DD2418 Language Engineering
5a: Text classification

Johan Boye, KTH

March 30, 2020

Is this spam?

Hello

Want to gain good money with no effort and spend little time on that?

As of now, you can gain good money with our binary options service. Remember that for this month we have paid $11,000,000 to our users

For your awesome, we have prepared a bonus for you: $1,000.

Everything is ready. Just click SIGN UP and start earning!

SIGN UP

Is this spam?

Hello

Want to gain good money with no effort and spend little time on that?

As of now, y_____od money with our binary options service. Remember that for this month we have pai___$11,000,000___o our users

For your awesome, we have prepared a bonus for you $1,000.

Everything is ready. Just click SIGN UP and sta___earning!

SIGN UP

© 2018 All Rights Reserved

Is this spam?

From: Mihaela Ioana <contact@icef4.info>
Sent: den 23 september 2018 22:11
To: info@icef4.org
Subject: You are invited to Amsterdam Science Park on 14 December

Dear colleagues

You are now welcome to participate to 4th International Conference on Engineering in the premises of **Amsterdam University Science Park**, with your abstract, full paper and poster, as in-person or virtual.

The conference topics include **Engineering** fields in Aerospace, Biomedical, Chemical, Civil, Electrical, Computer Science, Environmental, Industrial, Materials, Mechanical, Nuclear as well as Formal Sciences such as Mathematics and Physics.

The accepted full papers will be published in the proceedings book with ISBN and in indexed journals by **De Gruyter**.

You can register with your **abstract** from now until 25 November 2018. The abstract book with ISBN will be published before the conference. Sending paper is optional. Full Proceedings and Journal will be published within 2 months after the conference.

The **fee starts from €170** and it varies depending on your preferences such as participation type, journal publishing etc.

You will receive an acceptance letter with an invoice within 7 days after submission based on your abstract. You will receive the receipt and an invitation letter upon the payment. The committee makes sure to accept your optional full paper afterwards. This is helpful for you to make early travel planning.

For more details and registration, please visit http://icef4.org

Looking forward to meeting you at Amsterdam Science Park.

Best Regards
Mihaela Ioana,
Secretary ICEF IV

Is this spam?

Dear colleagues

You are now welcome to participate to 4th International Conference on Engineering in the premises of **Amsterdam University Science Park**, with your abstract, full paper and poster, as in-person or virtual.

The conference topics include **Engineering** fields in Aerospace, Biomedical, Chemical, Civil, Electrical, Computer Science, Environmental, Industrial, Materials, Mechanical, Nuclear as well as Formal Sciences such as Mathematics and Physics.

The accepted full papers will be published in the proceedings book with ISBN and in indexed journals by **De Gruyter**.

You can register with your **abstract** from now until 25 November 2018. The abstract book with ISBN will be published before the conference. Sending paper is optional. Full Proceedings and Journal will be published within 2 months after the conference.

The **fee starts from €170** and it varies depending on your preferences such as participation type, journal publishing etc.

You will receive an acceptance letter with an invoice within 7 days after submission based on your abstract. You will receive the receipt and an invitation letter upon the payment. The committee makes sure to accept your optional full paper afterwards. This is helpful for you to make early travel planning.

For more details and registration, please visit http://icef4.org

Looking forward to meeting you at Amsterdam Science Park.

Best Regards
Mihaela Ioana,
Secretary ICEF IV

Uncanny, haunting, I must have read this novel at the right time for me as it found a sure spot under my skin and disturbed my normally peaceful sleep. positive review

If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut.

humouristic negative

Who wrote Henry VI, part 3?


Henry VI, Part III
William Shakespeare
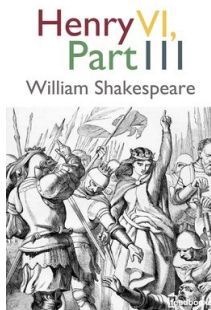
Shakespeare or Marlowe? In 1994, two researchers using neural network techniques found strong support that Marlowe wrote the original.

Merriam, T. V., & Matthews, R. A. (1994). Neural computation in stylometry II: An application to the works of Shakespeare and Marlowe. *Literary and Linguistic Computing*, 9(1), 1-6.

⬤CBS NEWS   NEWS   SHOWS   VIDEO   CBSN   MORE   🔍

By **MAYA KAUFMAN** / **CBS NEWS** / *August 6, 2015, 6:00 AM*

Language detectives make the web less anonymous
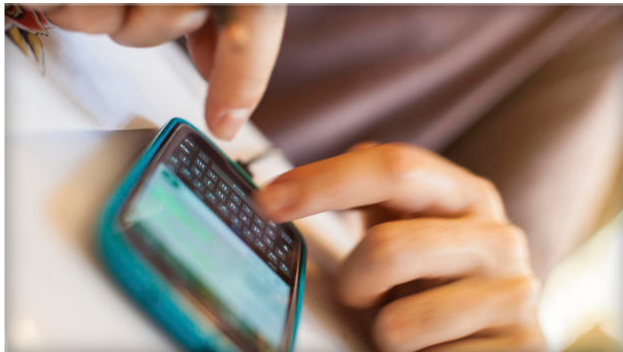
/ ISTOCKPHOTO

Text classification

We want to automatically assign a *label* to a document (=book, article, web page, paragraph, sentence).

It is assumed that every document has exactly one label.

A machine learning formulation

Given:

- a set of documents $\{d_1, d_2, \ldots, d_m\}$
- a set of labels (classes) $\{y_1, y_2, \ldots, y_n\}$
- a training set of labeled documents
  $\{(d_1, y_{d_1}), \ldots, (d_m, y_{d_m})\}$

produce:

- a classifier $h(d) = y$

How should we represent d?

Bag-of-words representation

A vector of word counts is called a *bag of words*.

e.g 2 occurences of "man"
3 occurences of "dog" $(2, 3, 1, 0, 0, 0, 1, \ldots, 0, 1)$

The vector has length n = the number of unique words in the training set.

Note that word order is not represented!

Bag-of-words representation

You have the following corpus. Create word vectors for all documents!

to be is to do	Sartre
to do is to be	Kant
do be do be do	Sinatra
to be or not to be	Shakespeare
I do I do I do I do I do	ABBA

Bag-of-words representation

You have the following corpus. Create word vectors for all documents!

	to	be	is	do	or	not	I

to be is to do	(2,1,1,1,0,0,0)
to do is to be	(2,1,1,1,0,0,0)
do be do be do	(0,2,0,3,0,0,0)
to be or not to be	(2,2,0,0,1,1,0)
I do I do I do I do I do	(0,0,0,5,0,0,5)

Comparing documents

Through the bag-of-word representation, documents can be seen as *points* or *vectors* in a multi-dimensional space.

ham

spam

ham

want to find the label of this document

can draw in more than 2 dimensions.

?

spam

ham
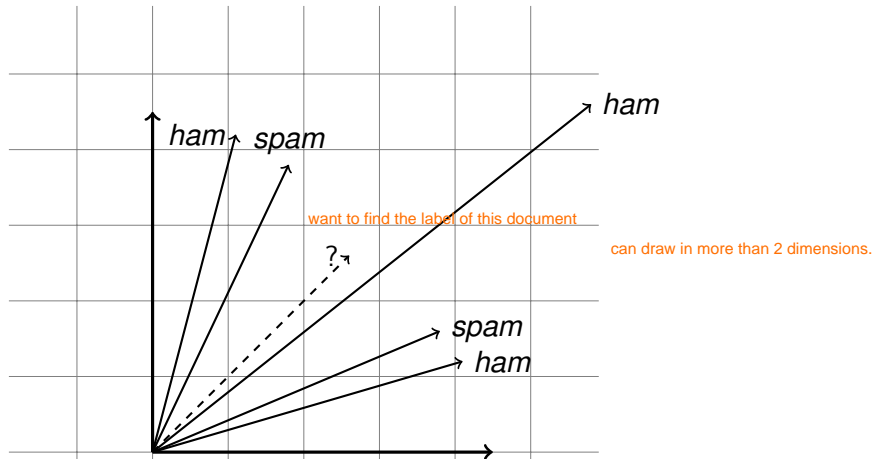
Comparing vectors

We can compare our documents vectors in different ways.

Euclidean distance:

$$\|(x - y)\|_2 = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Manhattan distance:

$$\|(x - y)\|_1 = \sum_{i=1}^{n}|x_i - y_i|$$

Cosine similarity :

can see it as a measure of differences in angle.
Smaller angle -> More similar

$$\cos(x, y) = \frac{x \cdot y}{\|x\|\|y\|}$$

more fruitful to compare similarity than differences
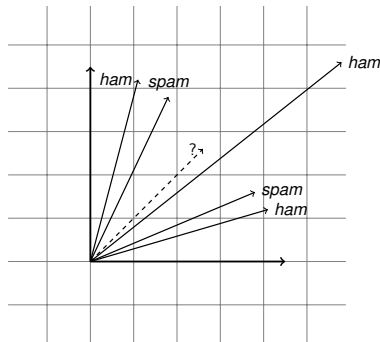
will get a number between 0 and 1.

Comparing vectors

$\cos(\text{Sartre}, \text{Kant}) =$

$\cos((2, 1, 1, 1, 0, 0, 0), (2, 1, 1, 1, 0, 0, 0)) =$

$\frac{2 \cdot 2 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 0 + 0 \cdot 0}{\sqrt{7}\sqrt{7}} =$

1

$\cos(\text{Sartre}, \text{Sinatra}) =$

$\cos((2, 1, 1, 1, 0, 0, 0), (0, 2, 0, 3, 0, 0, 0)) =$

$\frac{2 \cdot 0 + 1 \cdot 2 + 1 \cdot 0 + 1 \cdot 3 + 0 \cdot 0 + 0 \cdot 0 + 0 \cdot 0}{\sqrt{7}\sqrt{13}} =$

0.52

$\cos(\text{Shakespeare}, \text{ABBA}) =$

$\cos((2, 2, 0, 0, 1, 1, 0), (0, 0, 0, 5, 0, 0, 5)) =$

$\frac{2 \cdot 0 + 2 \cdot 0 + 0 \cdot 0 + 0 \cdot 5 + 1 \cdot 0 + 1 \cdot 0 + 0 \cdot 5}{\sqrt{10}\sqrt{50}} =$

0

k-nearest neighbour

Using cosine similarity, we have:

- 1-nearest neighbour ⇒ ?=*ham*
- 3-nearest neighbour ⇒ ?=*spam*
- 5-nearest neighbour ⇒ ?=*ham*

Dimensionality reduction

Do we have to consider all words?

No! Usually it is beneficial to remove so-called very common words (*stop words*). doesn't add very much to semantics

These can be defined *grammatically* (pronouns, prepositions, determiners, ...), or *statistically* (words that are common in the corpus).

To Sherlock Holmes she is always *the* woman. I have seldom heard him mention her under any other name. In his eyes she eclipses and predominates the whole of her sex. It was not that he felt any emotion akin to love for Irene Adler. All emotions, and that one particularly, were abhorrent to his cold, precise but admirably balanced mind. He was, I take it, the most perfect reasoning and observing machine that the world has seen, but as a lover he would have placed himself in a false position. He never spoke of the softer passions, save with a gibe and a sneer. They were admirable things for the observer—excellent for drawing the veil from men's motives and actions. But for the trained reasoner to admit such intrusions into his own delicate and finely adjusted temperament was to introduce a distracting factor which might throw a doubt upon all his mental results. Grit in a sensitive instrument, or a crack in one of his own high-power lenses, would not be more disturbing than a strong emotion in a nature such as his. And yet there was but one woman to him, and that woman was the late Irene Adler, of dubious and questionable memory.

To Sherlock Holmes she is always *the* woman. I have seldom heard him mention her under any other name. In his eyes she eclipses and predominates the whole of her sex. It was not that he felt any emotion akin to love for Irene Adler. All emotions, and that one particularly, were abhorrent to his cold, precise but admirably balanced mind. He was, I take it, the most perfect reasoning and observing machine that the world has seen, but as a lover he would have placed himself in a false position. He never spoke of the softer passions, save with a gibe and a sneer. They were admirable things for the observer—excellent for drawing the veil from men's motives and actions. But for the trained reasoner to admit such intrusions into his own delicate and finely adjusted temperament was to introduce a distracting factor which might throw a doubt upon all his mental results. Grit in a sensitive instrument, or a crack in one of his own high-power lenses, would not be more disturbing than a strong emotion in a nature such as his. And yet there was but one woman to him, and that woman was the late Irene Adler, of dubious and questionable memory.
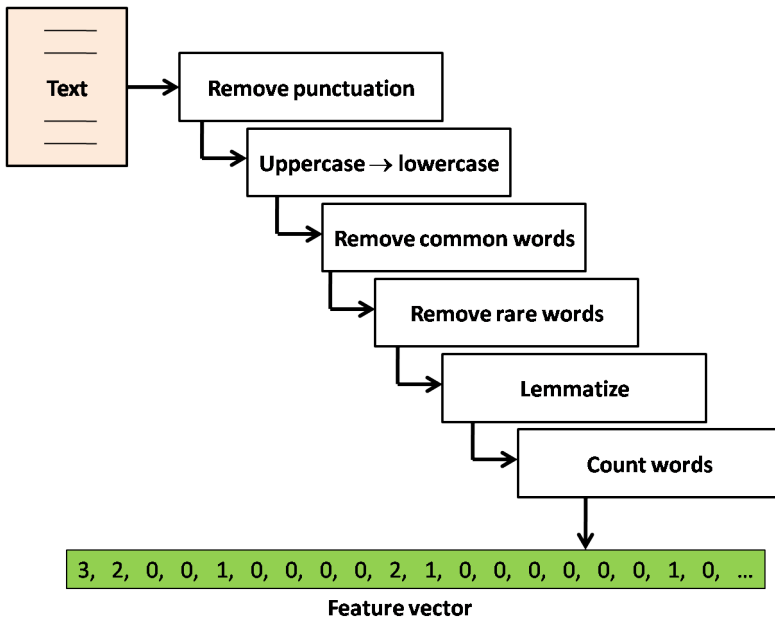
| 10 a |
| 9 the |
| 9 and |
| 6 to |
| 6 his |
| 5 was |
| 5 in |
| 4 that |
| 4 of |
| 4 he |
| 4 for |
| 4 but |
| 3 woman |
| 3 one |
| 2 would |
| 2 were |
| 2 such |
| 2 she |
| 2 own |
| 2 not |

To **Sherlock Holmes** she is **always** *the* **woman**. I have **seldom heard** him **mention** her under any other **name**. In his **eyes** she eclipses and **predominates** the **whole** of her **sex**. It was not that he **felt** any **emotion akin** to **love** for **Irene Adler**. All **emotions**, and that one **particularly**, were **abhorrent** to his **cold, precise** but **admirably balanced mind**. He was, I **take** it, the **most perfect reasoning** and **observing machine** that the **world** has **seen**, but as a **lover** he would have **placed himself** in a **false position**. He **never spoke** of the **softer passions, save** with a **gibe** and a **sneer**. They were **admirable things** for the **observer**—**excellent** for **drawing** the **veil** from **men's motives** and **actions**. But for the **trained reasoner** to **admit such intrusions** into his **own delicate** and **finely adjusted temperament** was to **introduce** a **distracting factor** which might **throw** a **doubt** upon all his **mental results**. **Grit** in a **sensitive instrument**, or a **crack** in one of his **own high-power lenses**, would not be **more disturbing** than a **strong emotion** in a **nature** such as his. And **yet** there was but one **woman** to him, and that **woman** was the **late Irene Adler**, of **dubious** and **questionable memory**.

3 woman
2 own
2 irene
2 emotion
2 adler
1 yet
1 world
1 whole
1 veil
1 trained
1 throw
1 things
1 the
1 temperament
1 take
1 such
1 strong
1 spoke
1 softer
1 sneer

To **Sherlock Holmes** she is **always** *the* **woman**. I have **seldom** hear him **mention** her under any other **name**. In his **eye** she eclipse and **predominate** the **whole** of her **sex**. It was not that he **feel** any **emotion akin** to **love** for **Irene Adler**. All **emotion** , and that one **particularly**, were **abhorrent** to his **cold**, **precise** but **admirably balanced mind**. He was, I **take** it, the **most perfect reasoning** and **observing machine** that the **world** has see , but as a **lover** he would have **place** himself in a **false position**. He **never speak** of the **soft passion** , **save** with a **gibe** and a **sneer**. They were **admirable things** for the **observer**— **excellent** for **draw** the **veil** from **man motive** and **action** . But for the **train reasoner** to **admit such intrusion** into his **own delicate** and **finely adjust temperament** was to **introduce** a **distracting factor** which might **throw** a **doubt** upon all his **mental result** . **Grit** in a **sensitive instrument**, or a **crack** in one of his **own high-power lens** , would not be **more disturbing** than a **strong emotion** in a **nature** such as his. And **yet** there was but one **woman** to him, and that **woman** was the **late Irene Adler**, of **dubious** and **questionable memory**.
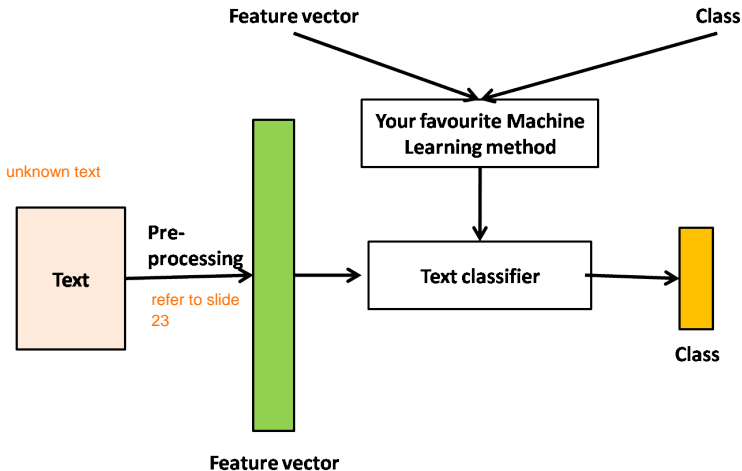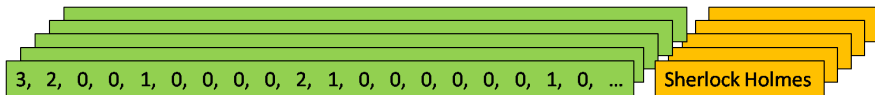
3 woman
2 own
2 irene
2 emotion
2 adler
1 yet
1 world
1 whole
1 veil
1 train
1 throw
1 things
1 the
1 temperament
1 take
1 such
1 strong
1 speak
1 soft
1 sneer

Text → Remove punctuation → Uppercase → lowercase → Remove common words → Remove rare words → Lemmatize → Count words

3, 2, 0, 0, 1, 0, 0, 0, 0, 2, 1, 0, 0, 0, 0, 0, 0, 1, 0, ...

Feature vector

3, 2, 0, 0, 1, 0, 0, 0, 0, 2, 1, 0, 0, 0, 0, 0, 0, 1, 0, ...

Sherlock Holmes

Feature vector

Class

Your favourite Machine Learning method

unknown text

Text

Pre-processing

refer to slide 23

Feature vector

Text classifier

Class

Feature vector
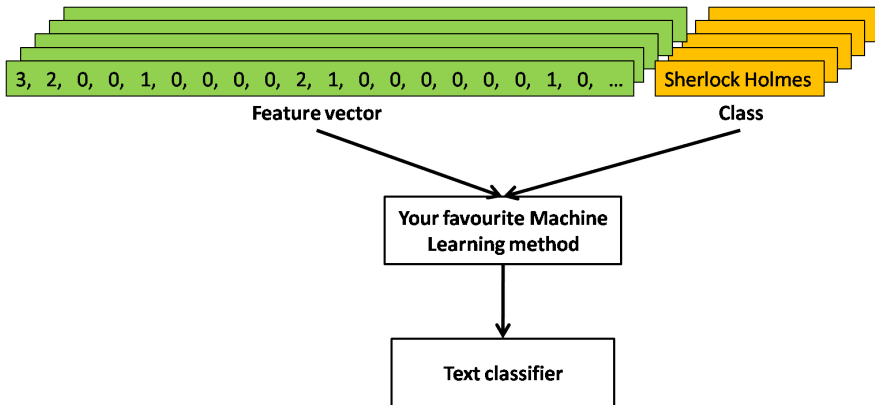
More dimensionality reduction

- Remove stop words
- Remove rare words
- Case folding turn everything into lower case
- Remove diacritics
- Spell checking is a dimensionality reduction technique. Mispelling creates a lot of version and hence vectors of word
- Lemmatization
- Replace numbers with `<number>` special tag
- Replace mail adresses with `<mail>`, web addresses with `<web>`, etc.
- Normalize punctuation (e.g. " becomes ")

There are also more advanced dimensionality reduction methods. More about them later in the course.

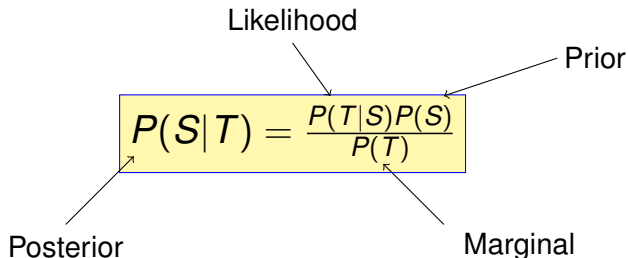# DD2418 Language Engineering
5b: Naive Bayes text classification

Johan Boye, KTH

Bayes' theorem expresses how beliefs should be updated in the light of new evidence.

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)}$$

Likelihood

Prior

Posterior

Marginal

Naive Bayes classifier

Given a document with words $w_1 \ldots w_n$, what is the most likely label y?

$$\arg \max_y P(y|w_1 \ldots w_n)$$ want to find the class label y that maximize[the probability

Rewrite and simplify:

$$\arg \max_y P(y|w_1 \ldots w_n) = \swarrow \text{Bayes' theorem}$$

$$\arg \max_y \frac{P(w_1 \ldots w_n|y)P(y)}{P(w_1 \ldots w_n)} =$$

$$\arg \max_y P(w_1 \ldots w_n|y)P(y) = \swarrow \text{conditional independence assumption}$$

$$\arg \max_y P(w_1|y) \ldots P(w_n|y)P(y)$$

Parameter estimation

Maximum likelihood estimation:

$$P(y) = \frac{\text{\# of documents of label } y}{\text{total number of documents}}$$

$$P(w_i|y) = \frac{\text{\# of times } w_i \text{ appears in } y}{\text{\# of tokens in all } y\text{-documents}}$$

Example

2 labels: $+$, $-$.
4 documents: A, B, C, D.

	cold	cough	fever	flu	influenza	pneumonia	temperature	
A	0	1	1	1	0	0	1	+
B	0	1	0	1	0	1	2	−
C	0	1	0	2	0	0	1	+
D	1	0	0	1	0	0	0	+

Construct a Naive Bayes classifier, and use it to classify the document "*temperature*".

Example

2 labels: $+, -$.
4 documents: A, B, C, D.

	cold	cough	fever	flu	influenza	pneumonia	temperature	
A	0	1	1	1	0	0	1	+
B	0	1	0	1	0	1	2	−
C	0	1	0	2	0	0	1	+
D	1	0	0	1	0	0	0	+

$P(+) = \dfrac{3}{4}$, $P(-) = \dfrac{1}{4}$, $P(\text{temperature}|+) = \dfrac{2}{10}$, $P(\text{temperature}|-) = \dfrac{2}{5}$

Predicted label

$$P(+|\text{temperature}) = \frac{1}{Z}P(+)P(\text{temperature}|+) =$$
$$\frac{1}{Z} \cdot \frac{3}{4} \cdot \frac{2}{10} = \frac{1}{Z} \cdot \underline{0.15}$$

$$P(-|\text{temperature}) = \frac{1}{Z}P(-)P(\text{temperature}|-) =$$
$$\frac{1}{Z} \cdot \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{Z} \cdot \underline{0.1}$$

where $Z = P(\text{temperature})$, which can be ignored (why?).

The predicted label is the one with the highest probability:

$$\arg\max_{y \in \{+,-\}} P(y|\text{temperature}) = +$$

Another example

2 labels: $+$, $-$.
4 documents: A, B, C, D.

	cold	cough	fever	flu	influenza	pneumonia	temperature	
A	0	1	1	1	0	0	1	+
B	0	1	0	1	0	1	2	−
C	0	1	0	2	0	0	1	+
D	1	0	0	1	0	0	0	+

Construct a Naive Bayes classifier, and use it to classify the
document "*flu influenza*".

	cold	cough	fever	flu	influenza	pneumonia	temperature	
A	0	1	1	1	0	0	1	+
B	0	1	0	1	0	1	2	−
C	0	1	0	2	0	0	1	+
D	1	0	0	1	0	0	0	+

$$P(+) = \frac{3}{4}, \ P(-) = \frac{1}{4}, \ P(\text{flu}|+) = \frac{4}{10}, \ P(\text{flu}|-) = \frac{1}{5}$$

However:

$$P(\text{influenza}|+) = P(\text{influenza}|-) = 0$$

Laplace smoothing

	cold	cough	fever	flu	influenza	pneumonia	temperature	
A	0	1	1	1	0	0	1	+
B	0	1	0	1	0	1	2	−
C	0	1	0	2	0	0	1	+
D	1	0	0	1	0	0	0	+

Laplace smoothing: Add 1 occurrence of every word to every label

$$P(+)P(\text{flu}|+)P(\text{influenza}|+) = \frac{3}{4} \cdot \frac{4+1}{10+7} \cdot \frac{1}{10+7} = \underline{0.013}$$

$$P(-)P(\text{flu}|-)P(\text{influenza}|-) = \frac{1}{4} \cdot \frac{1+1}{5+7} \cdot \frac{1}{5+7} = \underline{0.003}$$

Predicted label

$$P(+|\text{flu influenza}) = \frac{1}{Z}P(+)P(\text{flu}|+)P(\text{influenza}|+) =$$
$$\frac{1}{Z} \cdot \frac{3}{4} \cdot \frac{4+1}{10+7} \cdot \frac{1}{10+7} = \frac{1}{Z} \cdot \underline{0.013}$$

$$P(-|\text{flu influenza}) = \frac{1}{Z}P(-)P(\text{flu}|-)P(\text{influenza}|-) =$$
$$\frac{1}{Z} \cdot \frac{1}{4} \cdot \frac{1+1}{5+7} \cdot \frac{1}{5+7} = \frac{1}{Z} \cdot \underline{0.003}$$

where $Z = P(\text{flu influenza})$, which can be ignored (why?).

The predicted label is the one with the highest probability:

$$\arg\max_{y \in \{+,-\}} P(y|\text{flu influenza}) = +$$

Laplace smoothing again

The extra smoothing words are sometimes called *pseudowords*.

Pseudowords don't have to be integers; one might e.g. add 0.01 pseudowords per class.

In our example:

$$P(+)P(\text{flu}|+)P(\text{influenza}|+) = \frac{3}{4} \cdot \frac{4+0.01}{10+0.07} \cdot \frac{0.01}{10+0.07} = \underline{0.0003}$$

$$P(-)P(\text{flu}|-)P(\text{influenza}|-) = \frac{1}{4} \cdot \frac{1+0.01}{5+0.07} \cdot \frac{0.01}{5+0.07} = \underline{0.00001}$$

At test time, one might encounter words that do not appear in the training data.
(Many ML packages do not allow this.)

Recommended approach: Ignore those unknown words.

will be equally unknown for each class

DD2418 Language Engineering
5c: Evaluation of text classification

Johan Boye, KTH

Evaluation

Evaluation involves testing the hypothesis h on a development set $\{(x_1, y_1), \ldots, (x_k, y_k)\}$.

Confusion matrix:

		Real label y	
		spam	ham
$h(x)$	spam	102	48
	ham	27	323

Positive examples (for *spam*): 102+48 = 150
Negative examples: 27+323 = 350
True positives: 102
False positives: 48
True negatives: 323
False negatives: 27

Evaluation metrics

Accuracy (binary case):

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy (generally):

$$\frac{\text{correctly classified examples}}{\text{all examples}}$$

Precision:

$$\frac{TP}{TP + FP}$$

Recall:

$$\frac{TP}{TP + FN}$$

Evaluation involves testing the hypothesis h on a development set $\{(x_1, y_1), \ldots, (x_k, y_k)\}$.

Confusion matrix:

		Real label y	
		spam	ham
$h(x)$	spam	102	48
	ham	27	323

Compute accuracy of the classifier, as well as precision and recall (for the spam class).

Evaluation involves testing the hypothesis *h* on a development set $\{(x_1, y_1), \ldots, (x_k, y_k)\}$.

Confusion matrix:

		Real label *y*	
		spam	*ham*
$h(x)$	*spam*	102	48
	ham	27	323

Accuracy: (102+323) / (102+323+48+27) = .85
Precision: 102/(102+48) = .68
Recall: 102/(102+27) = .79

Precision-recall trade-off

What is most important, precision or recall?

(Arithmetic) average precision/recall:

$$\frac{P + R}{2}$$

More useful: Geometric mean (F-score):

$$\frac{1}{\frac{1}{2}(\frac{1}{P} + \frac{1}{R})} = \frac{2PR}{P + R}$$

Precision-recall trade-off

(Arithmetic) average precision/recall:

$$\frac{P + R}{2}$$

If $P = R = 0.5$, then $\frac{P+R}{2} = 0.5$.

If $P = 0.01$ and $R = 0.99$, then $\frac{P+R}{2} = 0.5$.

More useful: Geometric mean (F-score):

$$\frac{1}{\frac{1}{2}(\frac{1}{P} + \frac{1}{R})} = \frac{2PR}{P + R}$$

If $P = R = 0.5$, then $\frac{2PR}{P+R} = 0.5$.

If $P = 0.01$ and $R = 0.99$, then $\frac{2PR}{P+R} = 0.0198$.

Non-binary case

Classify news as *Sports, Politics*, or *Economy*.

x	h(x)	y
1	S	S
2	P	S
3	E	S
4	S	S
5	P	P
6	E	P
7	S	P
8	P	P
9	E	E
10	S	E

Confusion matrix

		Real label *y*		
		S	P	E
h(x)	S	2	1	1
	P	1	2	0
	E	1	1	1

Given the confusion matrix on the previous slide, compute
these quantities:

Accuracy   _____
Precision, class S _____
Precision, class P _____
Precision, class E _____
Recall, class S _____
Recall, class P _____
Recall, class E _____

Given the confusion matrix on the previous slide, compute these quantities:

Accuracy	0.5
Precision, class S	0.5
Precision, class P	0.67
Precision, class E	0.33
Recall, class S	0.5
Recall, class P	0.5
Recall, class E	0.5

Evaluation methodology

3 sets: *training, development* and *test* sets.

The three sets should all be distinct!

- Otherwise the model will be *over-trained*.
- Results will be unrealistically good.
- Model will perform poorly on new data.

First, set aside 10% of the data as the final test set.

Then train on 80% of the data, evaluate on 10%. Iterate.
validation set

To minimize random effects, *n-fold cross-validation* can be used.

Finally, evaluate on the test set. This is **the** result.

Hyperparameters

Typically, a machine learning method has a number of hyperparameters that influence the result.

- e.g. the number of layers in a neural network, and the number of neurons in each layer

Typical choices in a language engineering setting:

- Case folding, yes/no?
- Lemmatization, yes/no?
- Remove certain words (stop words)?
- Context window size?
- Normalization of numbers, years, dates, ...?
- Tokenization, what is a word?
- ...

How good is 80% accuracy?

If your data contains 90% ham and 10% spam, then a classifier returning "*ham*" all the time has 90% accuracy.

There might be other obvious base lines (e.g. check for the term "*weight loss*"), which beats the random baseline.

Make sure that you compare your results to baseline results!

- ... otherwise you risk fooling yourself!

Gold standard

The *gold standard* is the correct labeling according to a human expert.

Very useful, but can be hard/expensive to obtain.