

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 suspope pleasuroal, and the ageleate from Lard 2 special role. I revuid slop he special role, Securate the right and safety action is where the face secolis. This total, so securimended by other Polpers was amonting. It's more commended by other Polpers was amonting. It's more commended by other Polpers was amonting. It's more contribution of the commended of the commended

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

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 free first first and 2 mass free first first and 2 mass free first f

This is probably my feworths place to eat Japanese in bestite. My boyfered and Lordered right of society, goggenee mayope beaution, and the appeared to during the proposal refs. Livedd sight the special risk, secause the gift and societies can swhere this place easing a first disk and the case where this place easing a first disk and the case where this place easing a first place of the case of the

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 in forth and skip this special risk. Declares the right and soldhin could skip the special risk, because the right and soldhin could skip the special risk, because the special could be a sincernment of the other shapes and the soldhings by the perfect amount of their for the deficials total.

7/21/2015
This is protebly my flooring place to est Japanese in Sestera My boyleng place to est Japanese in Sestera My boyleng and cliedlered nigit of exalize, Japanese snapone beasonal, and the agedabil frost used 10% place special most invacid sight the special instals. Invacid sight the special instals, because the nigit and seatherin curs is where this place seeds. The total on socientered code you've five the view send analoge. It's more cliedly and the saucokygany is the perfect amount of flowor for the delicious flow.

TA LODIS

This is problety hysterile plant to set Japanese in Seattle, My boyfried and Lordend rigin of scallege, Japanese snapped pleasacelle, and the seagethat fell and 2 special rolls, because the rigin and seatiline is as where the Seattle, because the rigin and seatiline is as where the Seattle, because the rigin and seatiline is as where the Seattle, because the rigin and seatiline is as where the Seattle, because the rigin and seatiline is a where the Seattle, because the rigin and seatiline class where the Seattle seattle is a seattle s

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 appeal rolls, because right and seather cut in where this place seeds.) Lecture of the appeal rolls, because right and seather cut in where this place seeds. The association of the seather cut is where the place seeds in charge of the seather cut is where the place seeds in the charge of the seather cut is where the period amount of 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 specific tool and of seattle for land of right and seaterini can't a where this place excells. The fort, an encorrected by other Yelpes was amazing. It's more chewy and the seatury large yield the perfect amount of flavor for the desicate tool. 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

Subset of positive predictions that are actually positive

Positive sentences (correct predictions) $y_i=+1$

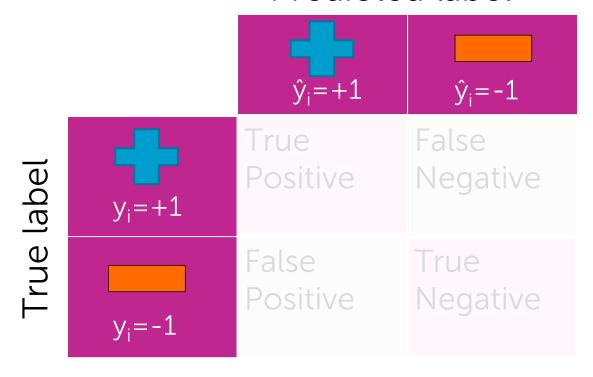
Negative sentences (incorrect predictions) $y_i = -1$

All sentences predicted to be positive $\hat{y}_i = +1$

Types of error: Review

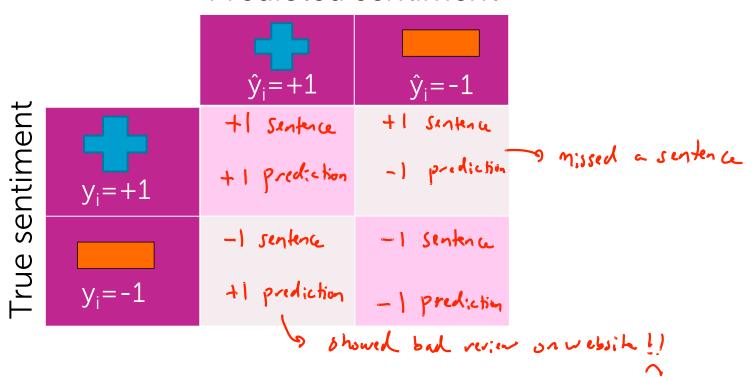
Predicted label

©2015-2016 Emily Fox & Carlos Guestrin



Confusion matrix for sentiment analysis

Predicted sentiment



Precision - Formula

Fraction of positive predictions that are correct

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

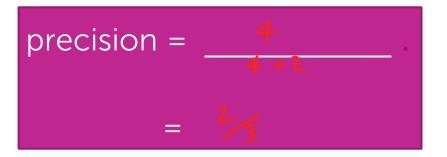
- Best possible value : 1.0

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

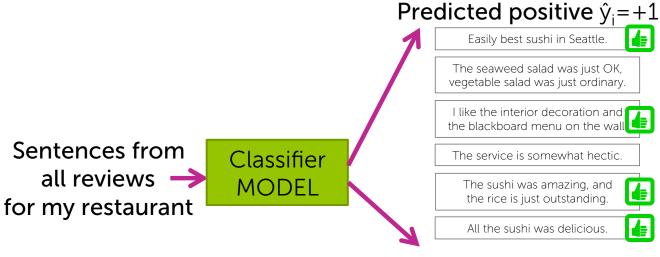


2 negative sentences shown to potential customers... 🙁 🔽

High precision means positive predictions actually likely to be positive!

Recall: Fraction of positive data predicted to be positive

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.

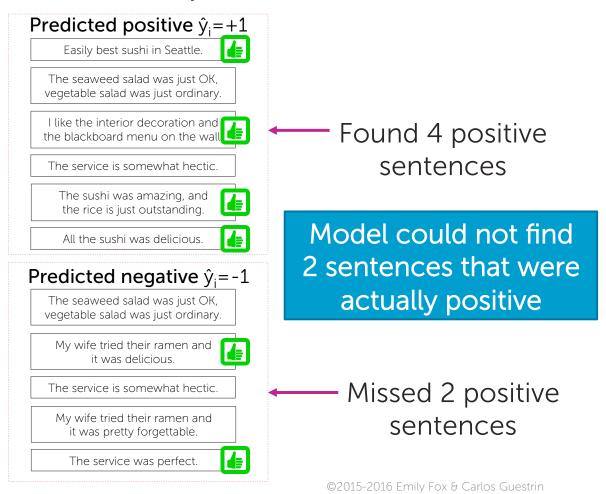
My wife tried their ramen and it was pretty forgettable.

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

All positive data points y=+1

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

Recall - Formula

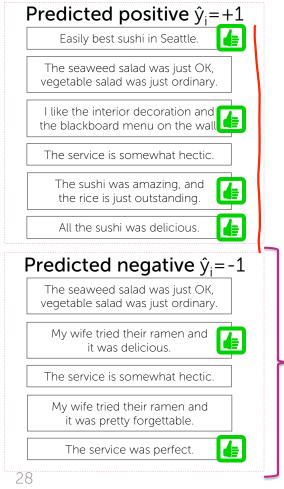
Fraction of positive data points correctly classified

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

- Best possible value : 1.0

- Worst possible value : 0.0

Why is recall important?



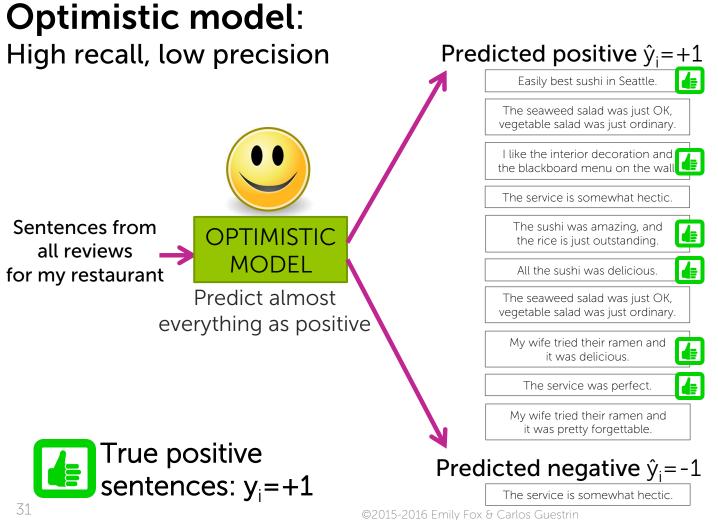
Want to show positive sentences on website

2 positive sentences not shown to potential customers... 🕾 High recall means positive data points are very likely to be discovered!

©2015-2016 Emily Fox & Carlos Guestrin

Machine Learning Specialization

Precision-recall extremes





High precision, low recall

00

Sentences from all reviews for my restaurant

PESSIMISTIC MODEL

Predict positive only when very sure

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.



True positive sentences: y_i=+1

My wife tried their ramen and it was pretty forgettable.

The service was perfect.

My wife tried their ramen and it was delicious.

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

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

Easily best sushi in Seattle.

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

The service is somewhat hectic.

32

©2015-2016 Emily Fox & Carlos Guestrin

Machine Learning Specialization

Balancing precision & recall



Finds all positive sentences, but includes many false positives

Finds few positive sentences, but includes no false positives

The description for each model is for the opposite. Need to swap them to ensure the image is correct.

Tradeoff precision and recall

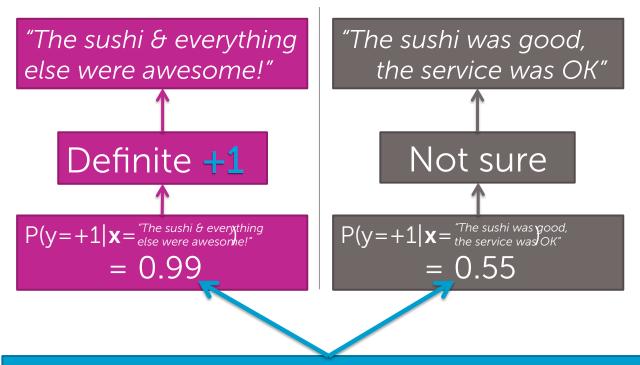
Can we tradeoff precision & recall?

Low precision,
high recall

Optimistic Model
Predict almost
everything as positive

Pessimistic Model
Predict positive only
when very sure

How confident is your prediction?



Can be used to tradeoff precision and recall

Basic classifier



Sentence from review

Input: **x**_i

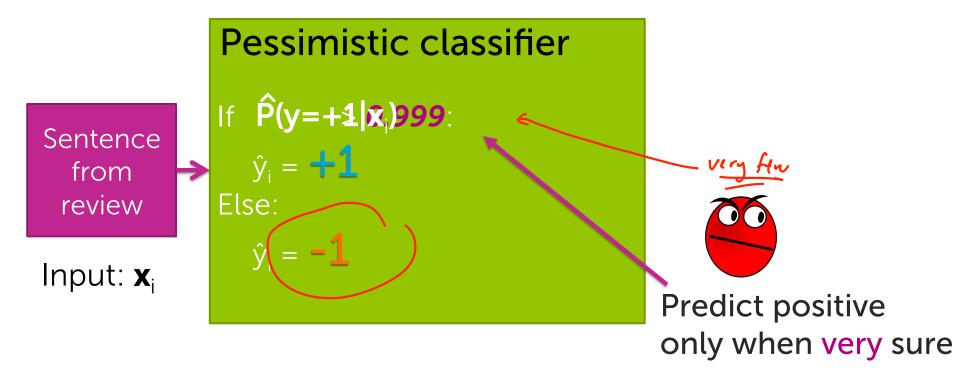
If
$$\hat{P}(y=+1|x|)5$$

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

Else:

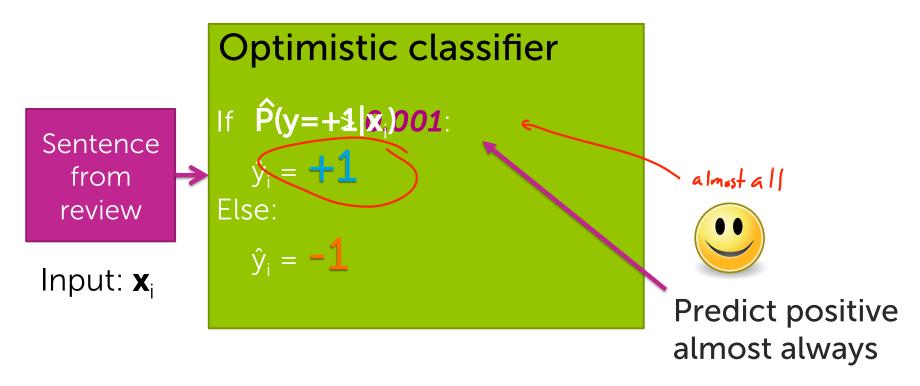
$$\hat{y}_i = -1$$

Pessimistic: High precision, low recall



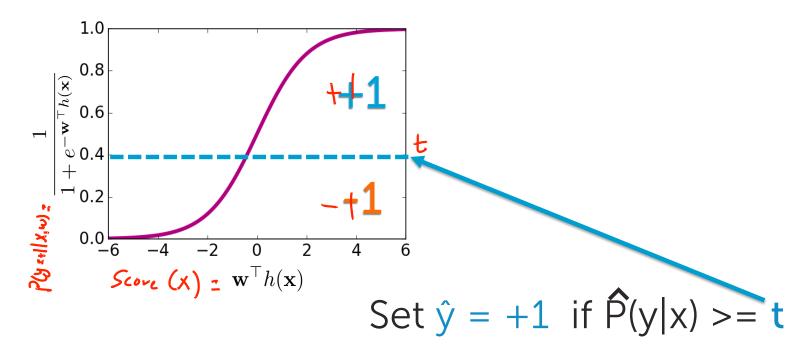
©2015-2016 Emily Fox & Carlos Guestrin

Optimistic: Low precision, high recall

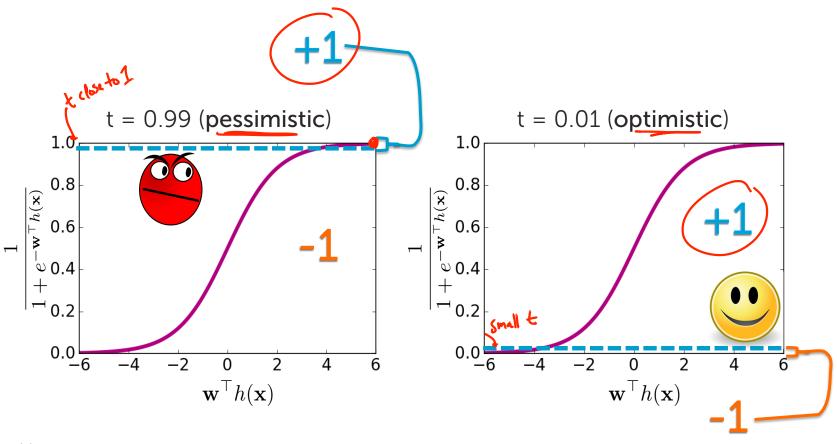


Prediction probability threshold

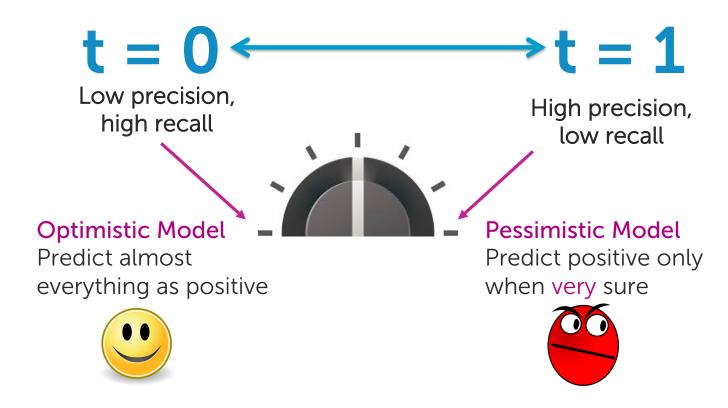
Probability t above which model predicts true



Example threshold values

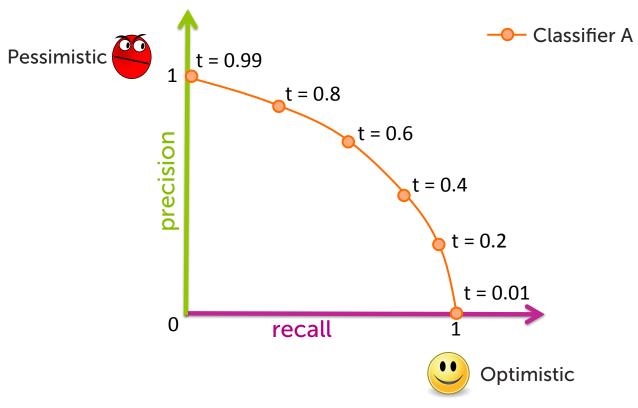


Tradeoff precision & recall with threshold

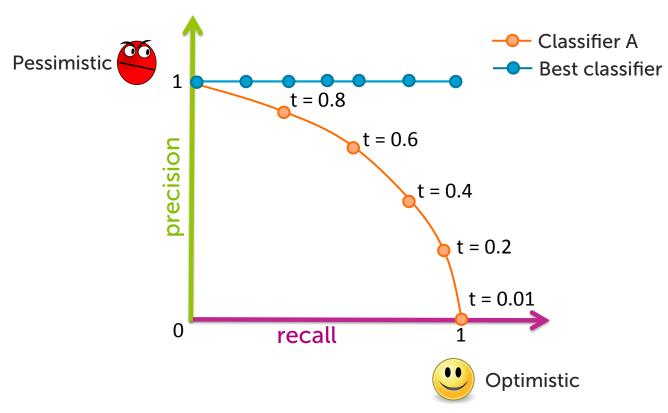


Precision-recall curve

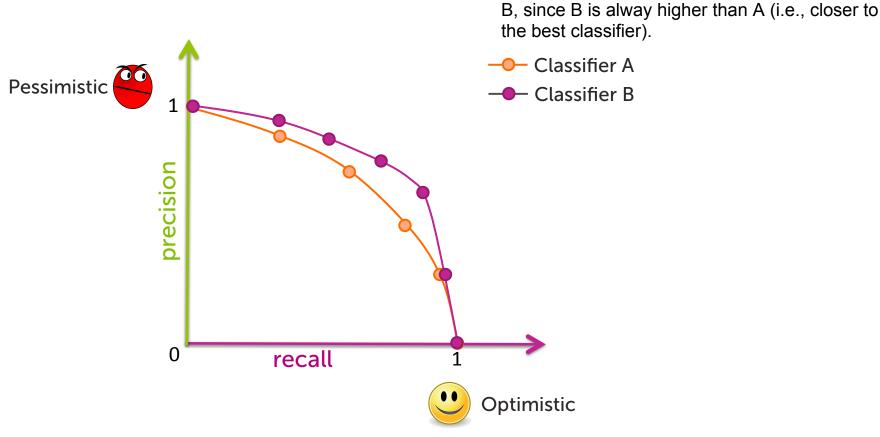
The precision-recall curve



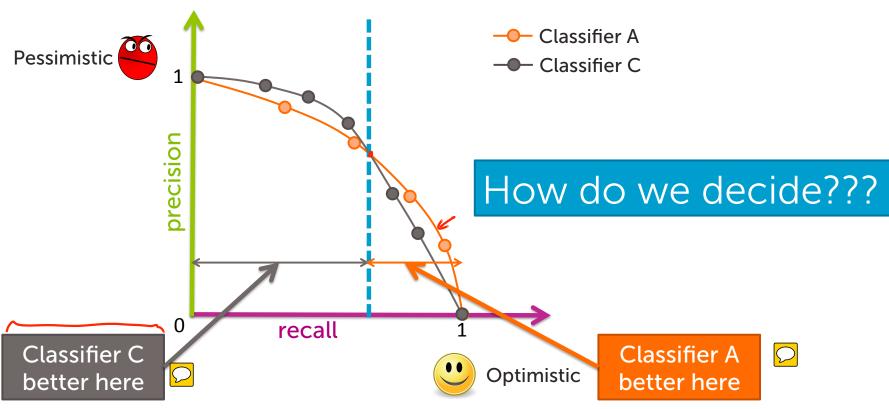
What does the perfect algorithm look like?



Which classifier is better? A or B?



Which classifier is better? A or C?



52

Compare algorithms

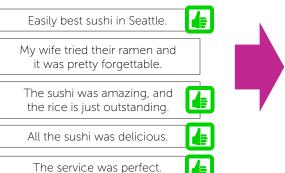
- Often, reduce precision-recall to single number to compare algorithms
 - F1 measure, area-under-the-curve (AUC),...

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.