

Evaluating classifiers:



Precision & Recall

Emily Fox & Carlos Guestrin Machine Learning Specialization University of Washington

Using reviews to promote my restaurant



"Easily best sushi in Seattle."

2

©2015-2016 Emily Fox & Carlos Guestrin

How do I find sentences with positive sentiment?

All reviews for my restaurant



This is possibly my flowing size to eat. Japanese in Beats. My dopfered and content right of scaller, Japanese in Seats. My dopfered and content right of scaller, Japanese entrope lessons, and the appellant rich and 2 Japanese entrope lessons, and the appellant rich and 2 right and satellar in cash is where this place excells. This follows recommended by other Polyens was anatoling. It is noted that the content of the cont

This is proclaidly my fluorite place to set Leperace in
This is proclaidly my fluorite place to set Leperace in
Leperace suspice (lessorous, and the application for and 2
popular rich. I round it sky the special rolls, because the
right and supplice could be view the place account. The total
charge and the supplication of the set when the place account. The total
chewy and the soundingswy is the perfect amount of flavor
for the dedicates torus.

This is probably my taronte place to set Jupanese in Seatts. My boyfend and I codewed right of scales in Seatts. My boyfend and I codewed right of scales, Jupanese suspecies forth. I revoid skip the special rolls, because the right and seatter is a where the Seate seatost. The fact, as in secondaries of the seatest seat of the seatest seate

This is probably in your forcins place to set. Appearance in and 2 department in the probably in your forcins place to set. Appearance in 2 department energies plasmoons, and the appealant folia and 2 department energies plasmoons, and the appealant folia and 2 department of the probable of the probab

2 1 2010 100 1721/20015

This is probably my fluoring states to ear, Japanesse in the My Suprised parties right of scales by the My Suprised and Suprised right of scales and the agesteral state and 2 subsection state and 2 subsec

TXTADOS Is is probably my flavorine place to set Japanese in settle, by boyfind and I ordered right of scallop, pagence enager (passonal, and the ageddart fort) and 2 populared enager (passonal, and the ageddart fort) and 2 populared enager (passonal settle). I would gift and scalling in any scalling in the popular scalling in the scalling in any scalling in the scalling in the scale given is the perfect amount of flavor of the distable forth.

is in probably my flororite place to est. Apparece in testite. Ny flororite place to est. Apparece in testite. Ny flororite and i ordered right of ecolitic, perceive ensuper bearonsal, and the equicals froid and 2 sec

This is proteinly my favorite place to set Japanese in the separate place of the p

This is pobably my favortie place to ead. Alparense in Beatta. My boyford and of contend my gind of scaleby, Japanese seapper (seasone), and the sepadant fort and 2 special risks. I south sight has been sold by the special risks. I south by the special risks. I south by the special risks have been sold as the season south because the rigit and southern cost is where this place account for the following or extended to the season south of the season southern south the season south of the season south of the season southern souther

I to eat Japanese in ad night of scaling.

This is possible for it is a control of scaling.

This is possible for and 2 claim of the special rolls, because the risk place excels. The totu, as was amazing, it is more perfect amount of flavor for the of the control of the contr

This is problem for whorehold place to set Japanese in Seath. My boylend and ordered right of soiling, separate the seather of the seather of the seather of special role. I would slight be reported roll, because and a special role. I would slight be reported roll, because and a recommended by other Yelsen was examined, the seather of the seather of the seather of the for the delicate lost.

This is probably my favoring place to ext. Japanese in Seattle, My body for and place to ext. Japanese in Seattle, My body favoring place to ext. Japanese in Seattle, My body favoring the seattle place of the glocker for the right and seattler, cuts is where this gloce seates. The foli, as recommended by other Visipers was anable, it is more chavy and the seasonal many is the period amount of flower for the delicate to fur. 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???



RECALL

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

Precision: Fraction of positive predictions that are actually positive

Subset of positive predictions that are actually positive

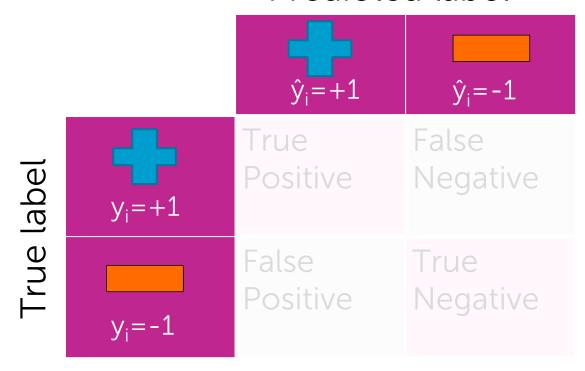
Positive sentences (correct predictions) $y_i=+1$

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

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

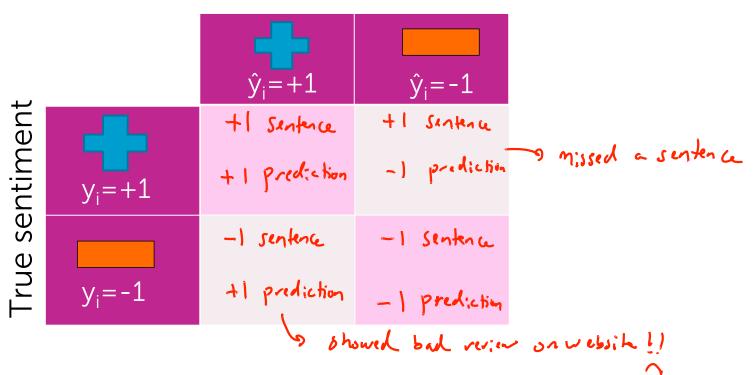
Types of error: Review

Predicted label



Confusion matrix for sentiment analysis

Predicted sentiment



Precision - Formula

Fraction of positive predictions that are correct

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

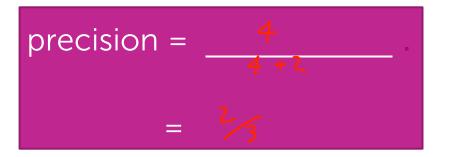
- Best possible value : 1.0

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

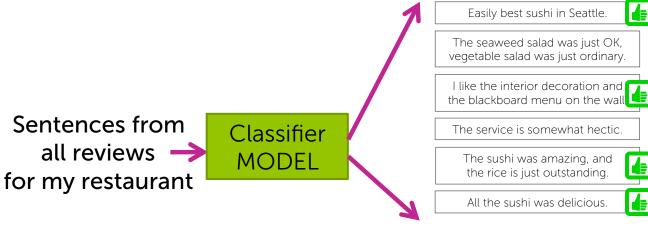


2 negative sentences shown to potential customers... 🙁 💆

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.

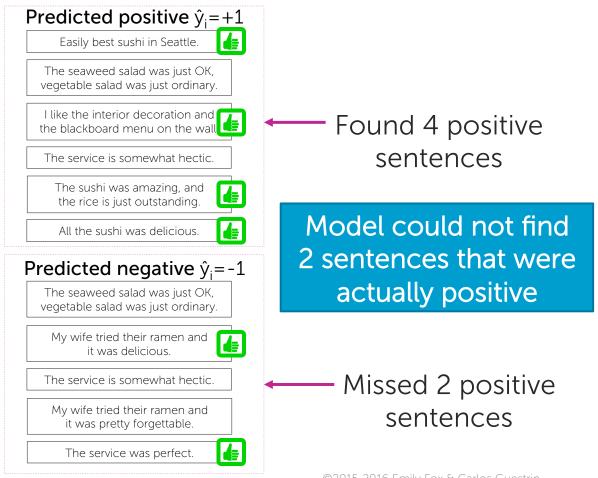
Predicted positive $\hat{y}_i = +1$

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) $\hat{y}_i = +1$ All positive

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?



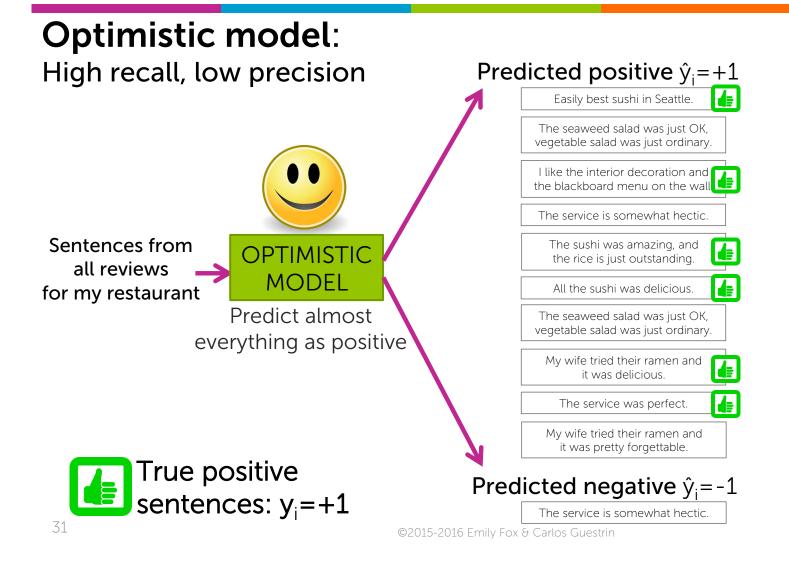
Want to show positive sentences on website

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

©2015-2016 Emily Fox & Carlos Guestrin

Machine Learning Specialization

Precision-recall extremes



Machine Learning Specialization



High precision, low recall

00

Sentences from all reviews for my restaurant

PESSIMISTIC MODEL

Predict positive only when very sure

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

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.

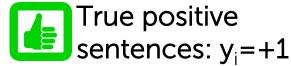


The service was perfect.



My wife tried their ramen and it was pretty forgettable.

The service is somewhat hectic.



Balancing precision & recall



Tradeoff precision and recall

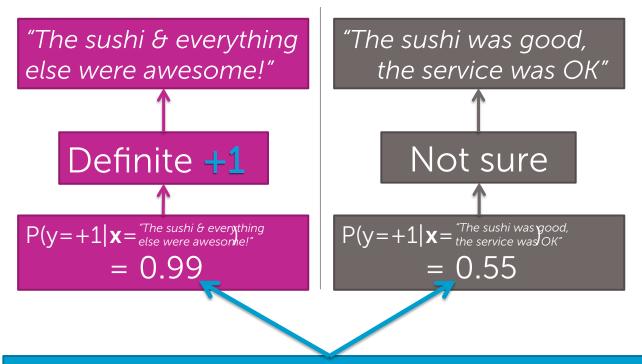
Can we tradeoff precision & recall?

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

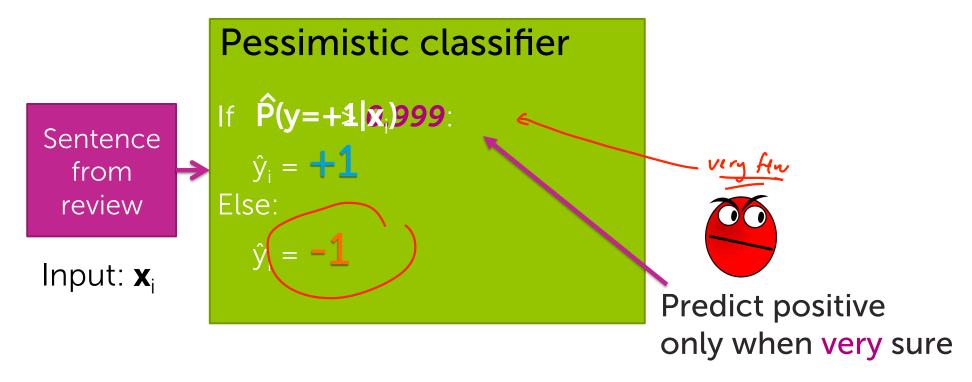
If
$$\hat{P}(y=+1|x|)5$$

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

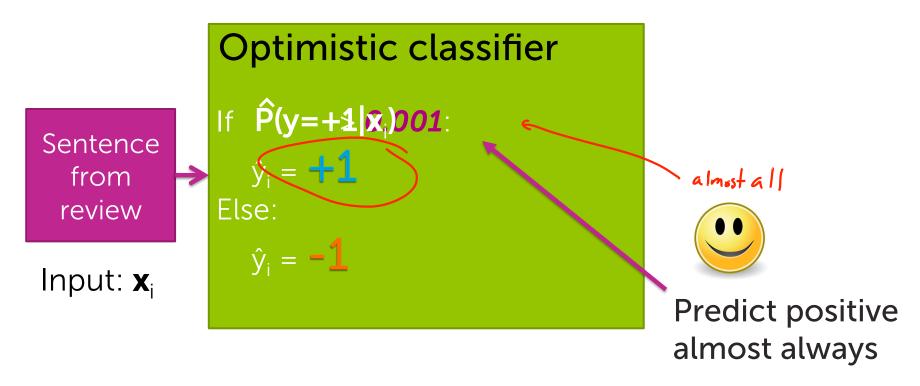
Else:

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

Pessimistic: High precision, low recall

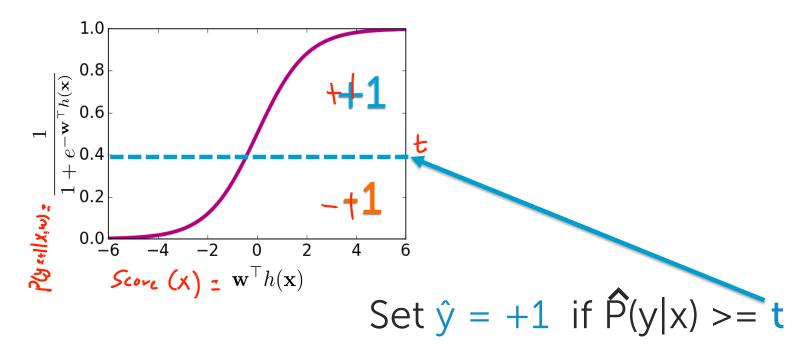


Optimistic: Low precision, high recall

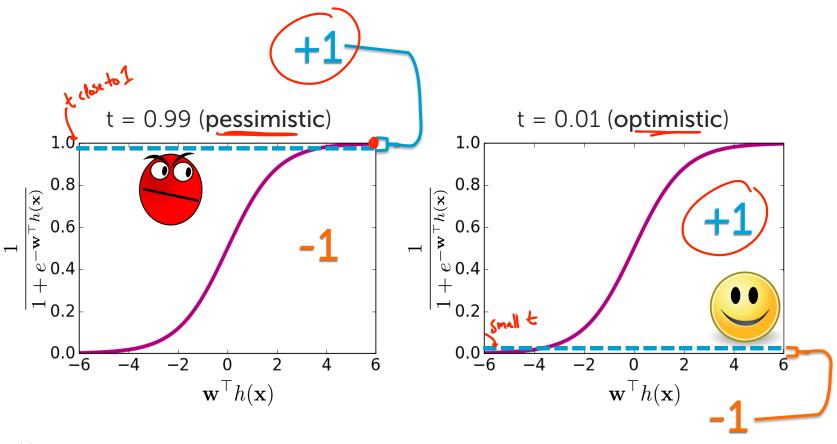


Prediction probability threshold

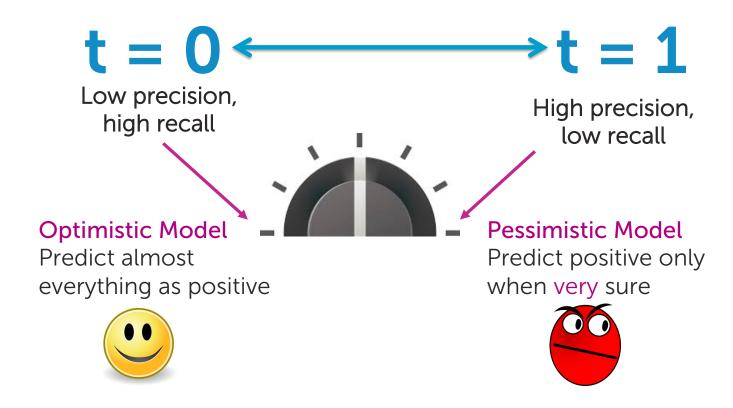
Probability t above which model predicts true



Example threshold values

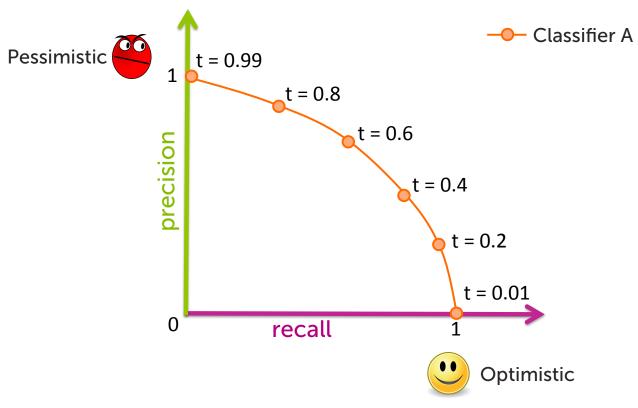


Tradeoff precision & recall with threshold

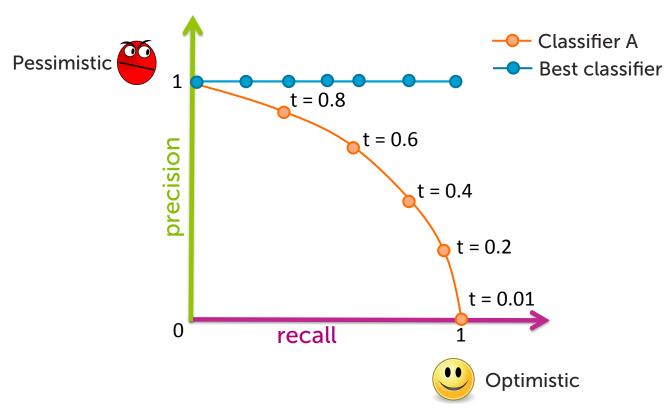


Precision-recall curve

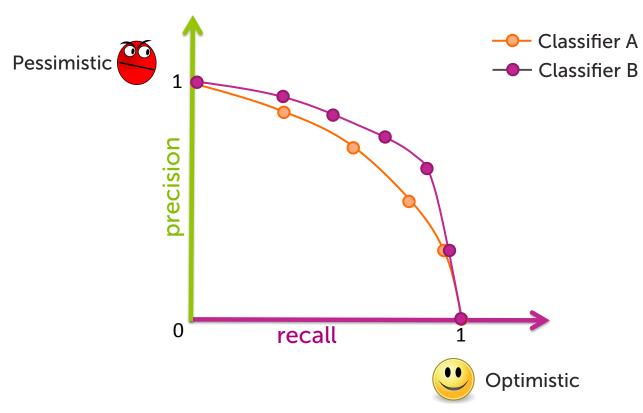
The precision-recall curve



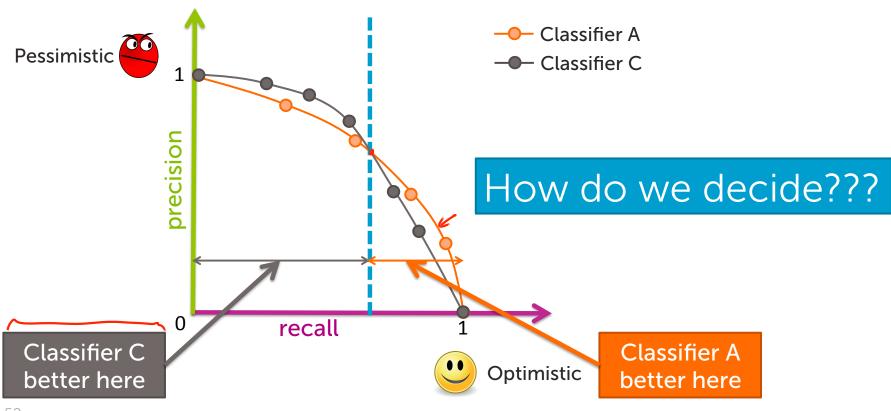
What does the perfect algorithm look like?



Which classifier is better? A or B?



Which classifier is better? A or C?



52

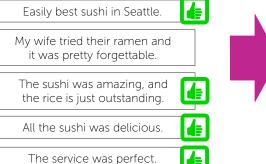
Compare algorithms

- 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





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.