

Precision - Recall

Case-study where accuracy is not best metric for classification

Using reviews to promote my restaurant

Goal: increase
guests by 30%



Reviews A hand-drawn pink arrow pointing upwards.

Need an automated,
"authentic"
marketing campaign



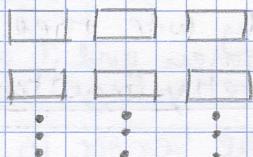
Great quotes
"Easily best sushi in
Seattle"



Great
spokespeople

How do I find sentences with positive sentiment?

All reviews for
my restaurant



"What are the positive
things being said
about my restaurant?"

Sentiment classifier

Input x : 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

Classifier Model

Show sentences with "positive" prediction on website

Sentences predicted to be positive
 $y = +1$

Easily best sushi in Seattle

I like the interior decoration and the blackboard menu

All Sushi was delicious

The sushi was amazing, and the rice is outstanding

Sentences predicted to be negative
 $y = -1$

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 is good performance for a classifier?

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)
→ classification error = 0.5
- For K classes, error = 1 - 1/K
 - error = 0.66% for 3 classes, 0.75 for 4 classes

At the very, very, very least,
you should hopefully 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 $\hat{y}_i = +1$

Easily best sushi in Seattle ✓

The seaweed salad was just OK
Vegetable salad was just ordinary X

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

The service is somewhat hectic X

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

All the sushi was delicious ✓

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

		$\hat{y}_i = +1$	$\hat{y}_i = -1$	
True label	$y_i = +1$	True positive	False negative	missed a sentence
	$y_i = -1$	False positive	True negative	showed bad review on website!!!

Precision - formula

- Fraction of positive predictions that are correct

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

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

Example: Calculating precision

4 correct 2 mistakes precision = $\frac{4}{4+2} = \frac{2}{3}$

Why precision is important

2 negative sentences shown to potential customers... if

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?
What fraction of positive sentences were missed out?

Predicted positive $\hat{y}_i = +\frac{1}{2}$

"Easily best sushi in Seattle"



"The seaweed salad was 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"

Found 4
positive
sentences

"All the sushi was delicious"

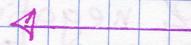
Predicted negative $\hat{y}_i = -\frac{1}{2}$

"The seaweed salad was just OK, Vegetable Salad was just ordinary"

"My wife tried their ramen and it was delicious"

Model could not find
2 sentences that
were actually
positive

"The service is somewhat hectic"



Missed 2
positive
sentences

"My wife tried their ramen and it was pretty forgettable"

"The service was perfect"



Recall: Fraction of positive data predicted to be positive

All positive data points
 $y = +1$

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

Subset of positive data points correctly identified

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

Recall - Formula

- Fraction of positive data points correctly classified

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

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

Why is recall important?

↳ positive sentences not shown to potential customers... :-)

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

Precision - Recall extremes

Optimistic model:

High recall, low precision

Sentences from all reviews for my restaurant

→ **OPTIMISTIC MODEL**

Predicts almost everything as positive

Pessimistic Model

High precision, low recall

Sentences from
all reviews
for my restaurant

PESSIMISTIC
MODEL

Predict positive
only when very sure

Balancing precision & recall

Want to find many positive
sentences, but minimize
risk of incorrect
predictions!!

PESSIMISTIC
MODEL

OPTIMISTIC
MODEL

Finds all positive
sentences, but includes
many false positives

Finds few positive
sentences, but includes
no false positives

Trading off precision and recall

How confident is your prediction?

"The sushi & everything
else were awesome!"

Definite +1



$P(y=+1 | x = \text{"The sushi & everything else were awesome!"})$

$$= 0.99$$

"The sushi was good,
the service was OK"

Not sure



$P(y=+1 | x = \text{"The sushi was good, the service was OK"})$

$$= 0.55$$

Can be used to trade off
precision and recall

Pessimistic: high precision, low recall

Pessimistic classifier

Sentence from review

If $P(\hat{y} = +1 | x_i) > 0.999$:

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

very few

Input: x_i

Else:

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

Optimistic Low precision, high recall

Optimistic classifier

Sentence from review

If $\hat{P}(y = +1 | x_i) > 0.001$:

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

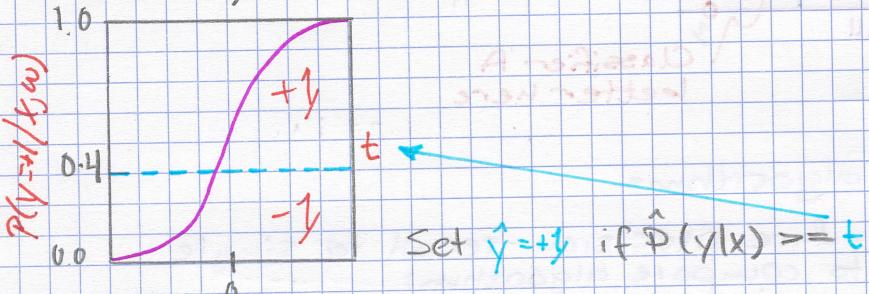
almost all

Input: x_i

Else

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

Prediction probability threshold
Probability t above which model predicts true



$$\text{Score}(x) = w^T x$$

Example threshold values:

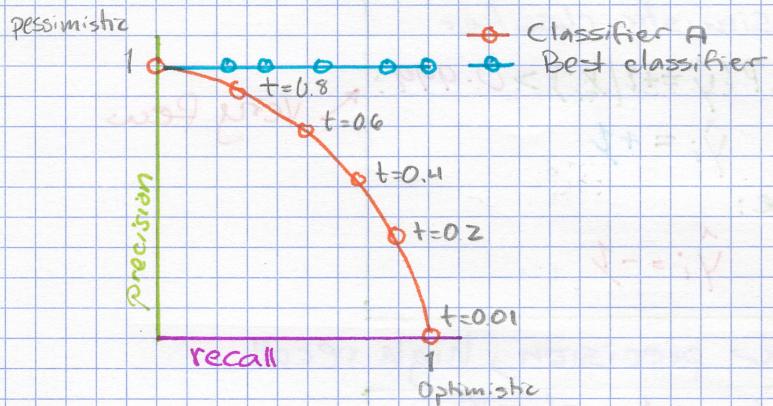
$t = 0.99$ (pessimistic)

t close to 1

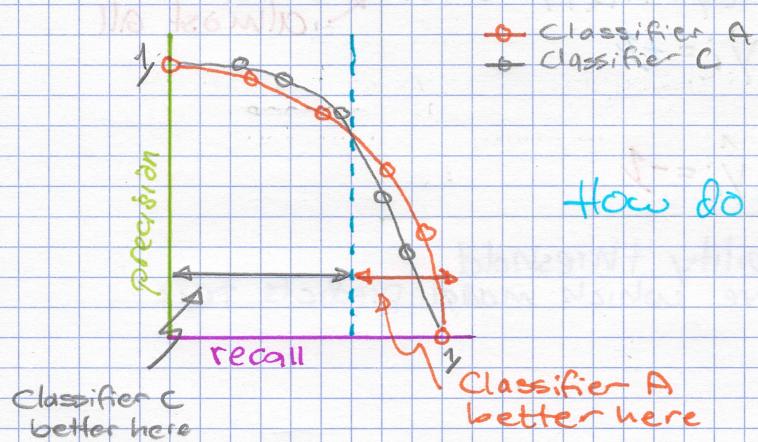
$t = 0.01$ (optimistic)

small t

Precision-recall curve



Which classifier is better? A or C?



How do we decide??

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

"Easily best sushi in Seattle" ✓



precision
at $k=5$

"My wife tried their ramen and it was pretty forgettable"

"The service was perfect" ✓