

### Evaluating classifiers:



#### Precision & Recall

Emily Fox & Carlos Guestrin Machine Learning Specialization University of Washington

#### Using reviews to promote my restaurant



# How do I find sentences with positive sentiment?

#### All reviews for my restaurant



TRIADOS

This is probably my flooring inface to set Japonese in Seattle, My Explient and I Codenet right of coaling, Japonese in Seattle, My Explient and I Codenet right of coaling, Japanese import personancel, and the ageleate from Lard 2 special role. I revoid slop the special role, Securate the right and seattle roles in Security in the Seattle Reprise and Seattle roles. Security in the Seattle Reprise and Seattle Reprise Reprise American Company.

This is probably my function plant to set Apparence in Seattle, My Oxylmid and I crisided rigin of scallop, Juparence snappor (seasors), and this appointed in the same of the special right. Set you have been right and seathern cost is where this plant set (so the scale see only in an exercised by other Vietne was artising it is more cheen yet and the season of the scale costs. The total, as recommended by other Vietne was artising it is more cheen yet of the support of the season of the scale of the season of the scale of the season of the season of the scale of the season of the season of the scale of the season of the season of the scale of the scal

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

7.21/2015

This is probledly my flavorise place to set Japanese in the land and 2 man for the first and 2 mass from the first and the first place section fits. Location the first place section for the first place for the first pl

This is probably my feworths place to eat Japanese in bestite. My boyfered and Lordered right of society, gapgeree mayope beaution, and the appeared to during the proposal refs. Livedd sight the special risk, secause the gap is related to the special refs. In this, and the special refs. Livedd sight the special risk, secause the special refs. Livedd sight the special risk, secause the special refs. Livedd sight the special risk, secause the special refs. Livedd sight the special risk, secause the special refs and secarcity of the special refs. The trivial refs. The special refs. Th

Training to the state of the st

This is probably my flavorite place to est Japanese in Seattle, My cylindria and crossed nigit of scalary, Japanese support lessoration, My cylindria and crossed nigit of scalary, Japanese support dessoration, and the appeals into Land skip the special risk. Declares the right and sophism routs is where the place codes. The for some common of the other seasons analogit, it is not change and the sounded by often risked seasons analogit, it is not change and the sounded by often risked seasons analogit, it is not change and the sounded by often risked seasons analogit, it is not change and the sounded seasons and the seasons are seasons are seasons are seasons are seasons and the seasons are seasons

7/21/2015
This is protebly my flowriter place to ext. Japanese in Sectite. My boyfers and clienter nigit of exalize, Japanese snapon beaconst, and the agedant found and 2 special must. Invade sight the special must. Decourse the nigit and cashimir curs is where this place excels. This totu, as occurrenced by other Velories was analog. It is more clienter of the example of the section of the section of the section of the decision state.

TAY LODGE

This is problety hy leverthe plant to set Japanese in Seattle, My boyfried and Lordend rigin of scallege, Japanese snapsper beaseroil, and the appoint folia and 2 special rolls, because the rigin and seatific is as where the Seate excells, rift has deep rigin and seatified is as where the Seate excells, It is more drivery and the seateroilly size in commercial by other Yolgens was arranting. It is more drivery and the seateroilgray is the perfect amount of flavor for the delicate total.

This is mobility my florother piace to est, Jacanese Seats, My boyfering the piace to est, Jacanese Seats, My boyfering and in ordered right of scaling Jupanese reapper (passionals, and the agricular for special rolls. I would old pit he special rolls, because right and seather cut in where this place seeds. I also recommended by other Velbers was associng. If otherway and the successfury is the period amount of for the delicate late.

This product is a second of the control of the cont

Tris is probably my fevorite place to set Japanese in Seattle. My boyfered and collected right of scallor, Japanese images (scallor), and the agender for land of judgments of the seattle collection of the right and seaterini can't a where this place excells. The trist, an economic deal by other Yelpen was amazing. It's more chewy and the seatury large yield the perfect amount of flavor for the desicate total. What are the positive things being said about my restaurant?



#### 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.

#### Sentiment classifier

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



Sentence Sentiment Classifier

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







#### Use the sentiment classifier model!

Sentences from all reviews for my restaurant

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.

Show sentences with "positive" prediction on website

Classifier

MODEL

Sentences predictedto be positive

$$\hat{y} = +1$$

Easily best sushi in Seattle.

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

All the sushi was delicious.

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

Sentences predicted to be negative

$$\hat{y} = -1$$

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

My wife tried their ramen and it was pretty forgettable.

The service is somewhat hectic.

What does it mean for a classifier to be good?

### Previously, we asked the question: "What is good accuracy?"

#### We explored accuracy of random classifier as baseline

- For binary classification:
  - Half the time, you'll get it right! (on average)
    - $\rightarrow$  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...

# 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!

# Automated marketing campaign cares about something else...

Website shows 10 sentences from recent reviews



#### **PRECISION**

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



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

Accuracy doesn't capture these issues well...

# Precision: Fraction of positive predictions that are actually positive

# What fraction of the positive predictions are correct?

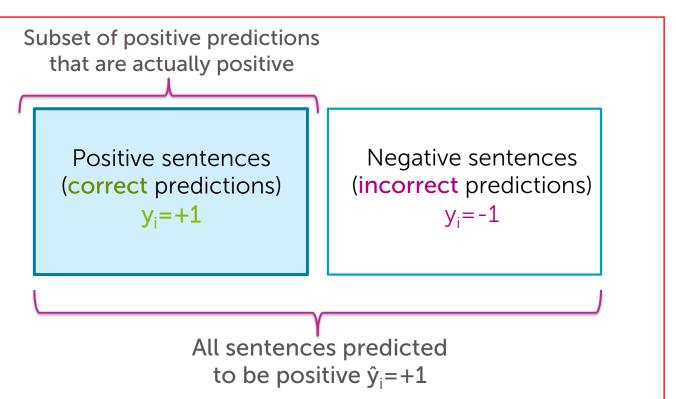
Sentences predicted to be positive:  $\hat{y}_i = +1$ 



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

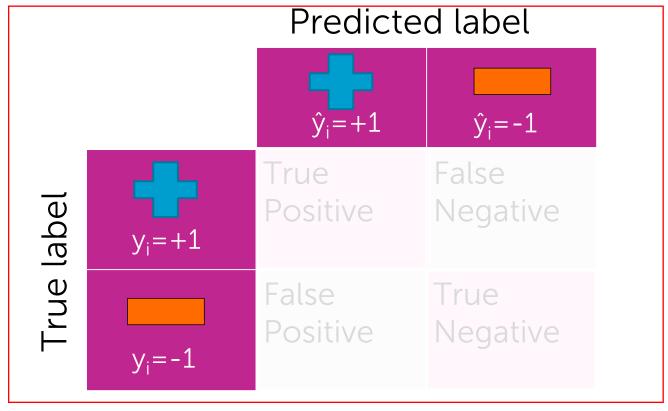
©2015-2016 Emily Fox & Carlos Guestrin

# **Precision**: Fraction of positive predictions that are actually positive



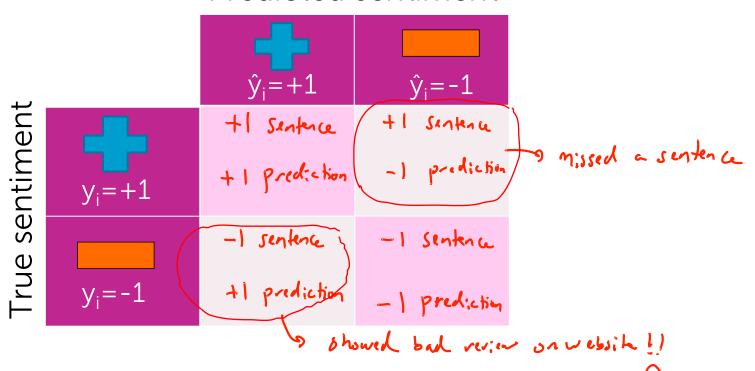
1 /

#### Types of error: Review



# Confusion matrix for sentiment analysis

Predicted sentiment



#### **Precision** - Formula

Fraction of positive predictions that are correct

```
precision = # true positives
# true positives + # false positives
```

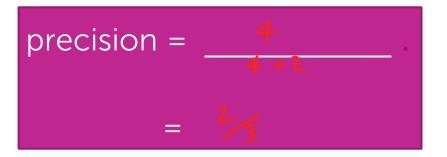
- Best possible value : 1.0

- Worst possible value : 0.0

#### Example: Calculating precision

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





#### Why precision is important

#### Shown on website

Sentences predicted to be 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.

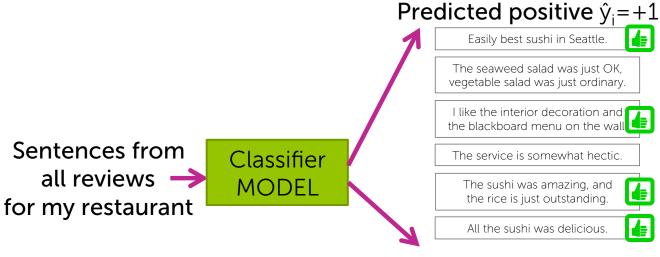
2 negative sentences shown to potential customers... 🙁

High precision means positive predictions actually likely to be positive!

#### Recall: Fraction of positive data predicted to be positive

©2015-2016 Emily Fox & Carlos Guestrin

#### Did I find all the positive sentences?





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.

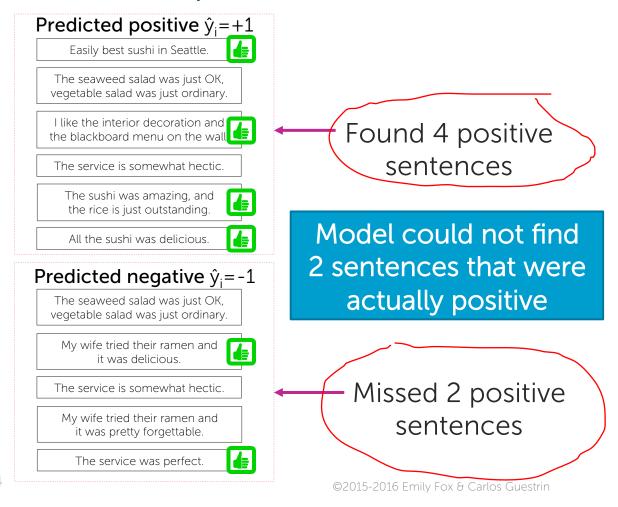
My wife tried their ramen and it was pretty forgettable.

23

©2015-2016 Emily Fox & Carlos Guestrin

Machine Learning Specialization

#### What fraction of positive sentences were missed out?



# Recall: Fraction of positive data predicted to be positive

Sentences predicted to be positive (correct predictions)  $\hat{y}_i$ =+1

Subset of positive data points correctly identified

data points y=+1

Sentences predicted to be negative (incorrect predictions)  $\hat{y}_i = -1$ 

#### **Recall** - Formula

Fraction of positive data points correctly classified

```
Recall = # true positives
# true positives + # false negatives
```

- Best possible value : 1.0

- Worst possible value : 0.0

#### Why is recall important?



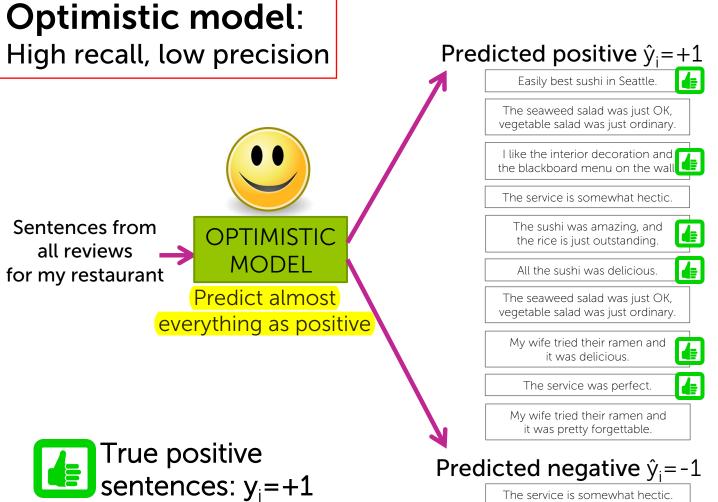
Want to show positive sentences on website

2 positive sentences not shown to potential customers...  $\otimes$ 

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

Machine Learning Specialization

# Precision-recall extremes

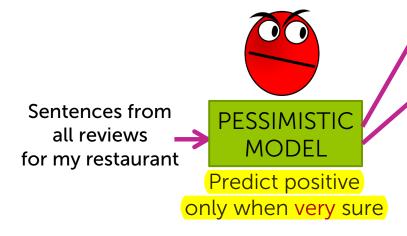


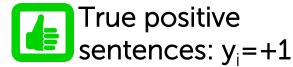
©2015-2016 Emily Fox & Carlos Guestrin

31



High precision, low recall





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.



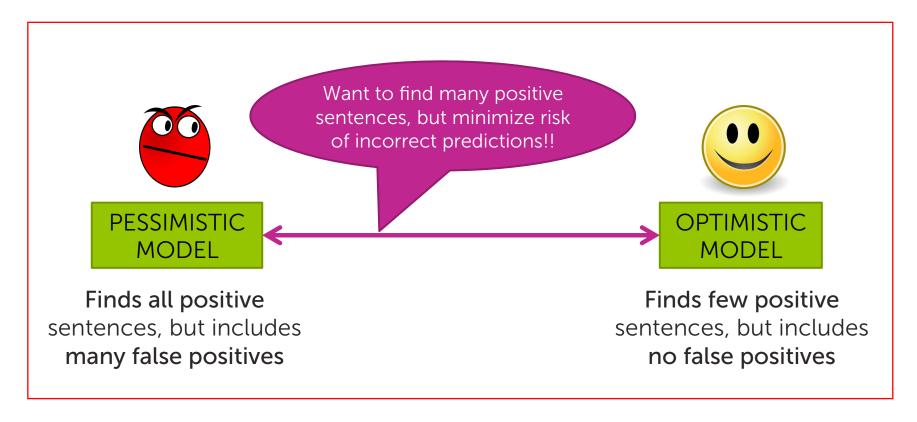
The service was perfect.



My wife tried their ramen and it was pretty forgettable.

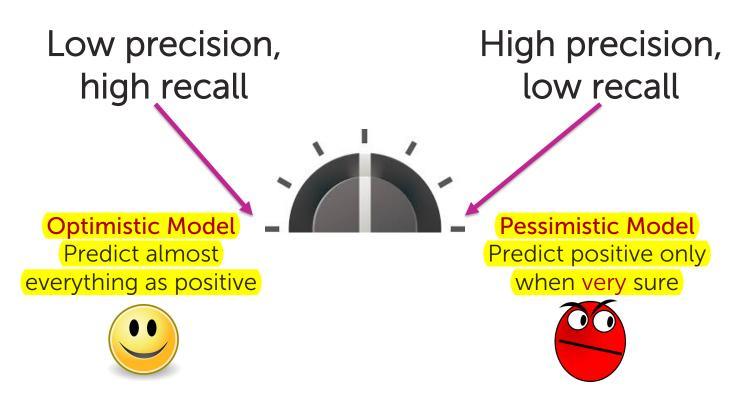
The service is somewhat hectic.

#### Balancing precision & recall



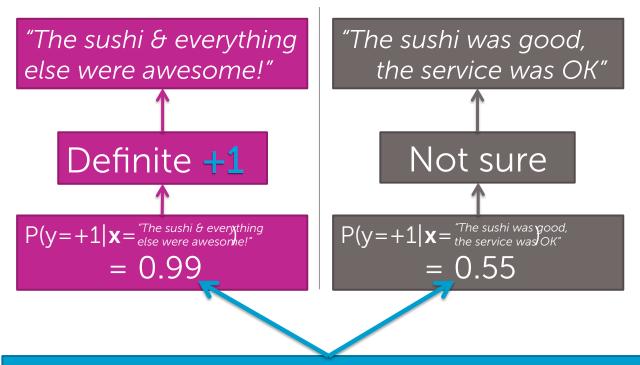
# Tradeoff precision and recall

#### Can we tradeoff precision & recall?



©2015-2016 Emily Fox & Carlos Guestrin

#### How confident is your prediction?



Can be used to tradeoff precision and recall

#### Basic classifier



Sentence from review

Input: **x**<sub>i</sub>

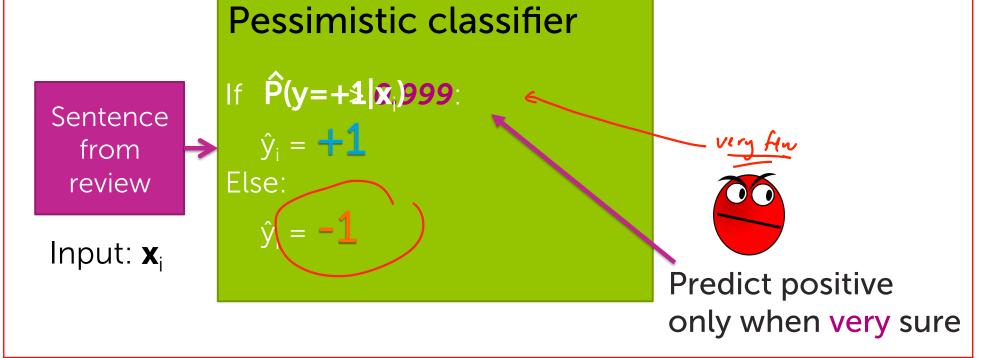
If 
$$\hat{P}(y=+1|x|)5$$

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

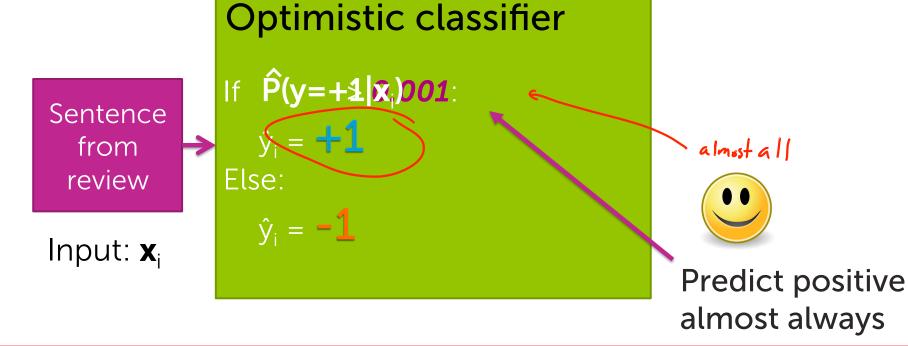
Else:

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

#### Pessimistic: High precision, low recall



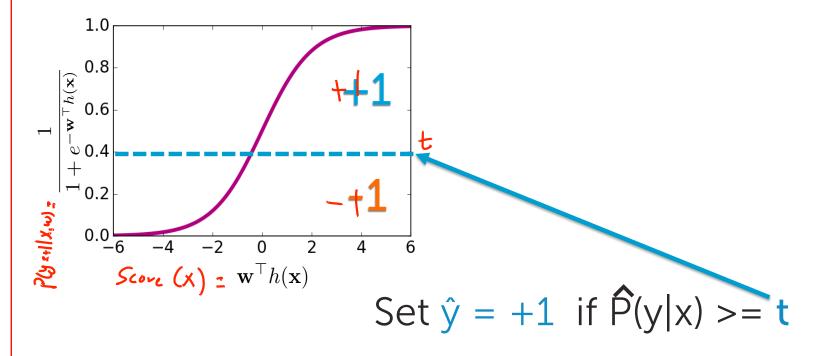
#### Optimistic: Low precision, high recall



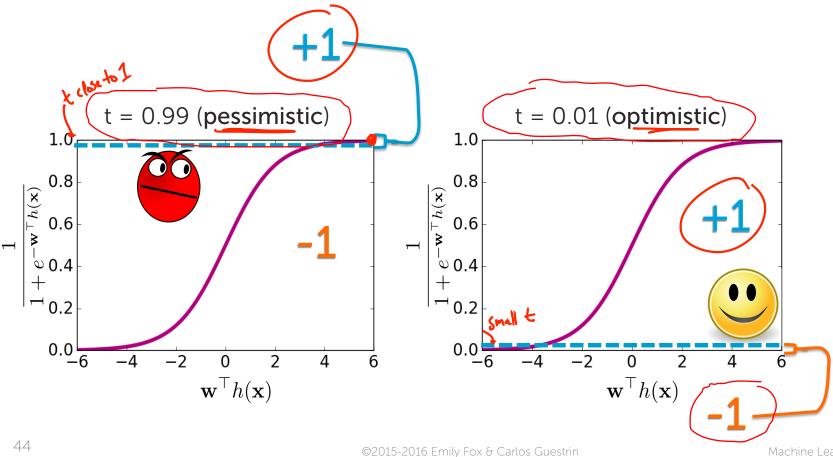
©2015-2016 Emily Fox & Carlos Guestrin

#### Prediction probability threshold

Probability t above which model predicts true

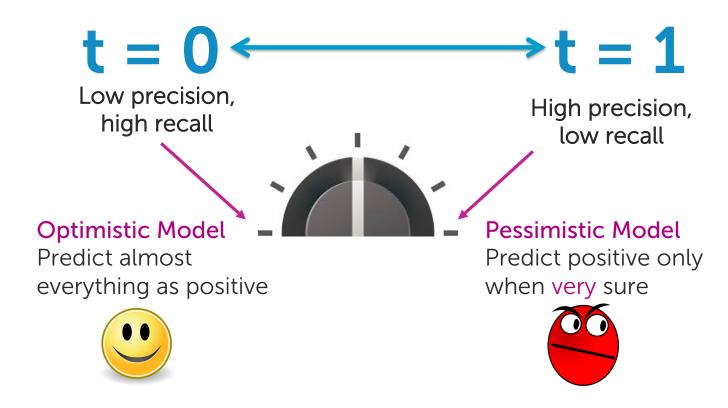


#### Example threshold values



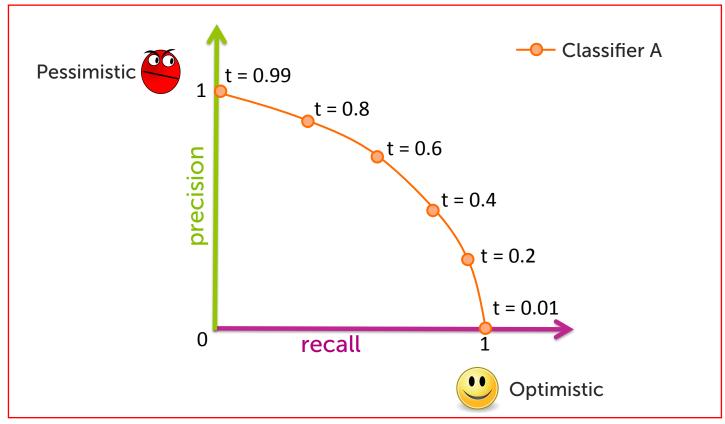
Machine Learning Specialization

#### Tradeoff precision & recall with threshold

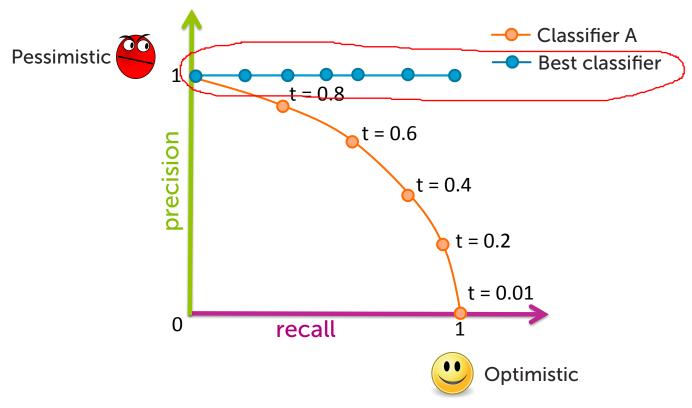


# Precision-recall curve

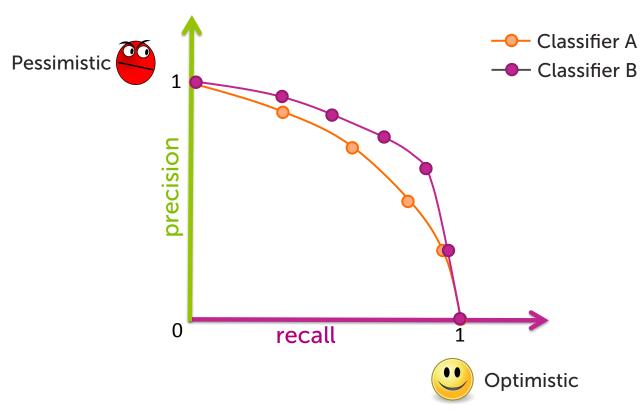
#### The precision-recall curve



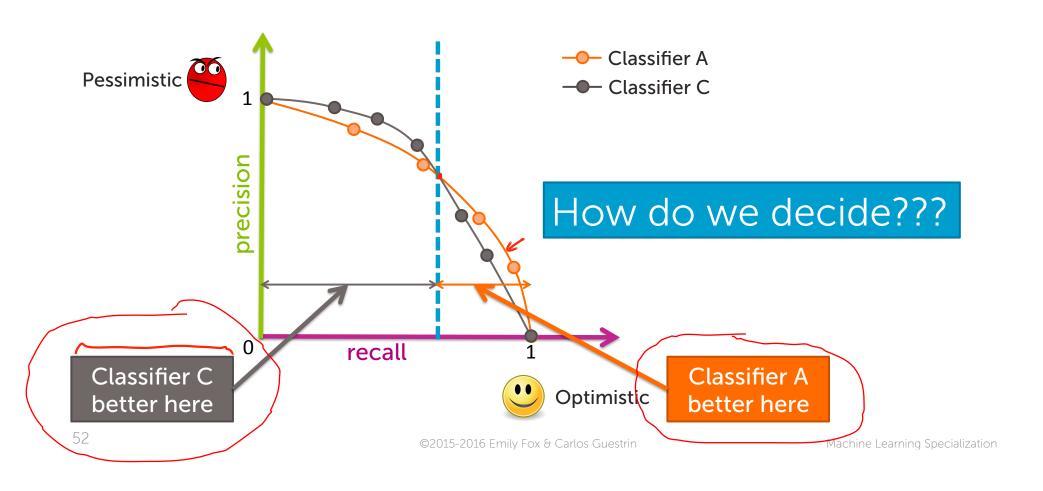
#### What does the perfect algorithm look like?



#### Which classifier is better? A or B?



#### Which classifier is better? A or C?



#### 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

# Summary of precision-recall

#### 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

#### Thank you to Dr. Krishna Sridhar



Dr. Krishna Sridhar Staff Data Scientist, Dato, Inc.