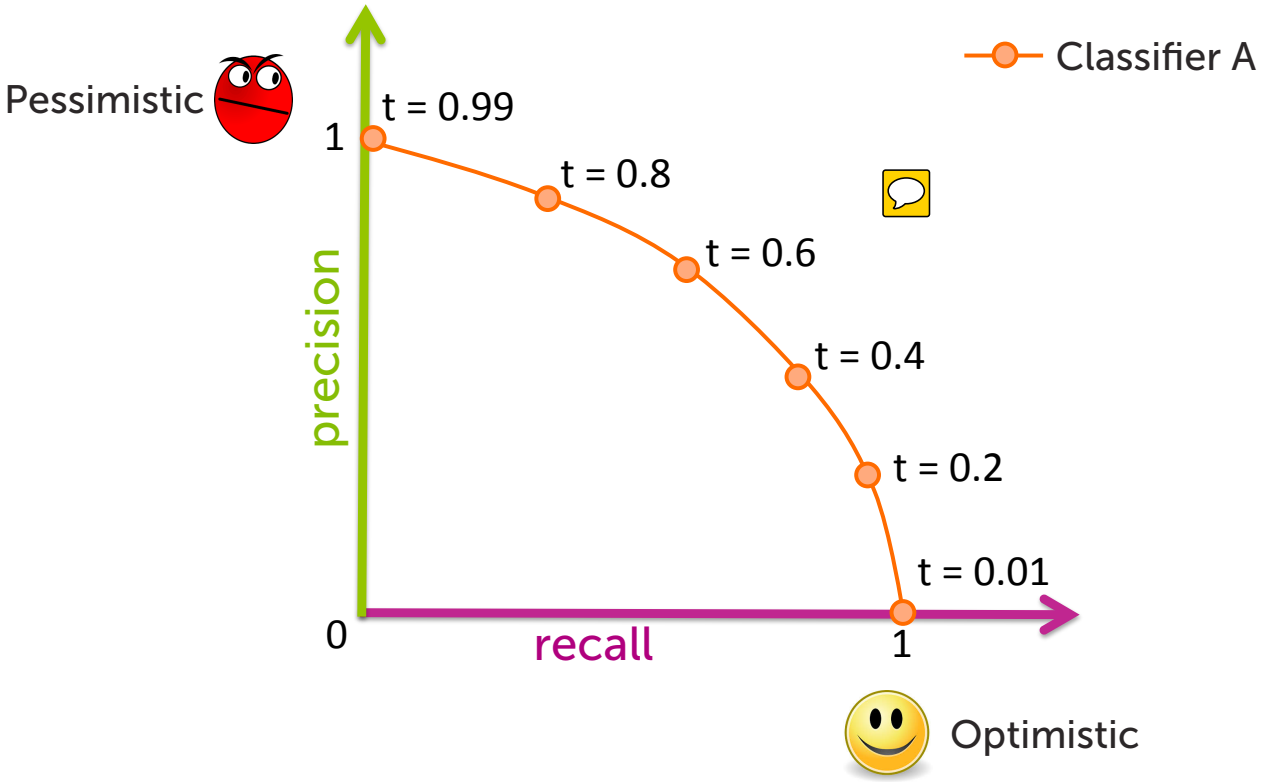


©2015-2016 Emily Fox & Carlos Guestrin

Precision-recall curve

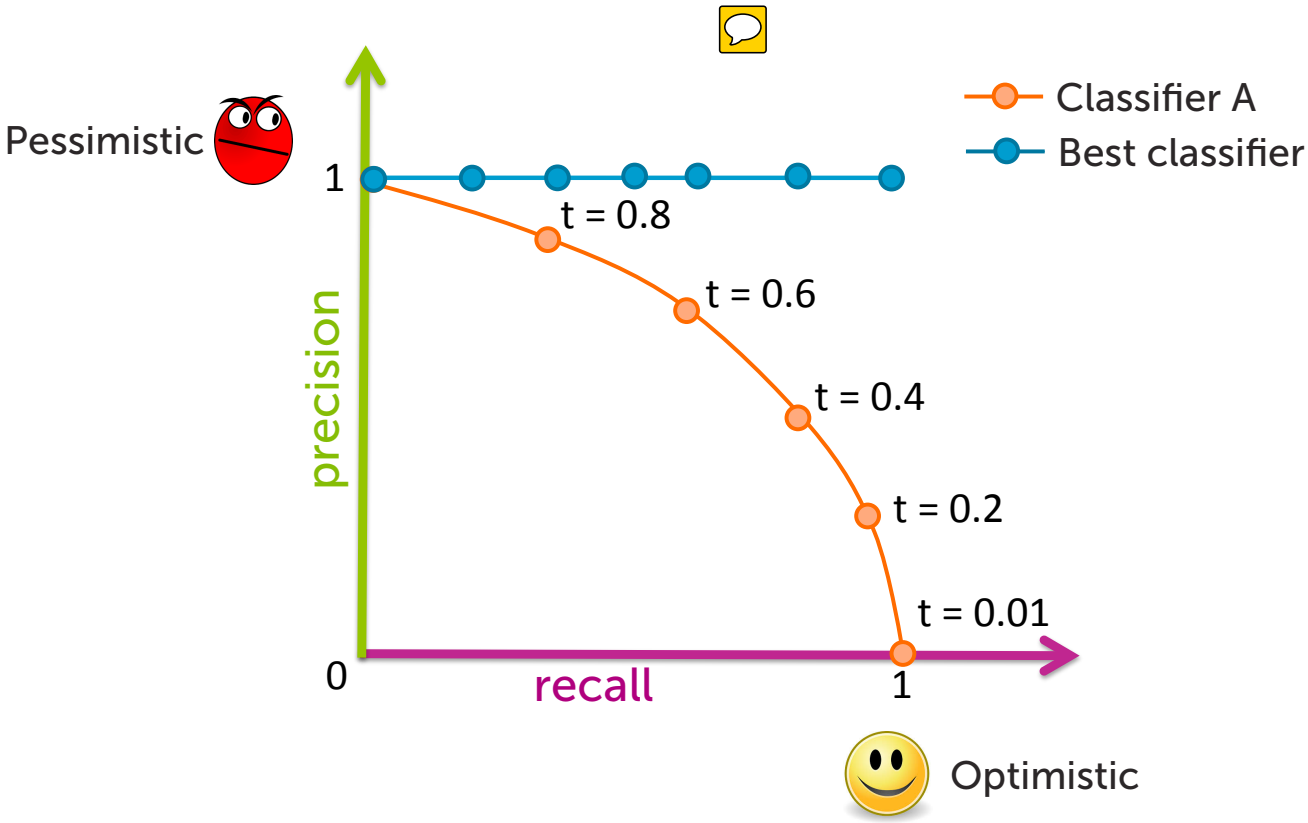
©2015-2016 Emily Fox & Carlos Guestrin

The precision-recall curve



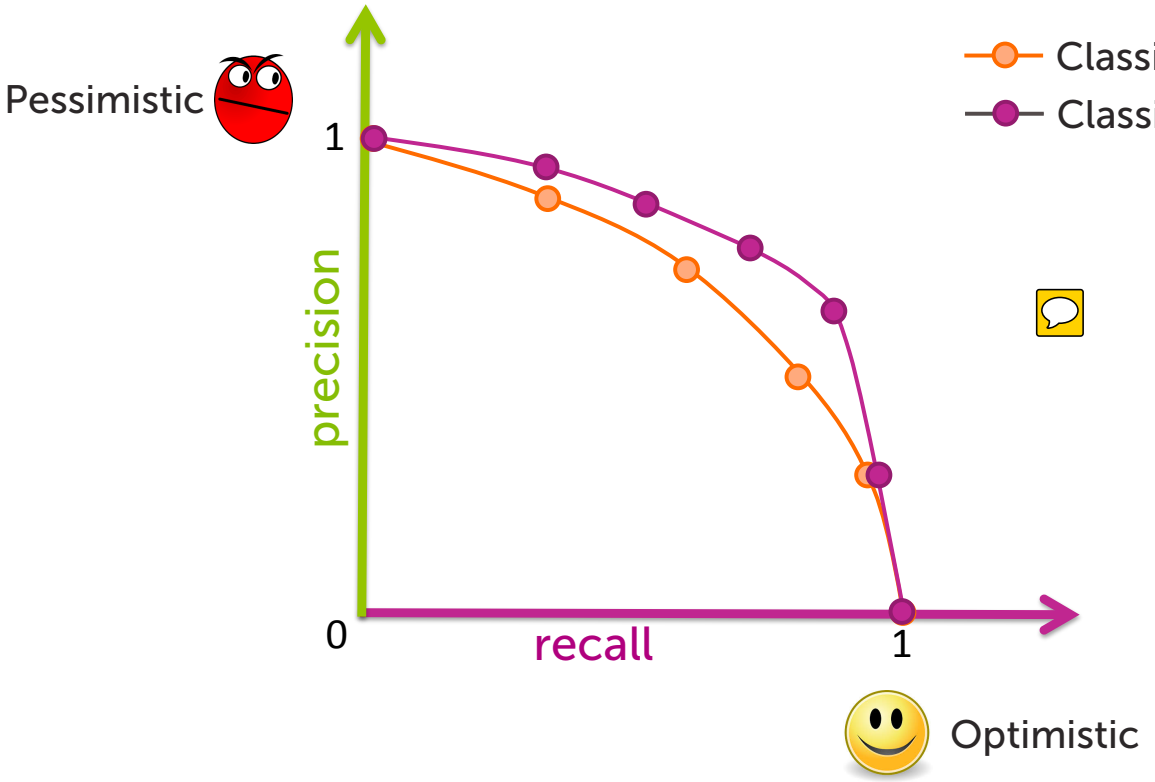
©2015-2016 Emily Fox & Carlos Guestrin

What does the perfect algorithm look like?



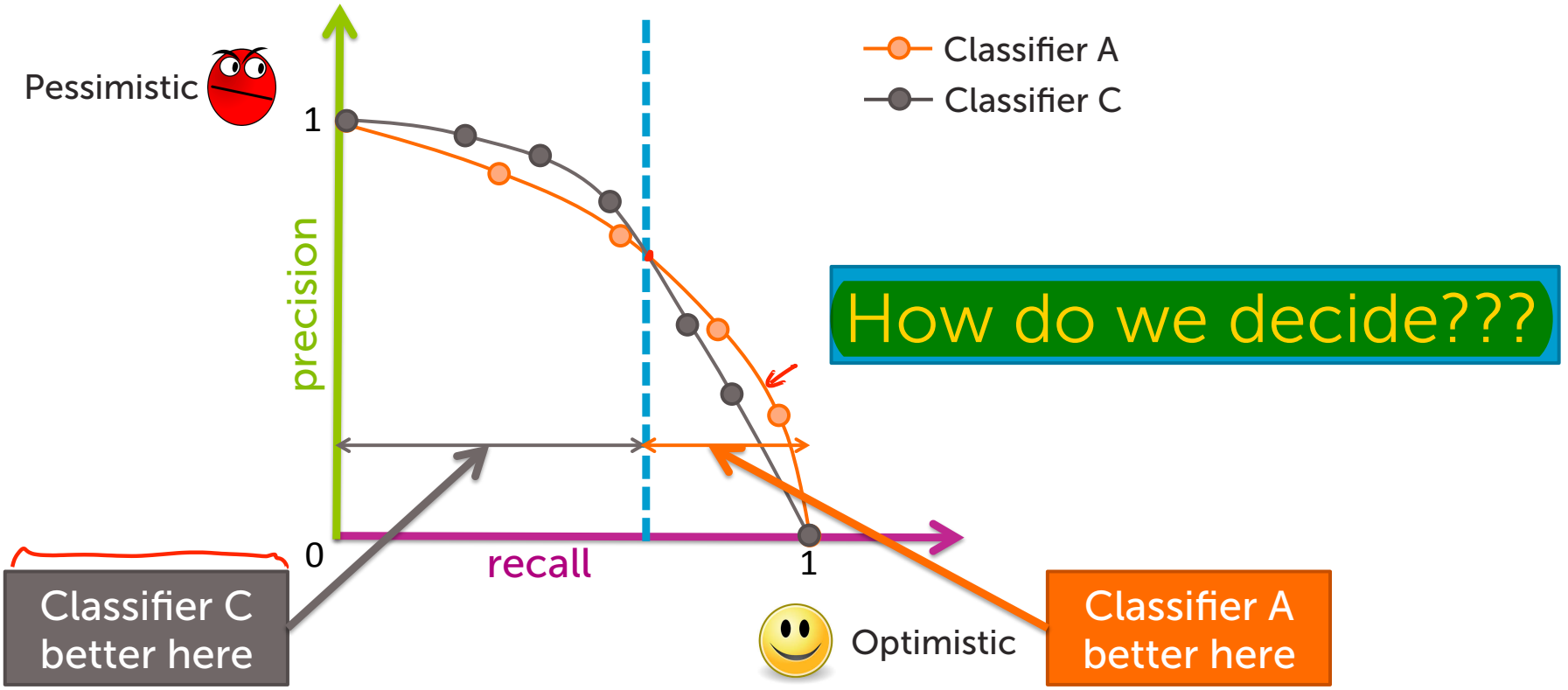
©2015-2016 Emily Fox & Carlos Guestrin

Which classifier is better? A or B?



©2015-2016 Emily Fox & Carlos Guestrin

Which classifier is better? A or C?



©2015-2016 Emily Fox & Carlos Guestrin

Compare algorithms

- Often, reduce precision-recall to single number to compare algorithms
  - F1 measure, area-under-the-curve (AUC),…

Precision at k

Showing k=5 sentences on website

Sentences model most sure are positive

precision at k = 0.8

©2015-2016 Emily Fox & Carlos Guestrin

Summary of precision-recall

©2015-2016 Emily Fox & Carlos Guestrin

What you can do now...

- Classification accuracy/error are not always right metrics
- Precision captures fraction of positive predictions that are correct
- Recall captures fraction of positive data correctly identified by the model
- Trade-off precision & recall by setting probability thresholds
- Plot precision-recall curves.
- Compare models by computing precision at k

©2015-2016 Emily Fox & Carlos Guestrin

Thank you to Dr. Krishna Sridhar

Dr. Krishna Sridhar

Staff Data Scientist, Dato, Inc.

©2015-2016 Emily Fox & Carlos Guestrin