Natural Language Processing

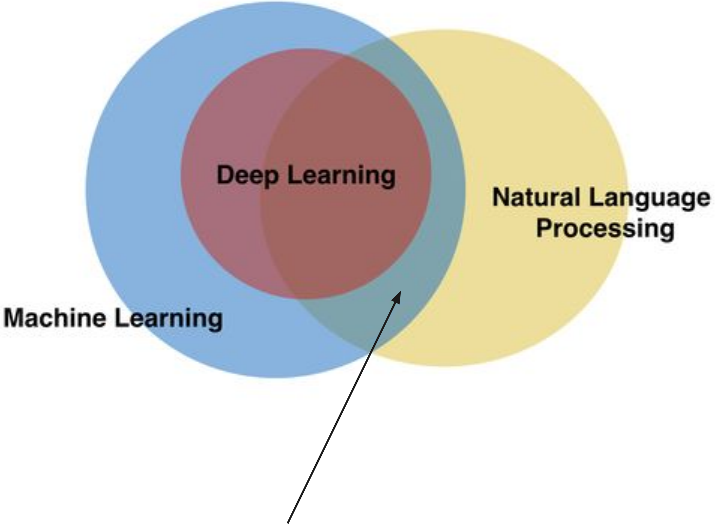
Introduction

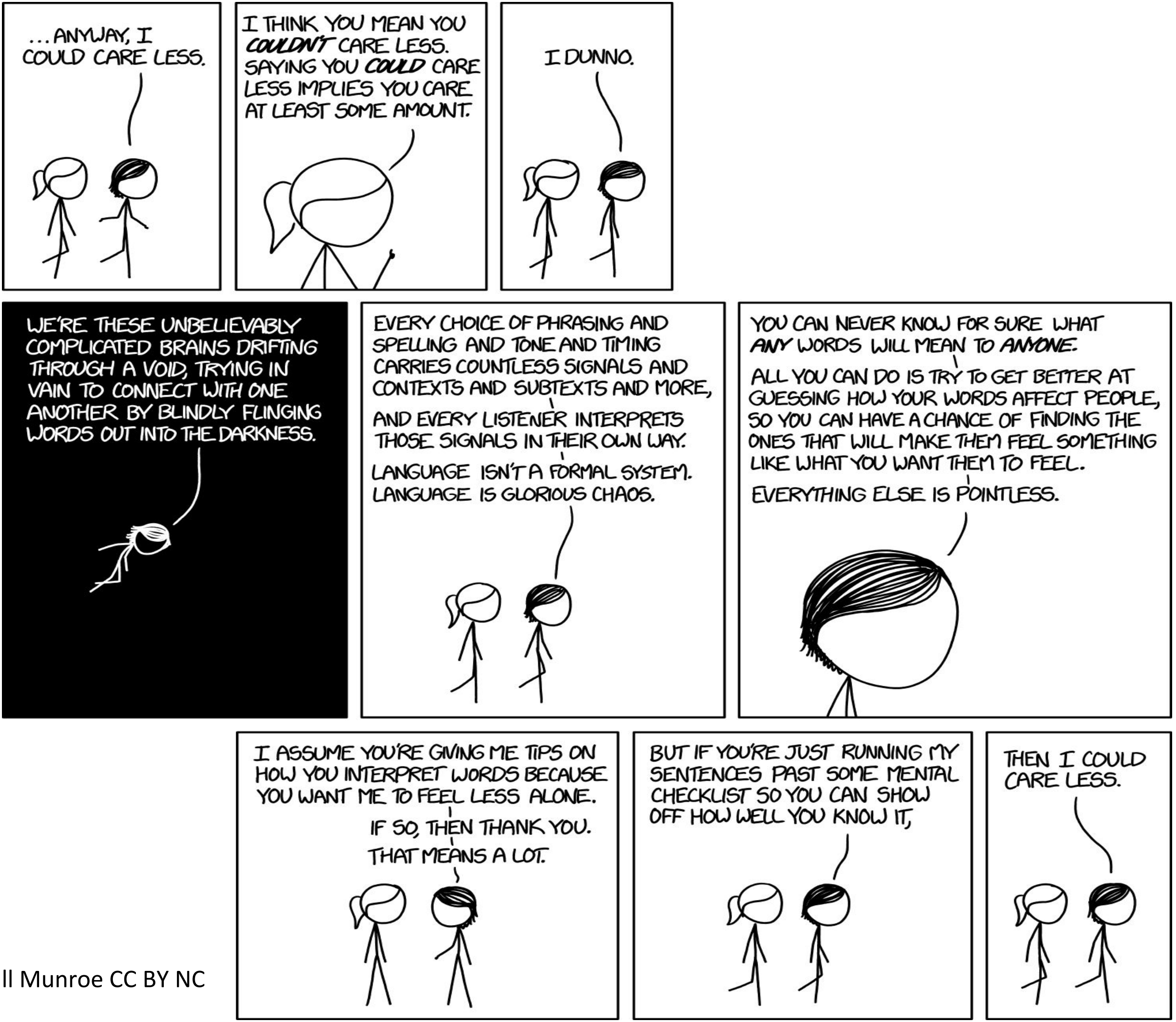
Agenda

- Language + NLP
- How to represent words?
- How do we learn these representations?
- Are these representations are good? Why?
- MATLAB + homework

Natural Language Processing

NLP + Machine Learning




...ANYWAY, I COULD CARE LESS.
I THINK YOU MEAN YOU ***COULDN'T*** CARE LESS. SAYING YOU ***COULD*** CARE LESS IMPLIES YOU CARE AT LEAST SOME AMOUNT.
I DUNNO.
WE'RE THESE UNBELIEVABLY COMPLICATED BRAINS DRIFTING THROUGH A VOID, TRYING IN VAIN TO CONNECT WITH ONE ANOTHER BY BLINDLY FLINGING WORDS OUT INTO THE DARKNESS.
EVERY CHOICE OF PHRASING AND SPELLING AND TONE AND TIMING CARRIES COUNTLESS SIGNALS AND CONTEXTS AND SUBTEXTS AND MORE,
AND EVERY LISTENER INTERPRETS THOSE SIGNALS IN THEIR OWN WAY.
LANGUAGE ISN'T A FORMAL SYSTEM. LANGUAGE IS GLORIOUS CHAOS.
YOU CAN NEVER KNOW FOR SURE WHAT ***ANY*** WORDS WILL MEAN TO ***ANYONE.***
ALL YOU CAN DO IS TRY TO GET BETTER AT GUESSING HOW YOUR WORDS AFFECT PEOPLE, SO YOU CAN HAVE A CHANCE OF FINDING THE ONES THAT WILL MAKE THEM FEEL SOMETHING LIKE WHAT YOU WANT THEM TO FEEL.
EVERYTHING ELSE IS POINTLESS.
I ASSUME YOU'RE GIVING ME TIPS ON HOW YOU INTERPRET WORDS BECAUSE YOU WANT ME TO FEEL LESS ALONE.
IF SO, THEN THANK YOU.
THAT MEANS A LOT.
BUT IF YOU'RE JUST RUNNING MY SENTENCES PAST SOME MENTAL CHECKLIST SO YOU CAN SHOW OFF HOW WELL YOU KNOW IT,
THEN I COULD CARE LESS.

https://xkcd.com/1576/ Randall Munroe CC BY NC 2.5

riverrun, past Eve and Adam's, from swerve of shore to bend of bay, brings us by a commodius vicus of recirculation back to Howth Castle and Environs.

Sir Tristram, violer d'amores, fr'over the short sea, had **passencore** rearrived from North Armorica on this side the scraggy isthmus of Europe Minor to wielderfight his penisolate war: nor had topsawyer's rocks by the stream Oconee exaggerated themselse to Laurens County's **gorgios** while they went **doublin** their mumper all the time: nor avoice from **afire** bellowsed mishe mishe to tauftauf thuartpeatrick: not yet, though **venissoon** after, had a kidscad buttended a bland old isaac: not yet, though all's fair in **vanessy**, were sosie sesthers wroth with **twone nathandjoe**.

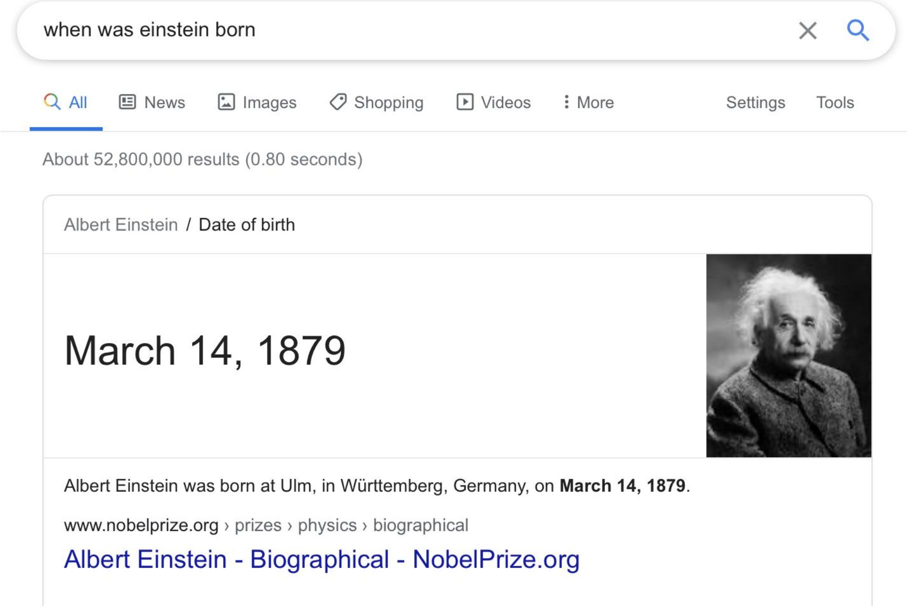
Joyce, James. Finnegans Wake.

Language Characteristics

- Not a formal system
- Is the hallmark of human success
    - How we encode knowledge
- It’s new
- It’s slow
- But information dense
- Spoken vs written

IRL

- Search / question answering
- Text summarization
- Resume filtering
- FB: hate speech, hate memes
- Identifying AI-generated text
- Chatbots
- Translation



when was einstein born

About 52,800,000 results (0.80 seconds)

Albert Einstein / Date of birth

March 14, 1879

Albert Einstein was born at Ulm, in Württemberg, Germany, on **March 14, 1879**.

www.nobelprize.org › prizes › physics › biographical

Albert Einstein - Biographical - NobelPrize.org

Word Representations

As discrete entities

- “Denotational semantics”
- Symbols that point to a meaning
    - A chair denotes all possible chairs

```
cat = [ 0 0 0 0 0 1 0 0 0 ]

dog = [ 0 0 0 0 1 0 0 0 0 ]

pet = [ 0 0 0 0 0 0 0 0 1 ]
```

WordNet

- Lexical database
- Thesaurus / dictionary
- Capture synonym sets and hypernyms (relationships)
- Capture different senses of a word
- A “network” of words
- Drawbacks:
    - In some cases, words may share some meaning but be in different synonym sets, so not an explicit representation
    - Built by human labor
    - Isn’t up to date new slang

Distributional Hypothesis

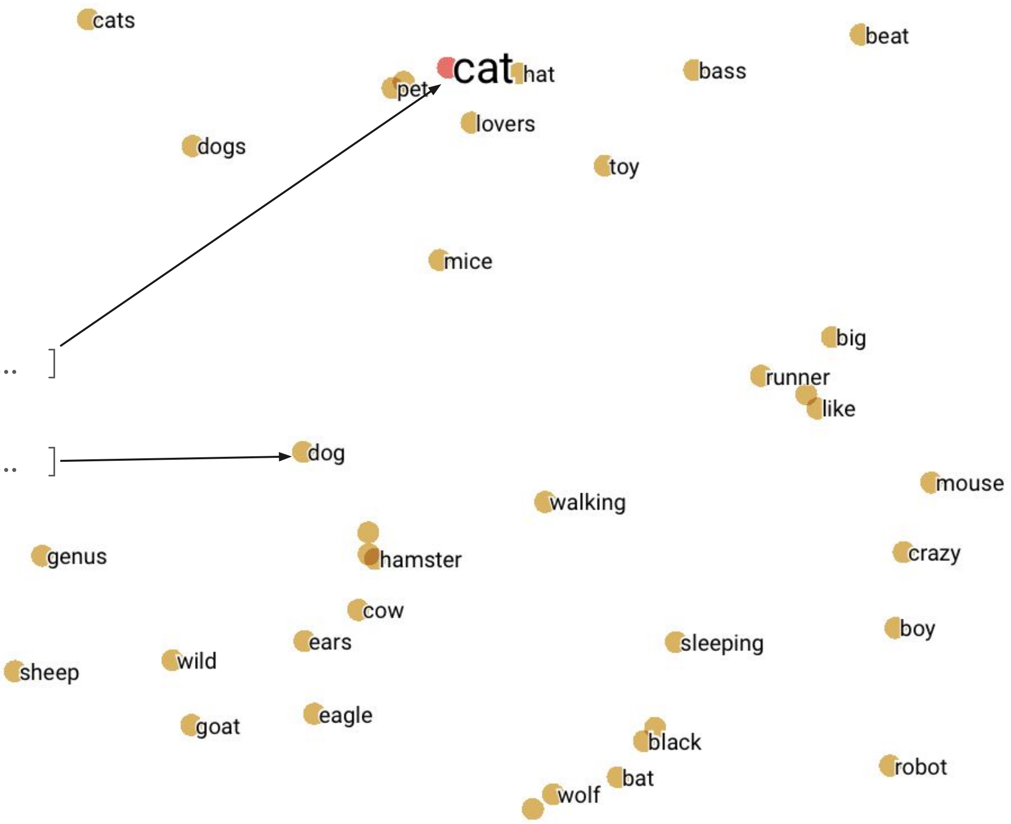
- “You shall know a word by the company it keeps”(J. R. Firth 1957: 11)
- Words that are used and occur in the same contexts tend to purport similar meanings (Harris, Z. 1954)
- Words that share a “context” have similar meanings

Here are a few facts that will ________ you.

Distributed representations

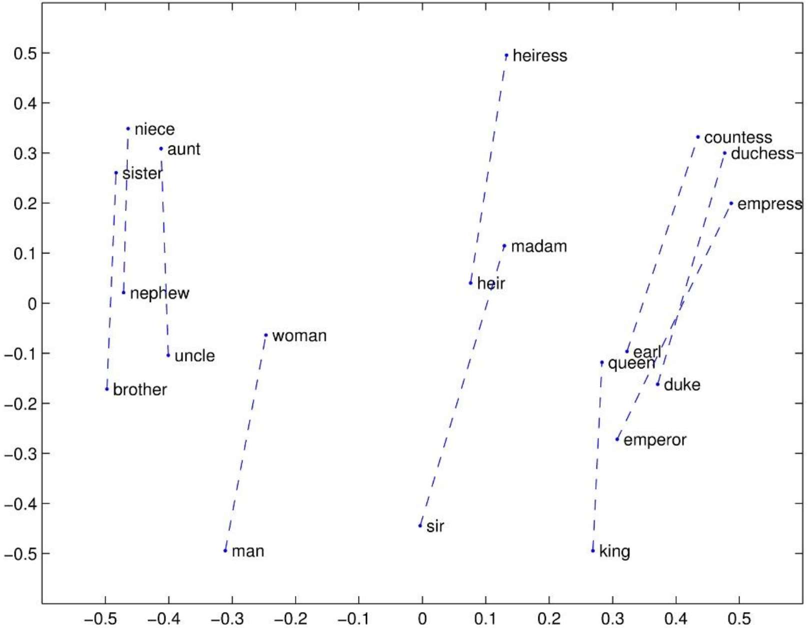
```
cat = [ … 0.212 0.23 0.345 … ]

dog = [ … 0.934 0.64 0.091 … ]
```

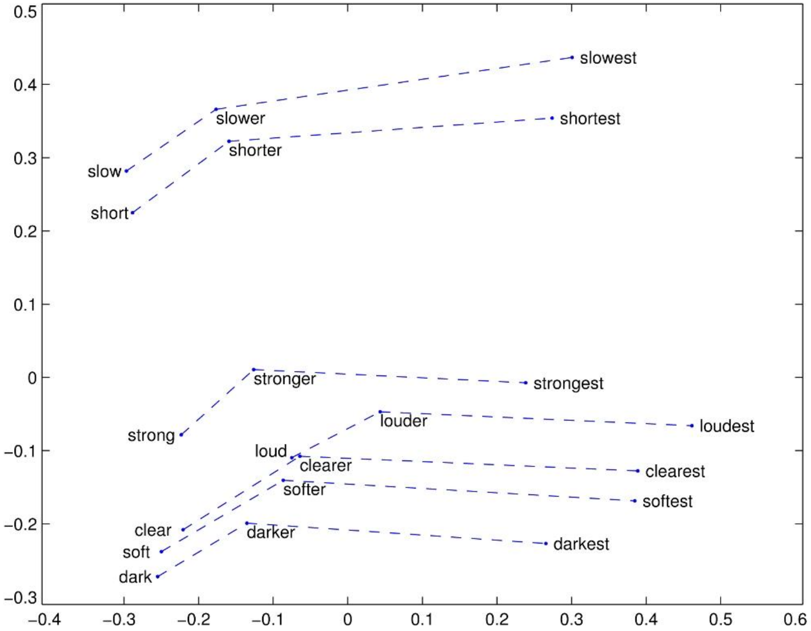


Meaningful Directions

Semantic



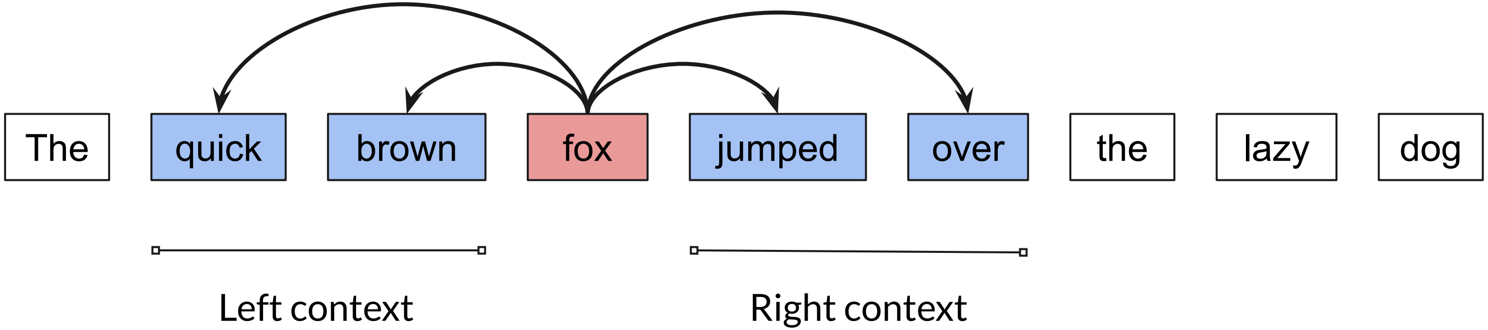
Syntactic



Learning Embeddings

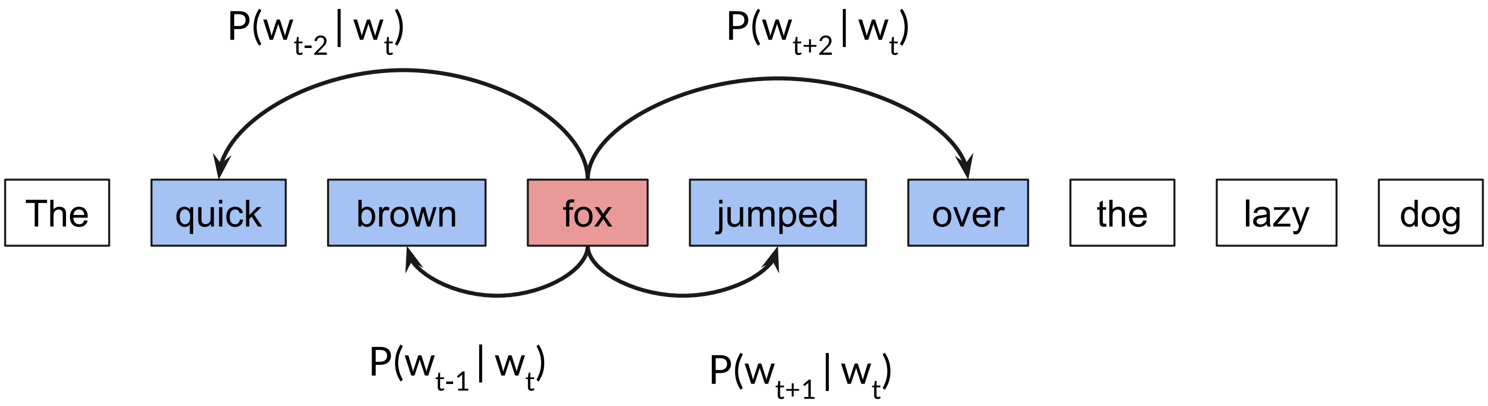
Word2Vec

- A word’s **embedding** is a d-dimensional vector that we learn (starts random)
- A word’s **context** is the set of words that occur near it (e.g. **n** words on either side)
- We want the model to be able to predict what the context words are given the content words



Mikolov et al. 2013

Word2Vec



Mikolov et al. 2013

Word2Vec

$$\underset{\uparrow}{L(\theta)} = \prod_{t=1}^{T} \prod_{\substack{-m \le j \le m \\ j \ne 0}} P\left(w_{t+j} \mid w_t; \theta\right)$$

$$P(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)}$$

Evaluating Embeddings

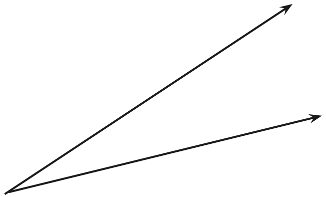
Evaluate Embeddings

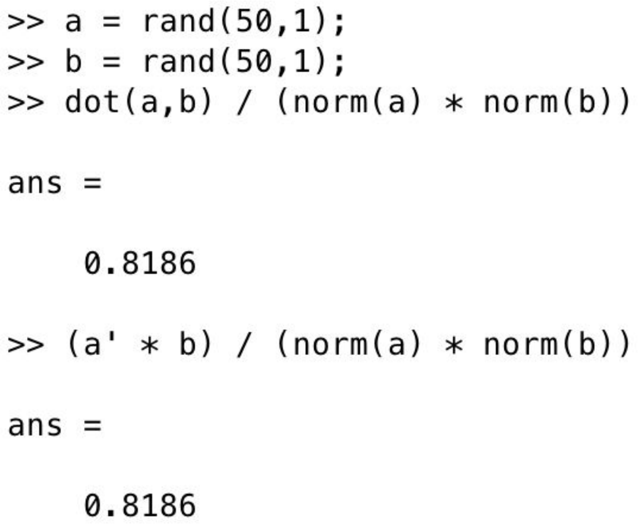
- Intrinsic
  - Specific, intermediate task
  - Help us understand vector space properties
  - Word analogies
  - Similarity correlation evaluation
- Extrinsic
  - On actual task
  - Can be slow
  - Sentiment classification

Cosine Similarity/Distance

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

cosine distance = 1 - cosine similarity





```
>> a = rand(50,1);
>> b = rand(50,1);
>> dot(a,b) / (norm(a) * norm(b))

ans =

    0.8186

>> (a' * b) / (norm(a) * norm(b))

ans =

    0.8186
```

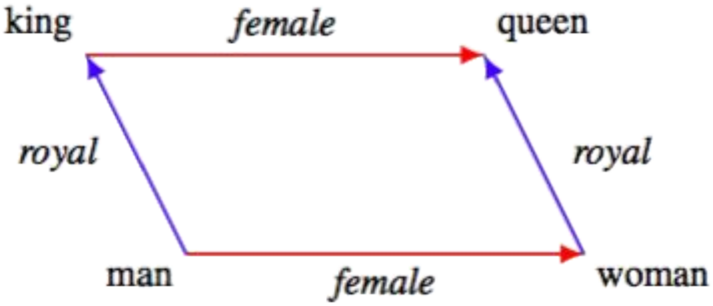
Word Similarity

- Compute similarity between two words
- Ask humans to rate how similar two words are
- Correlate the ratings

		cos-sim
Ring	Apple	0.2760
Dog	Cat	0.9218
Thumb	Tree	0.3564
Germany	Berlin	0.7985

Analogies

- a : b :: c : d
- B - a = d - c
- b - a + c = d



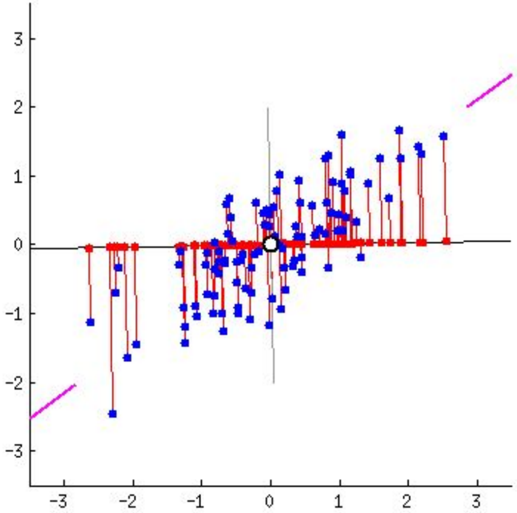
Queen - king = woman - man

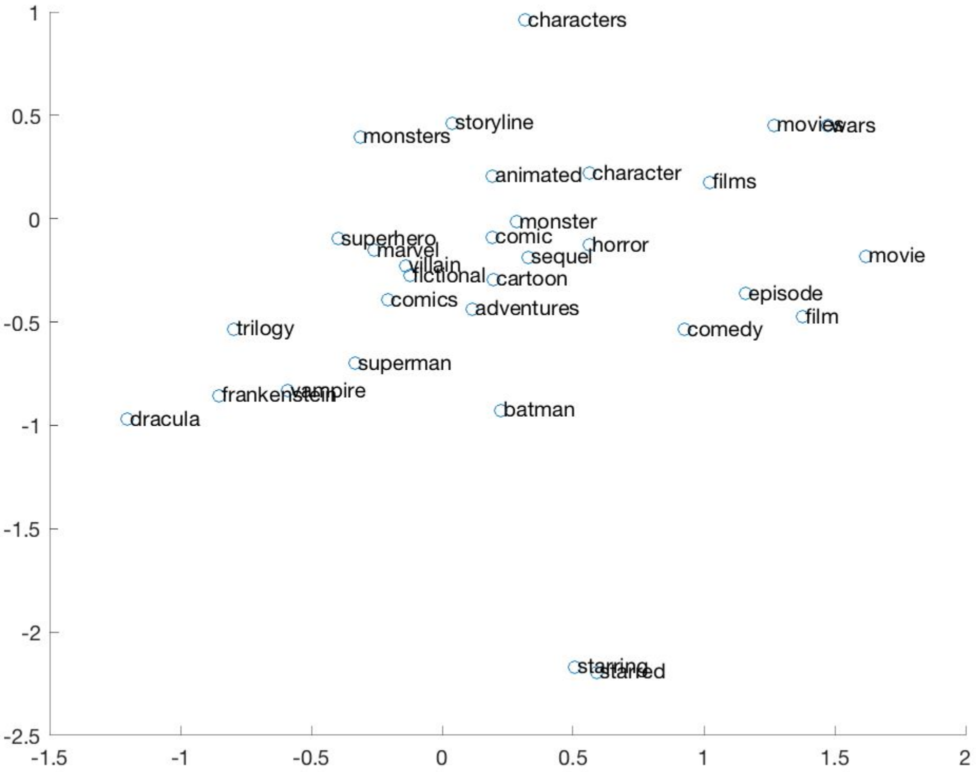
King - man + woman = queen

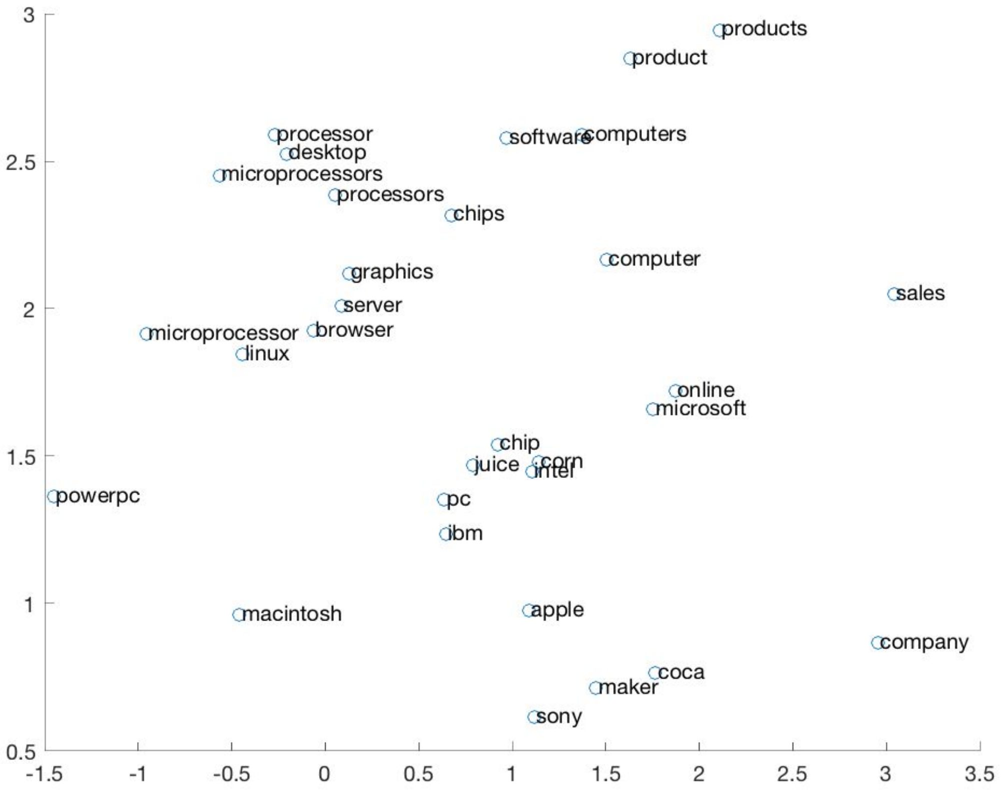
Visualization - PCA

- Principal component analysis
- Allows us to reduce dimensionality of the data
- While preserving much of the variance

- Procedure:
  - Compute principal components (coefficients)
  - Multiply original data by n components to reduce dimensions




characters
storyline
monsters
movies
wars
animated
character
films
monster
superhero
comic
horror
marvel
sequel
movie
villain
fictional
cartoon
episode
comics
adventures
film
trilogy
comedy
superman
frankenstein
vampire
batman
dracula
starring
starred
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-1.5
-1
-0.5
0
0.5
1
1.5
2


products
product
processor
desktop
microprocessors
processors
software
computers
chips
computer
graphics
server
sales
microprocessor
browser
linux
online
microsoft
chip
juice
corn
intel
powerpc
pc
ibm
macintosh
apple
company
coca
maker
sony
3
2.5
2
1.5
1
0.5
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5

PCA - MATLAB

$$\begin{array}{cccc} \text{First principal component} & \text{Second principal component} & & \\ \downarrow & \downarrow & & \\ \left[\begin{matrix} a_{1,1} \\ a_{2,1} \\ a_{3,1} \\ \vdots \end{matrix}\right. & \begin{matrix} a_{1,2} \\ a_{2,2} \\ a_{3,2} \\ \vdots \end{matrix} & \begin{matrix} a_{1,3} \\ a_{2,3} \\ a_{3,3} \\ \vdots \end{matrix} & \left.\begin{matrix} \cdots \\ \cdots \\ \cdots \\ \ddots \end{matrix}\right] \end{array} \text{(n x n)}$$

Column sorted by how much variance they explain

Clustering

- Hierarchical
- K-means
- What kind of clusters do you think you would get?

Polysemy

- A problem with word embeddings…
- The same word can have multiple meanings
- Context matters

"The word *good* has many meanings. For example, if a man were to shoot his grandmother at a range of five hundred yards, I should call him a good shot, but not *necessarily* a good man."

– G.K. Chesterton, "Orthodoxy," 1909

MATLAB / hw