Getting started on simple Sentiment Analysis

Natalie Schluter

Objectives

- Be able to understand the mini-project:
  - What is (automatic) sentiment analysis?
    - Context of *Natural Language Processing*
  - What is machine learning?
  - How does the perceptron algorithm work?
  - What are word embeddings?

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011!

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

→ Bram Stoker

Machine Translation

Killing Palestinians and wounding nine in the raids Sector
Nine Palestinians were wounded among civilians in an Israeli air raid in the neighborhood result in the Gaza Strip. This comes immediately after the killing of two prominent Al-Aqsa Martyrs Brigades in the Israeli occupying forces carried out air and infantry forces in the Balata camp in the West Bank.

Bashir meets Fraser, the Security Council will not impose forces Darfur
Is scheduled to meet with Sudanese President Omar al-Bashir Jenday Fraser Assistant Minister for Foreign Affairs of the American attempt to persuade officials in Khartoum, Sudanese Darfur deployment of the nationalities. For his part, US Ambassador to the United Nations that it has no intention of the Security Council to impose its forces in the province.

Rmsfield and Cheney insist on keeping the American forces in Iraq
Called American Defense Minister Donald Rmsfield Americans to show patience on Iraq. I take Vice President Dick Cheney calls Democrats withdrawal of American forces from Iraq link and the possibility of early withdrawal of attacks inside the United States.

Killing civilians and wounding officer suicide attack in Afghanistan
The international force to help establish security (ISAF) killed civilians and the wounding of an officer in an attack against Afghan forces convoy south Atlantic Afghanistan. In the capital Kabul, a hand grenade exploded at the passage of manufacture French patrol was not reported injuries or damage.

Information Extraction

Event: Curriculum mtg
Date: Jan-16-2017
Start: 10:00am
End: 11:30am
Where: 4A09

Subject: **curriculum meeting**
Date: January 15, 2017

To: Natalie Schluter

Hi Natalie, we've now scheduled the curriculum meeting.
It will be in 4A09 tomorrow from 10:00-11:30.
-Mette

Create new Calendar entry

5

Information Extraction & Sentiment Analysis

Attributes:
zoom
affordability
size and weight
flash
ease of use

Size and weight

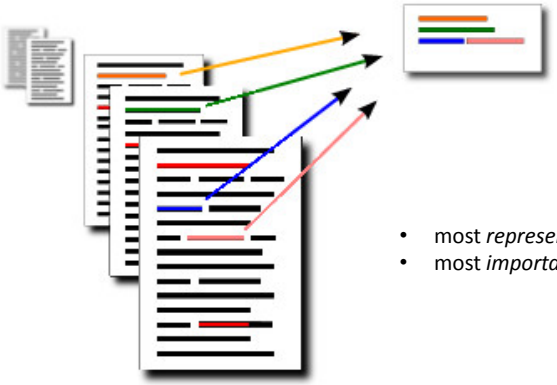
✓ • nice and compact to carry!

✓ • since the camera is small and light, I won't nee e heavy, bulky professional cameras either!

✗ • the camera feels flimsy, is plastic and very light be very delicate in the handling of this camera

6

Automatic Summarisation



- most *representative*
- most *important*

22-11-2018 · 7

Query-ese to Question Answering

Language Technology

mostly solved

Spam detection
Let's go to Agra! ✓
Buy V1AGRA … ✗

Part-of-speech (POS) tagging
ADJ ADJ NOUN VERB ADV
Colorless green ideas sleep furiously.

Named entity recognition (NER)
PERSON ORG LOC
Einstein met with UN officials in Princeton

making good progress

Sentiment analysis
Best roast chicken in San Francisco!
The waiter ignored us for 20 minutes.

Coreference resolution
Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)
I need new batteries for my *mouse*.

Parsing
I can see Alcatraz from the window!

Machine translation (MT)
第13届上海国际电影节开幕...
The 13th Shanghai International Film Festival...

Information extraction (IE)
You're invited to our dinner party, Friday May 27 at 8:30
Party May 27 add

still really hard

Question answering (QA)
Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase
XYZ acquired ABC yesterday
ABC has been taken over by XYZ

Summarization
The Dow Jones is up
The S&P500 jumped
Housing prices rose
→ Economy is good

Dialog
Where is Citizen Kane playing in SF?
Castro Theatre at 7:30. Do you want a ticket?

Commercial World

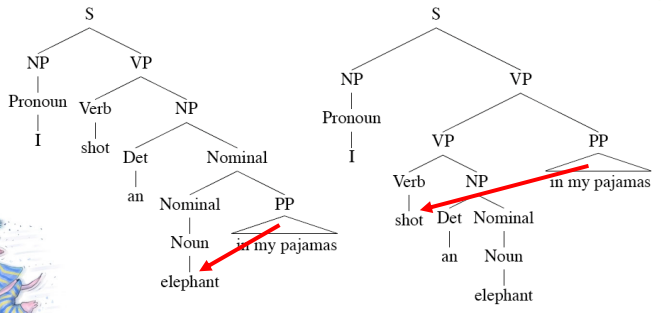
- Lot's of exciting stuff going on…

PEARSON Knowledge Technologies
Siri
Microsoft
Google
YAHOO!
amazon echo
Always ready, connected, and fast. Just ask.
Visible
Deeper Insights. Driving Results.
ORACLE | collective intellect
IBM Watson
BETA hakia search for meaning
J.D.POWER AND ASSOCIATES
Bloomberg Technology

Ambiguity

"One morning I shot an elephant in my pyjamas.
How he got into my pyjamas I don't know."
Groucho Marx, Animal Crackers, 1930

Ambiguity



Ambiguity Examples

Attachment ambiguity:

ex. We picked up John on the way to school.

Coordination ambiguity:

ex. There were fairy princesses and pirates at the kid's party.

ex. Only rich men and women are invited.

Ambiguity makes NLP hard: "Crash blossoms"

Violinist Linked to JAL Crash Blossoms
Teacher Strikes Idle Kids
Red Tape Holds Up New Bridges
Hospitals Are Sued by 7 Foot Doctors
Juvenile Court to Try Shooting Defendant
Local High School Dropouts Cut in Half

Why else is natural language understanding difficult?

non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

segmentation issues

the New York-New Haven Railroad
the New York-New Haven Railroad

idioms

dark horse
get cold feet
lose face
throw in the towel

neologisms

unfriend
Retweet
bromance

world knowledge

Mary and Sue are sisters.
Mary and Sue are mothers.

tricky entity names

Where is *A Bug's Life* playing …
Let It Be was recorded …
… a mutation on the *for* gene …

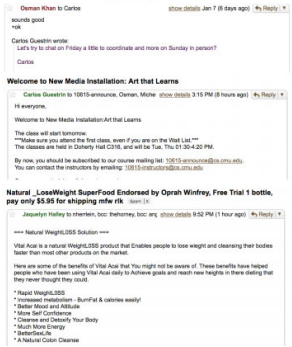
Making progress on this problem…

- The task is difficult! What tools do we need?
  - Knowledge about language
  - Knowledge about the world
  - A way to combine knowledge sources
- How we generally do this:
  - probabilistic models built from language data
    - P("maison" → "house") high
    - P("L'avocat général" → "the general avocado") low
  - Luckily, rough text features can often do half the job.

Introduction to Machine Learning

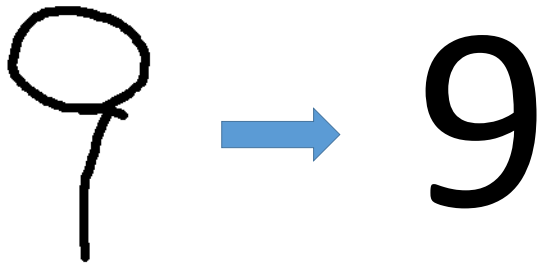
Spam filtering

data **prediction**



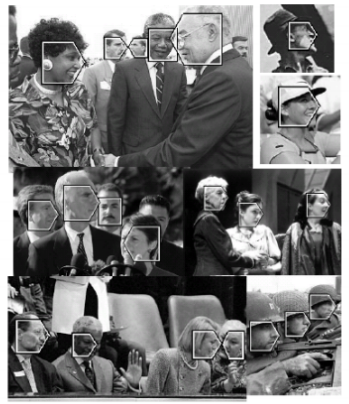
Spam
vs.
Not Spam

Digit recognition



Face recognition

Example training images
for each orientation



Weather prediction

Supervised Learning: find f

- Given: Training set $\{(x_i, y_i) \mid i = 1 \ldots N\}$
- Find: A good approximation to $f : X \rightarrow Y$

Examples: what are X and Y?

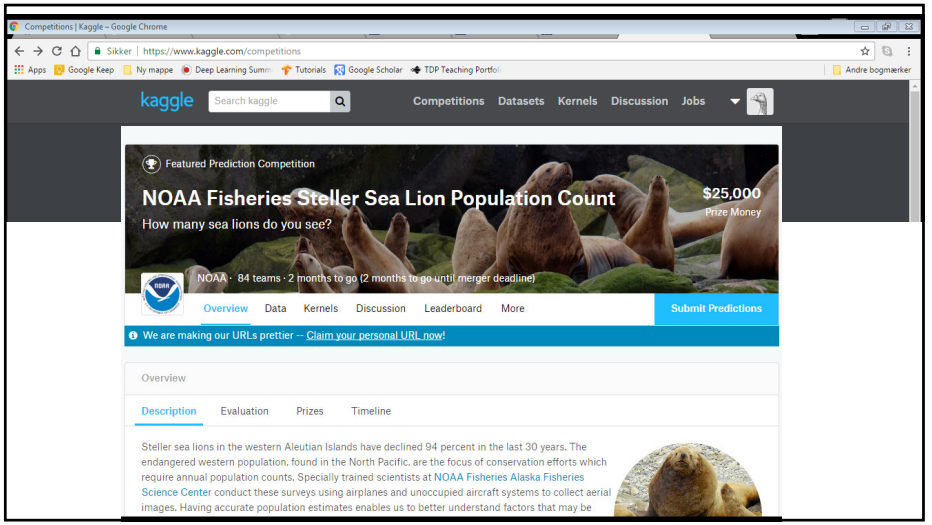
- Spam Detection
  - Map email to {Spam, Not Spam}
- Digit recognition — Classification
  - Map pixels to {0,1,2,3,4,5,6,7,8,9}
- Stock Prediction
  - Map new, historic prices, etc. to $\Re$ (the real numbers) — Regression

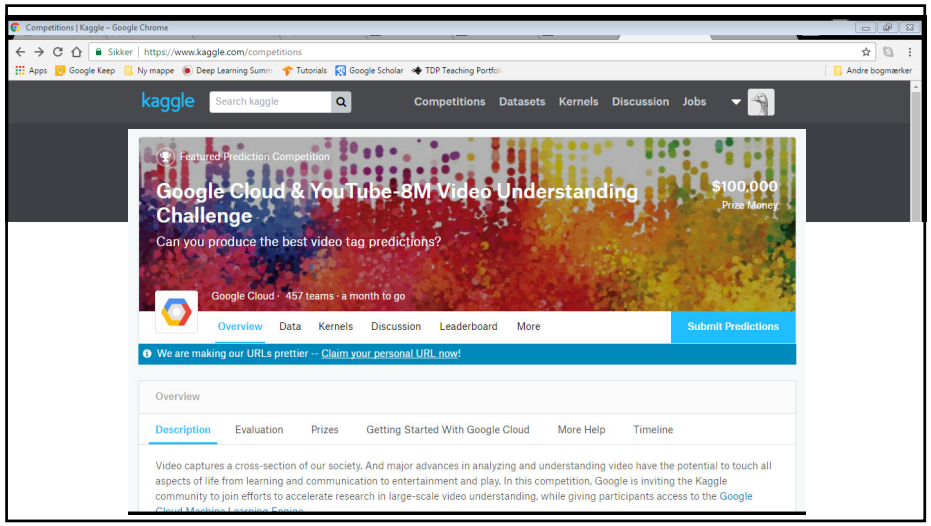
A Supervised Learning Problem

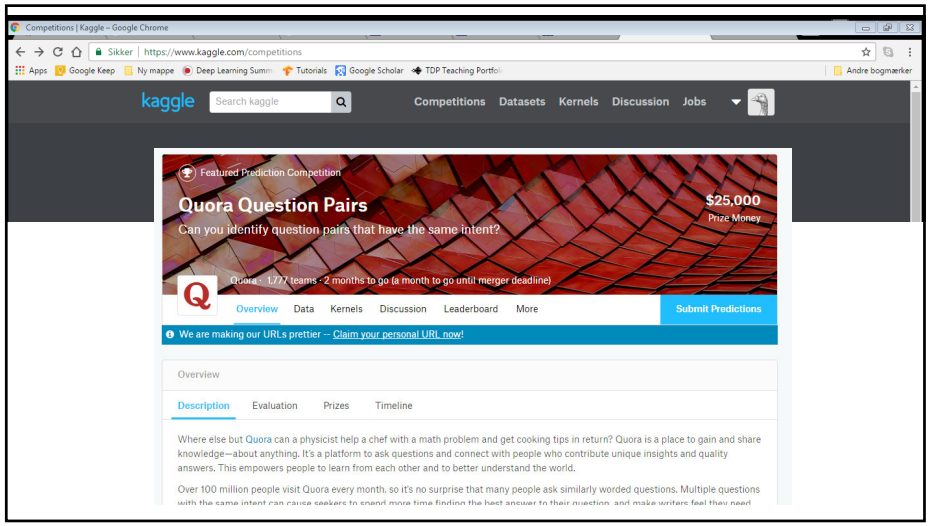
Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

- Our goal is to find a function $f : X \rightarrow Y$
  - $X = \{0,1\}^4$
  - $Y = \{0,1\}$
- Question 1: How should we pick the *hypothesis space*, the set of possible functions f?
- Question 2: How do we find the best f in the hypothesis space?







What is Machine Learning?

Positive or negative movie review?

- “unbelievably disappointing”
- “Full of zany characters and richly applied satire, and some great plot twists”
- “this is the greatest screwball comedy ever filmed”
- “It was pathetic. The worst part about it was the boxing scenes.”

28

Sentiment Analysis

Google Shopping

HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner
$89 online, **$100** nearby ★★★★☆ 377 reviews
September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sh

Reviews

Summary - Based on 377 reviews

1 star	2	3	4 stars	5 stars

What people are saying

ease of use	"This was very easy to setup to four computers."
value	"Appreciate good quality at a fair price."
setup	"Overall pretty easy setup."
customer service	"I DO like honest tech support people."
size	"Pretty Paper weight."
mode	"Photos were fair on the high quality mode."
colors	"Full color prints came out with great quality."

Amazon.com

The History of the Decline and Fall of the Roman Empire: v. 1 (Penguin Classics) Paperback – 7 Mar 1996
by Edward Gibbon (Author), David Womersley (Editor, Introduction)
★★★★½ 35 customer reviews

See all 187 formats and editions

Kindle Edition	Hardcover	Paperback	Audio CD
£0.00 kindleunlimited	£21.95	£19.99	£48.86
This title and over 1 million more are available with Kindle Unlimited £0.00 to buy	4 Used from £40.66 10 New from £21.95	16 Used from £9.49 31 New from £8.88 1 Collectible from £11.91	2 Used from £54.43 3 New from £24.03

Want it delivered by Friday, 21 Apr.? Order within 13 hrs 11 mins and choose **One-Day Delivery** at checkout. Details

Note: This item is eligible for **click and collect.** Details

Edward Gibbon's six-volume History of the Decline and Fall of the Roman Empire (1776-88) is among the most magnificent and ambitious narratives in European literature. Its subject is the fate of one of the world's greatest civilizations over thirteen centuries - its rulers, wars and society, and the events that led to its disastrous collapse. Here, in volumes one and two, Gibbon charts the vast extent and constitution of the Empire from the reign of Augustus to 395 ad. And in a controversial critique, he

Read more

See all 3 images

Most Recent Customer Reviews

★★★★★ **Five Stars**
Very deep read. The language is very dated but that makes the book for me.
Published 3 months ago by jay long

★★★★★ **An ideal gift for the amateur historian**
I bought this as a present for my father, who was very pleased to receive it.
Published 10 months ago by Deborah T K

★★★★★ **Five Stars**
Have not had a chance to read it yet, but is a classic
Published 10 months ago by SunnyAvalon

★☆☆☆☆ **History**
Very dated in its style, if you want to learn about Rome read one of the more modern historians such as Mary beard, this book is ready for the museum
Published 11 months ago by Super shopper

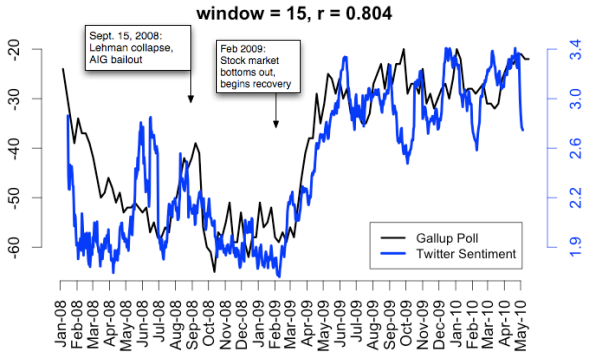
★★★★★ **Five Stars**
Informative.
Published 11 months ago by Kristian Myer

★★★★★ **Five Stars**
informative
Published 15 months ago by David I.

31

Twitter sentiment versus Gallup Poll of Consumer Confidence

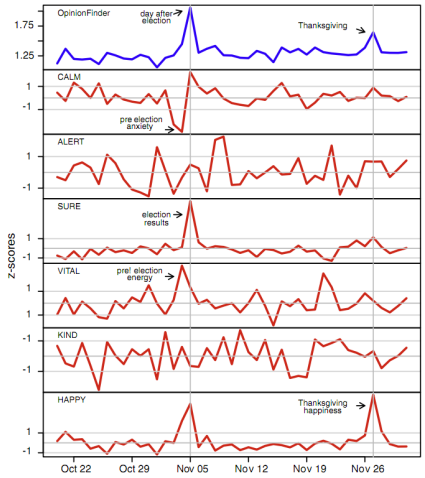
Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In ICWSM-2010



Twitter sentiment:

Johan Bollen, Huina Mao, Xiaojun Zeng. 2011. Twitter mood predicts the stock market,

Journal of Computational Science 2:1, 1-8. 10.1016/j.jocs.2010.12.007.

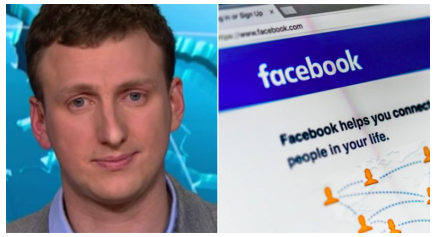


Facebook sentiment: what not to do

Facebook emotion study breached ethical guidelines, researchers say

Lack of 'informed consent' means that Facebook experiment on nearly 700,000 news feeds broke rules on tests on human subjects, say scientists

Poll: Facebook's secret mood experiment: have you lost trust in the social network?



Sentiment analysis has many other names

- Opinion extraction
- Opinion mining
- Sentiment mining
- Subjectivity analysis

35

Why sentiment analysis?

- *Movie*: is this review positive or negative?
- *Products*: what do people think about the new iPhone?
- *Public sentiment*: how is consumer confidence? Is despair increasing?
- *Politics*: what do people think about this candidate or issue?
- *Prediction*: predict election outcomes or market trends from sentiment

36

Scherer Typology of Affective States

- **Emotion**: brief organically synchronized … evaluation of a major event
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons**
  - *liking, loving, hating, valuing, desiring*
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious, reckless, morose, hostile, jealous*

Sentiment Analysis

- Sentiment analysis is the detection of **attitudes**

  "enduring, affectively colored beliefs, dispositions towards objects or persons"

  1. **Holder (source)** of attitude
  2. **Target (aspect)** of attitude
  3. **Type** of attitude
     - From a set of types
       - *Like, love, hate, value, desire,* etc.
     - Or (more commonly) simple weighted **polarity**:
       - *positive, negative, neutral,* together with *strength*
  4. **Text** containing the attitude
     - Sentence or entire document

38

Sentiment Analysis

- Simplest task:
  - Is the attitude of this text positive or negative?
- More complex:
  - Rank the attitude of this text from 1 to 5
- Advanced:
  - Detect the target, source, or complex attitude types

Sentiment Analysis

- Simplest task:
  - Is the attitude of this text positive or negative?
- More complex:
  - Rank the attitude of this text from 1 to 5
- Advanced:
  - Detect the target, source, or complex attitude types

Problems: What makes reviews hard to classify?

- Subtlety:
  - Perfume review in *Perfumes: the Guide*:
    - "If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut."
  - Dorothy Parker on Katherine Hepburn
    - "She runs the gamut of emotions from A to B"

41

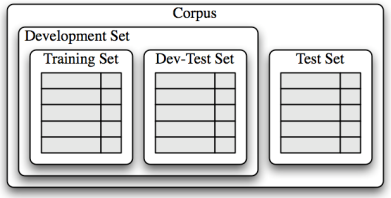
Thwarted Expectations and Ordering Effects

- "This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it **can't hold up**."
- Well as usual Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishbourne is **not so good** either, I was surprised.

42

What is the sentiment analysis task (for us)?

Evaluation

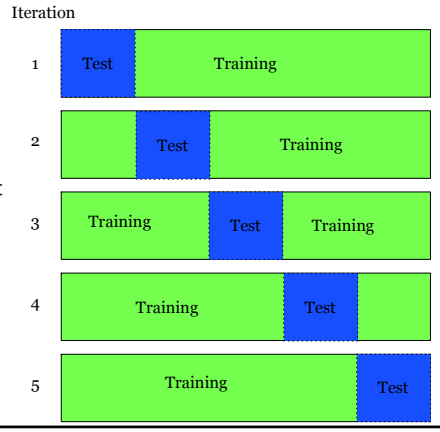


$$Accuracy = \frac{Correct\ Predictions}{All\ Predictions}$$

…over *unseen* data (test set)
…after
- training on the training set and
- model development on a separate holdout set of data (dev set)
- 80/10/10 split

Cross-Validation

- Break up data into 10 folds
  - (Equal positive and negative inside each fold?)
- For each fold
  - Choose the fold as a temporary test set
  - Train on 9 folds, compute performance on the test fold
- Report average performance of the 10 runs

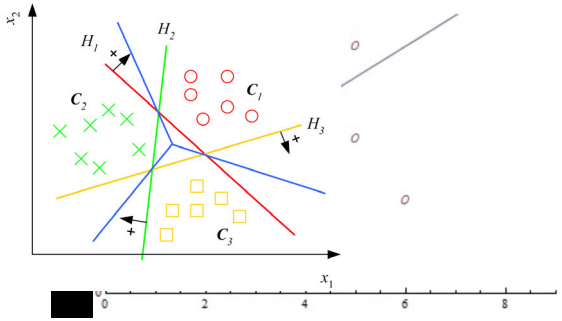


How to evaluate your classifier?

Perceptron learning algorithm

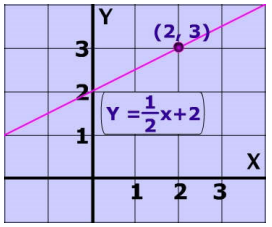
$$w_1x_1 + w_2x_2 + \ldots + w_nx_n > \theta \Rightarrow 1$$

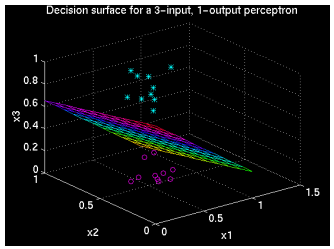
$$w_1x_1 + w_2x_2 + \ldots + w_nx_n \leq \theta \Rightarrow 0$$



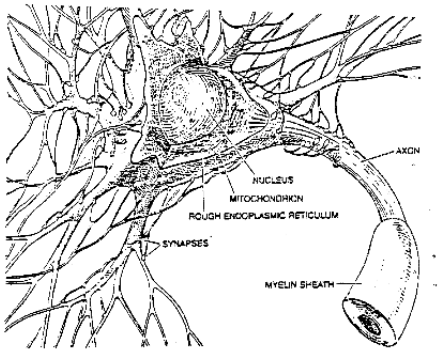
Pop-quiz

- Equation of a line in 2D:
  - With slope m and y-intercept
- 3D, plane?
- N dimensions, hyperplane?

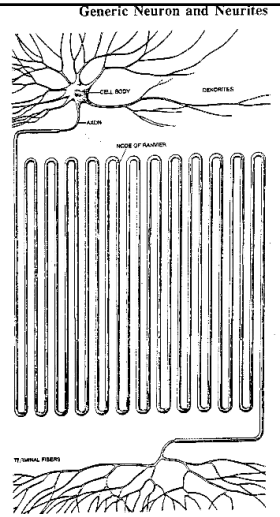




Basic Neuron



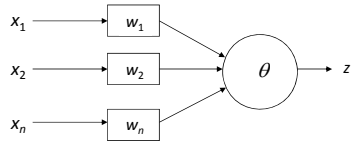
Expanded I



Perceptron Learning Algorithm

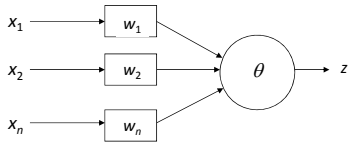
- First neural network learning model in the 1960's
- Simple and limited (single layer models)
- Basic concepts are similar for multi-layer models so this is a good learning tool (= Deep learning!)
- Still used in many current applications (modems, etc.)

Perceptron Node – Threshold Logic Unit



$$z = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i w_i \geq \theta \\ 0 & \text{if } \sum_{i=1}^{n} x_i w_i < \theta \end{cases}$$

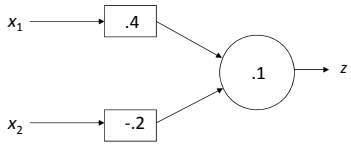
Perceptron Node – Threshold Logic Unit



- Learn weights such that an objective function is maximized.
- What objective function should we use?
- What learning algorithm should we use?

$$z = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i w_i \geq \theta \\ 0 & \text{if } \sum_{i=1}^{n} x_i w_i < \theta \end{cases}$$

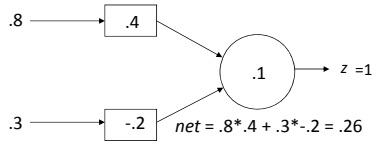
Perceptron Learning Algorithm



x_1	x_2	t
.8	.3	1
.4	.1	0

$$z = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i w_i \geq \theta \\ 0 & \text{if } \sum_{i=1}^{n} x_i w_i < \theta \end{cases}$$

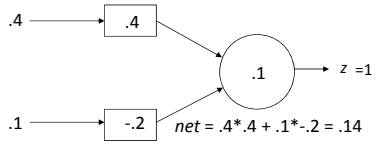
First Training Instance



x_1	x_2	t
.8	.3	1
.4	.1	0

$$z = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i w_i \geq \theta \\ 0 & \text{if } \sum_{i=1}^{n} x_i w_i < \theta \end{cases}$$

Second Training Instance



x_1	x_2	t
.8	.3	1
.4	.1	0

$$z = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i w_i \geq \theta \\ 0 & \text{if } \sum_{i=1}^{n} x_i w_i < \theta \end{cases}$$

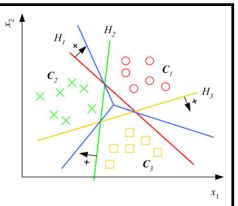
$$\Delta w_i = (t - z) * c * x_i$$

K Outputs

Regression:

$$y_i = \sum_{j=1}^{d} w_{ij} x_j + w_{i0} = \mathbf{w}_i^T \mathbf{x}$$

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

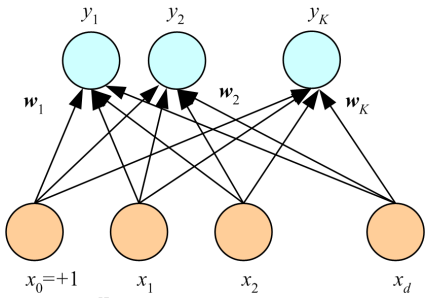


Classification:

$$o_i = \mathbf{w}_i^T \mathbf{x}$$

$$y_i = \frac{\exp o_i}{\sum_k \exp o_k}$$

choose C_i if $y_i = \max_k y_k$



Learning a Perceptron

Update weights as you read through the instances.

- Randomly initialise **w**.
- While overall error is too large (or no improvement): =convergence
- For each $j \in [n]$:

(1) Calculate $\hat{y_j} = \begin{cases} 1 \text{ (for } C_1) & \text{if } g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} > 0 \\ 0 \text{ (for } C_2) & \textit{otherwise} \end{cases}$

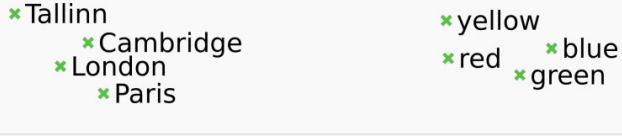
(2) Update each $w_i \leftarrow w_i + \eta(y_j - \hat{y_j})x_{j,i}$ and η is the learning factor (step size)

What is a Perceptron Classifier?

Representing words as vectors

Let's represent words (or any objects) as vectors.
We want to construct them so that similar words have similar vectors.

Sequence	Count
I live in Cambridge	19
I live in Paris	68
I live in Tallinn	0
I live in yellow	0



1-hot vectors

How can we represent words as vectors?

Option 1: each element represents a different word.

Also known as "1-hot" or "1-of-V" representation.

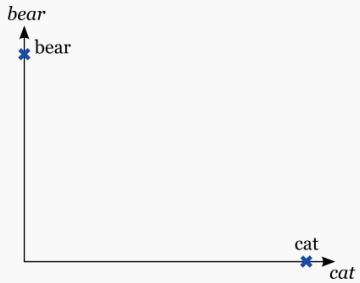
	bear	*cat*	*frog*
bear	1	0	0
cat	0	1	0
frog	0	0	1

```
bear=[1.0, 0.0, 0.0]        cat=[0.0, 1.0, 0.0]
```

1-hot vectors

When using 1-hot vectors, we can't fit many and they tell us very little.

Need a separate dimension for every word we want to represent.



Distributed vectors

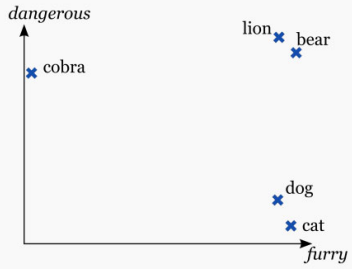
Option 2: each element represents a property, and they are shared between the words.

Also known as "distributed" representation.

	furry	*dangerous*	*mammal*
bear	0.9	0.85	1
cat	0.85	0.15	1
frog	0	0.05	0

```
bear = [0.9, 0.85, 1.0]     cat = [0.85, 0.15, 1.0]
```

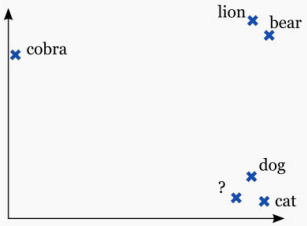
Distributed vectors



	furry	*dangerous*
bear	0.9	0.85
cat	0.85	0.15
cobra	0.0	0.8
lion	0.85	0.9
dog	0.8	0.15

Distributed vectors group similar words/objects together

Distributed vectors



We can infer some information, based only on the vector of the word

We don't even need to know the labels on the vector elements

Distributional hypothesis

Words which are similar in meaning occur in similar contexts.
(Harris, 1954)

You shall know a word by the company it keeps
(Firth, 1957)

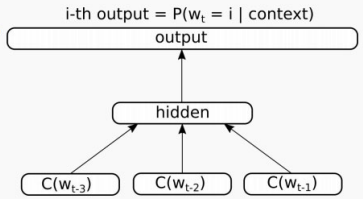
He is reading a **magazine**	I was reading a **newspaper**
This **magazine** published my story	The **newspaper** published an article
She buys a **magazine** every month	He buys this **newspaper** every day

Embeddings through language modelling

Predict the next word in a sequence, based on the previous words.

Use this to guide the training for word embeddings.

Bengio et. al. 2003. *A Neural Probabilistic Language Model.*



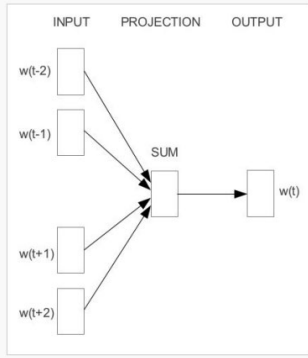
I **read** at my **desk**

I **study** at my **desk**

Continuous Bag-of-Words (CBOW) model

Predict the current word, based on the surrounding words

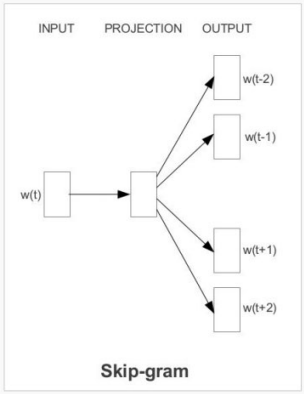
Mikolov et. al. 2013. *Efficient Estimation of Word Representations in Vector Space.*



Skip-gram model

Predict the surrounding words, based on the current word.

Mikolov et. al. 2013. *Efficient Estimation of Word Representations in Vector Space.*



Skip-gram

Representing texts simply (without further NN)

Ex. `The cat is nice.`

where

`cat`=(0.1,0.4), `the`=(0.5,0.5), `is`=(0.8,0.9), `nice`=(0.2, 0.3), .=(0.8,0.8)

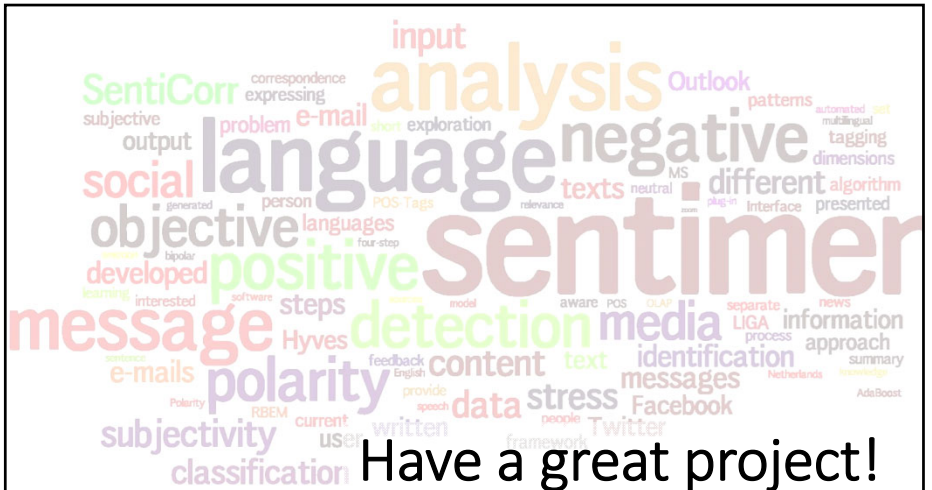
Steps

1) Segment the text into words
2) Make aggregate of vectors

`Cat, the, is, nice, .`

- Average? Standard-deviation? …
- What about ambiguity?
- What about unknown words?

What is a word embedding?



Have a great project!