Vector Space Models

Prof. Sameer Singh

CS 295: STATISTICAL NLP

WINTER 2017

January 19, 2017

Based on slides from Jacob Eisenstein, Noah Smith, Mohit Bansal, Richard Socher, and everyone else they copied from.

Outline

- Latent Semantic Analysis
- Vector Models for Words
- Reducing the Dimensions
- Direct Embeddings

Outline

- Latent Semantic Analysis
- Vector Models for Words
- Reducing the Dimensions
- Direct Embeddings

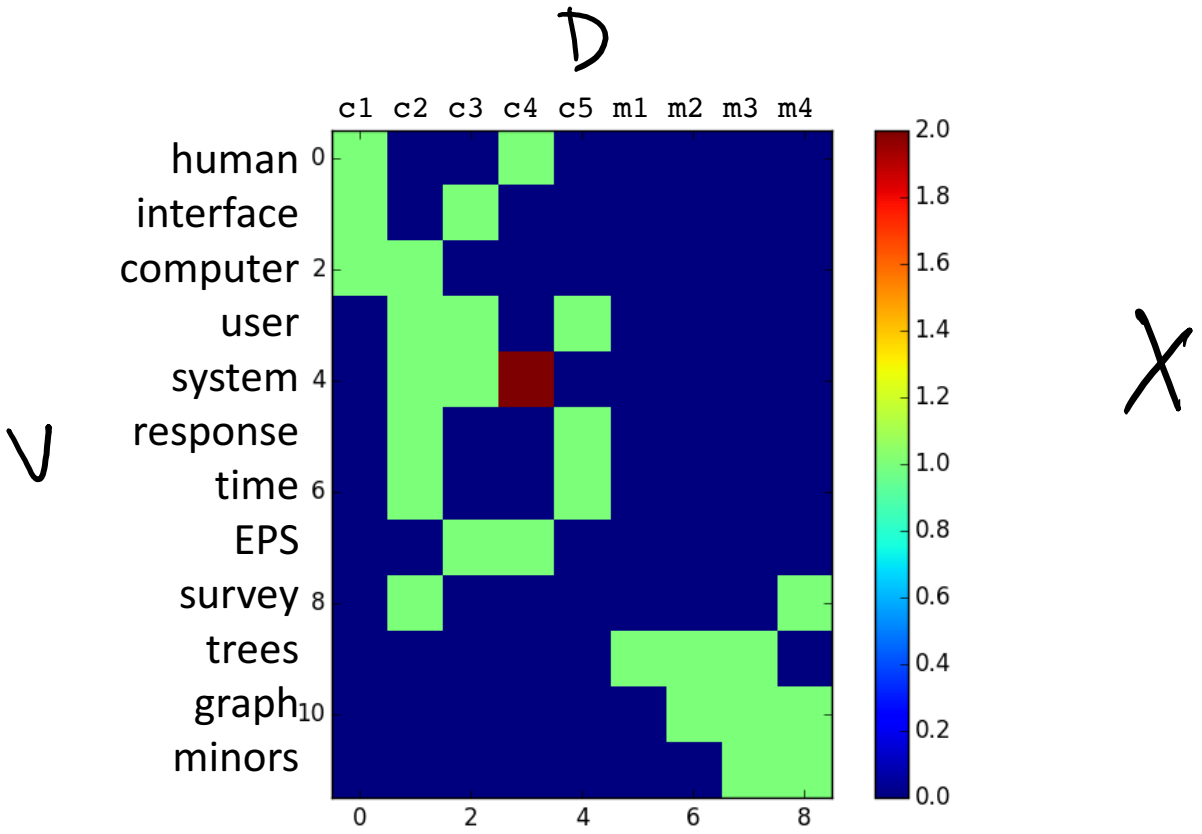
Example: Documents

c1: Human machine interface for ABC computer applications
c2: A survey of user opinion of computer system response time
c3: The EPS user interface management system
c4: System and human system engineering testing of EPS
c5: Relation of user perceived response time to error measurement

m1: The generation of random, binary, ordered trees
m2: The intersection graph of paths in trees
m3: Graph minors IV: Widths of trees and well-quasi-ordering
m4: Graph minors: A survey

From http://lsa.colorado.edu/papers/dp1.LSAintro.pdf

Example: Term-Doc Matrix



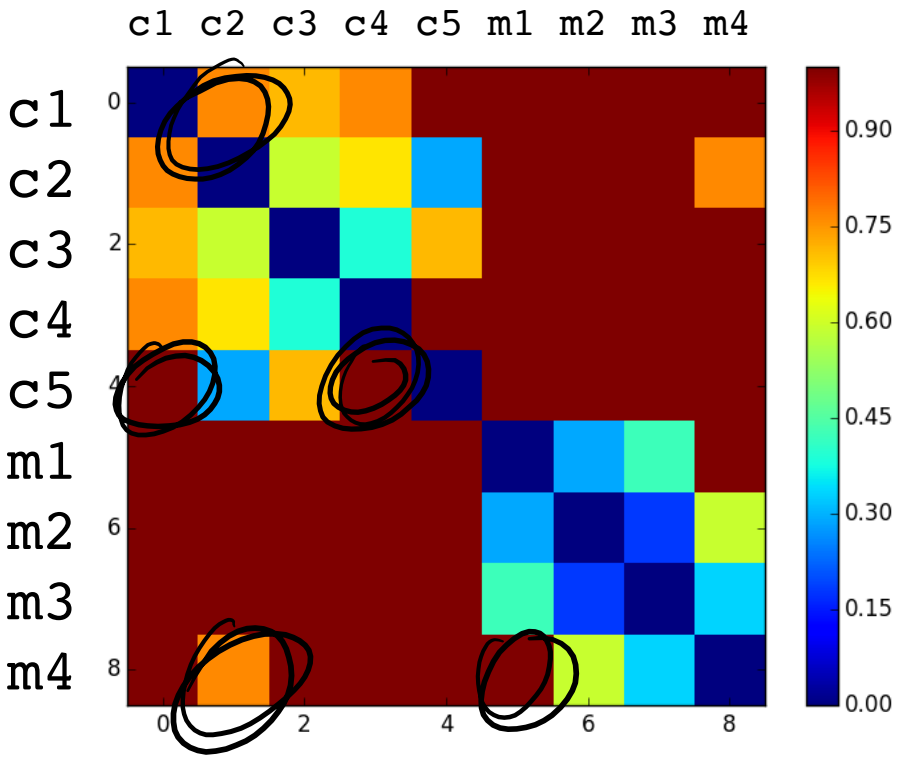
Problems with Sparse Vectors

c2: A survey of user opinion of computer system response time

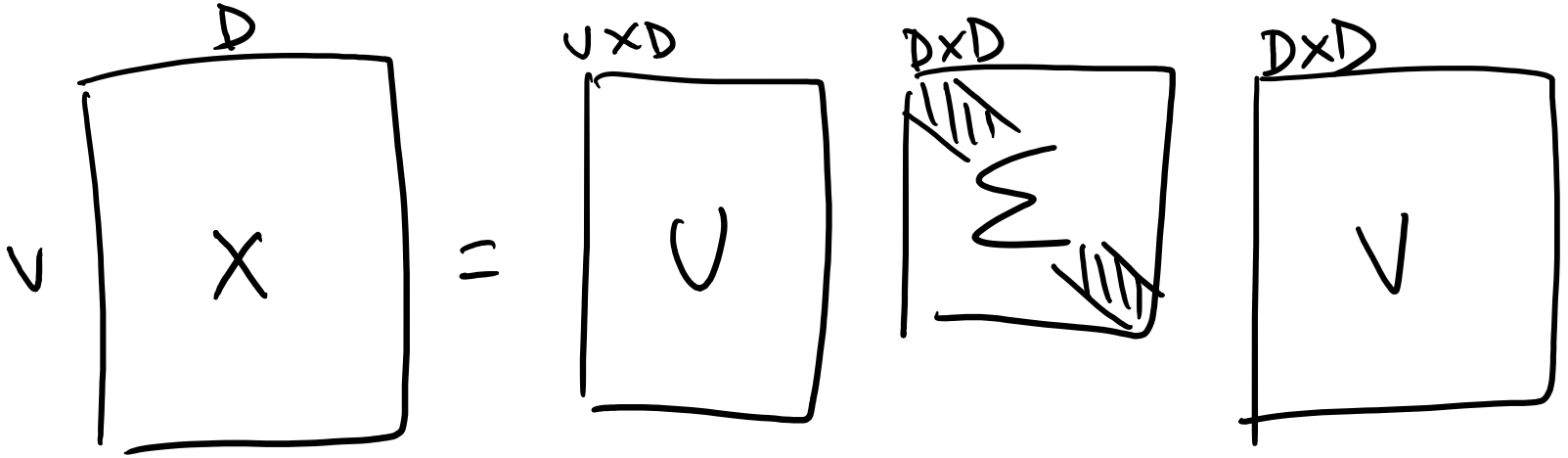
m4: Graph minors: A survey

c1: Human machine interface
for ABC computer applications

Example: Distance Matrix



Option 2: SVD ~ Singular Value Decomp.

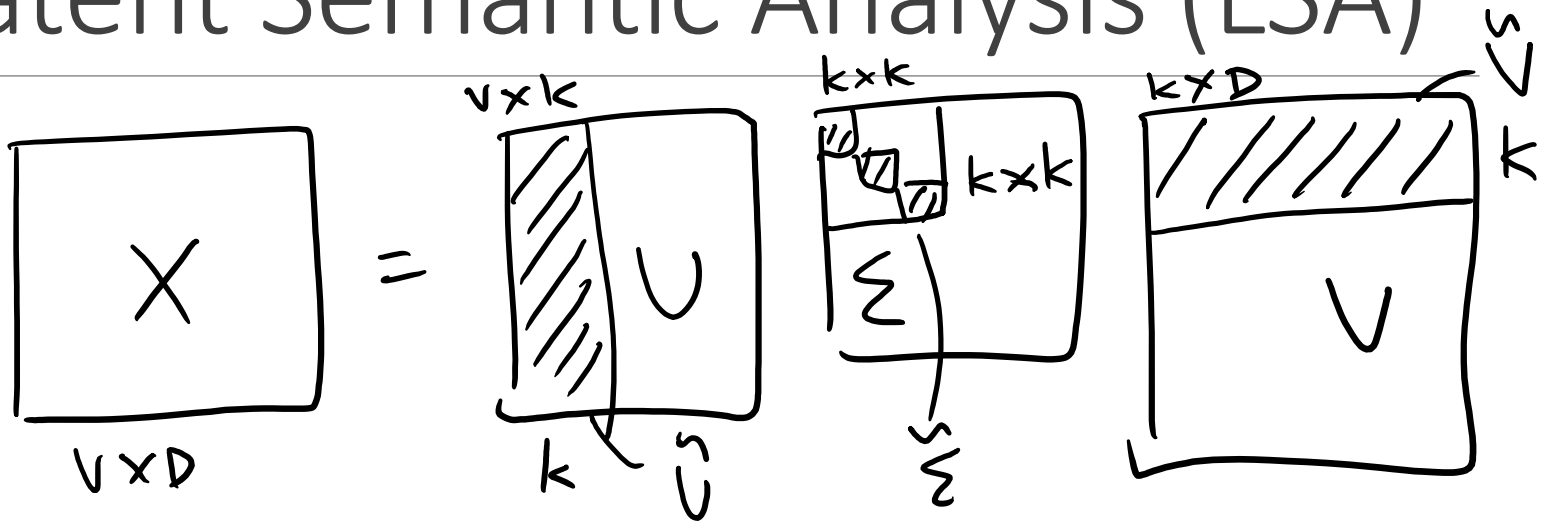


$X = U \times \sqrt{\Sigma} \; W \; C \; \sqrt{\Sigma} \times V$

U x D

D x D

Latent Semantic Analysis (LSA)

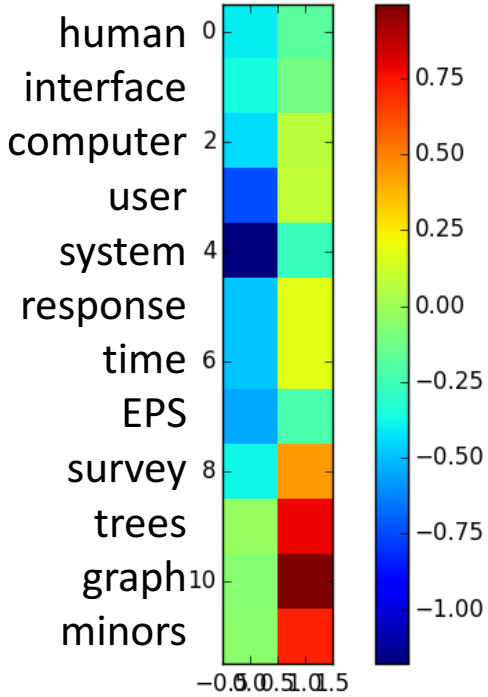


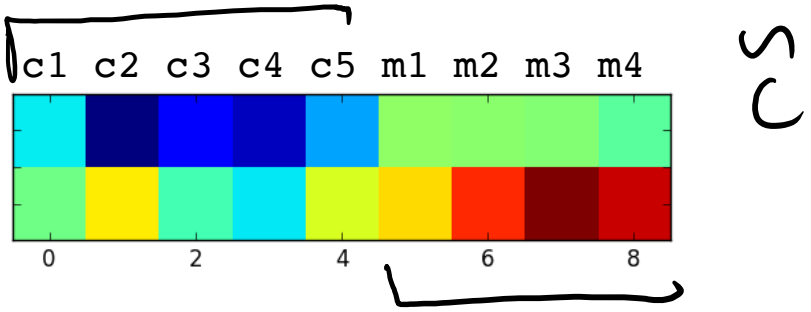
$$\hat{W} = \hat{U} \times \sqrt{\hat{\Sigma}} \qquad \hat{C} = \sqrt{\hat{\Sigma}} \times \hat{V}^T$$

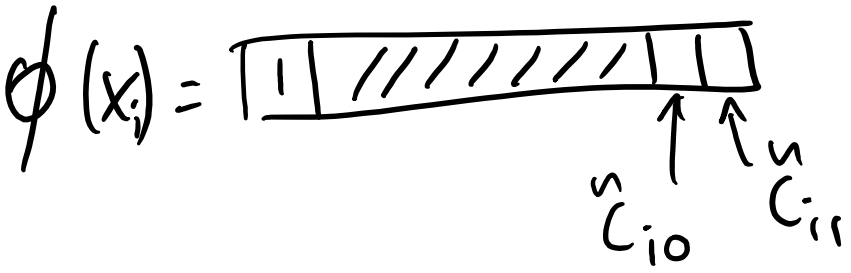
$V \times k \qquad k \times D$

$$\hat{X} = \hat{W} \times \hat{C}$$

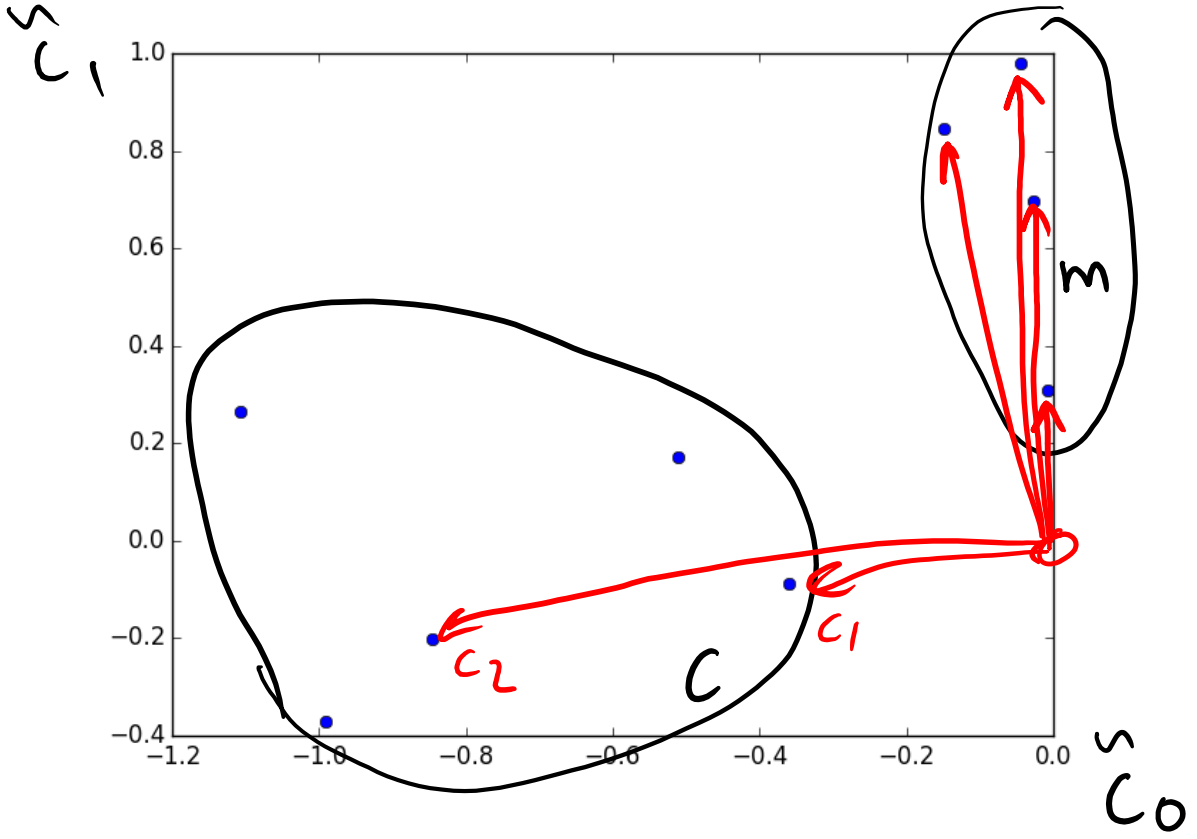
Example: Decomposition, $k = 2$



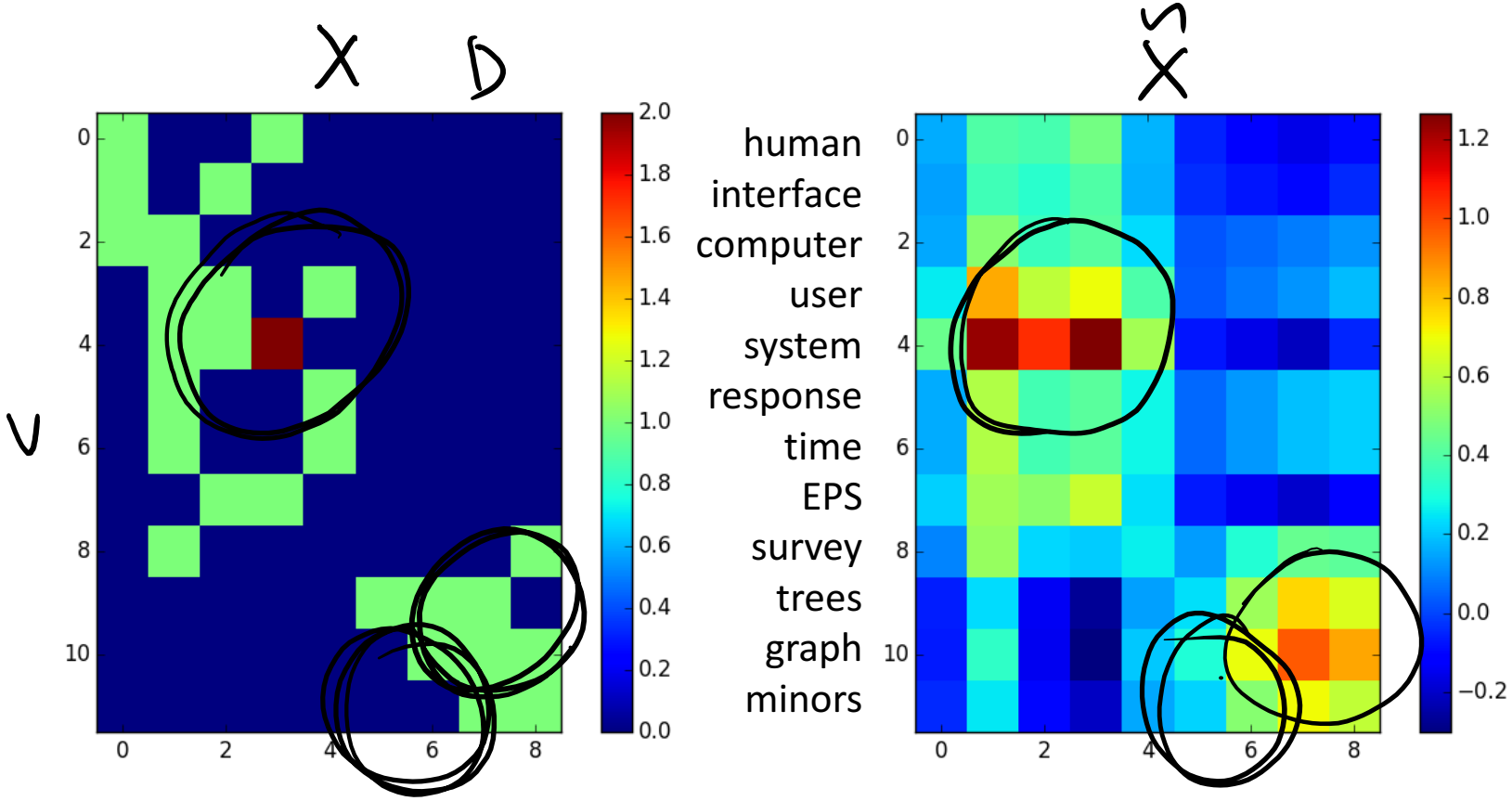




New Document Vectors

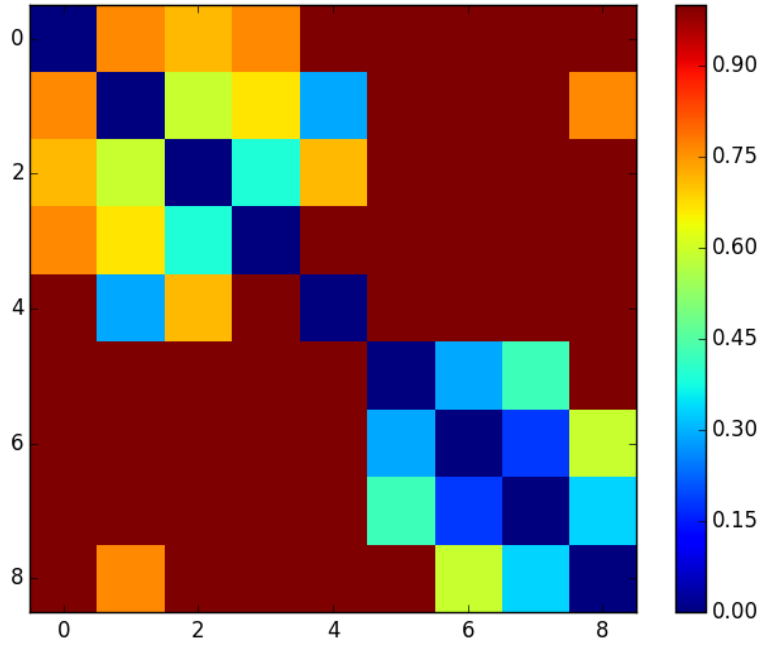


Example: Reconstruction

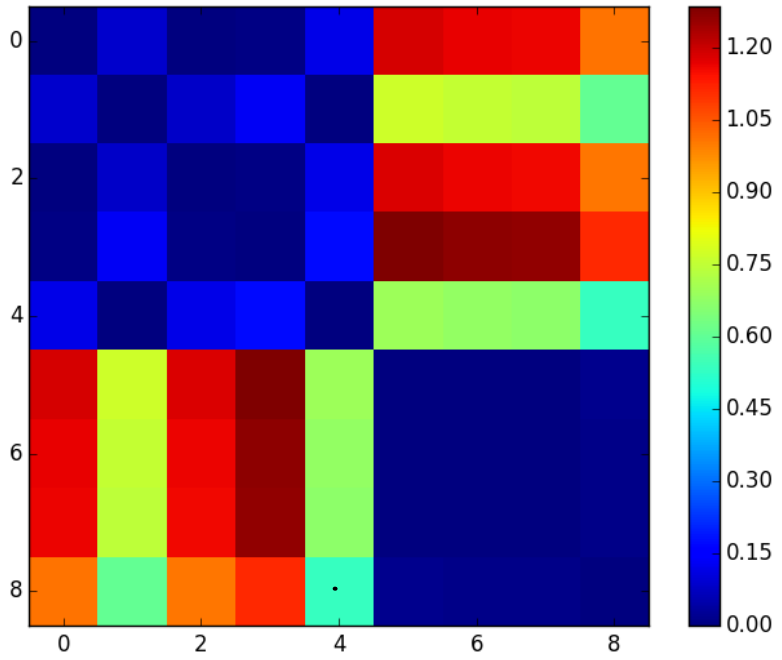


Example: Distance Matrix

$dist(x_i, x_j)$



$dist(\tilde{x}_i, \tilde{x}_j)$



Outline

Latent Semantic Analysis

Vector Models for Words

Reducing the Dimensions

Direct Embeddings

Let's look at words

A bottle of tezguino is on the table.
Everybody likes tezguino.
Tezguino makes you drunk.
We make tezguino out of corn.

What does tezguino mean?
Loud, motor oil, tortillas, choices, wine

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

Term-Context Matrix

C1: A bottle of _____ is on the table.
C2: Everybody likes _____.
C3: _____ makes you drunk.
C4: We make _____ out of corn.

	C1	C2	C3	C4
tezguino	1	1	1	1
loud	0	0	0	0
motor oil	1	0	0	0
tortillas	0	1	0	1
choices	0	1	1	0
wine	1	1	1	0

What is a “Context”?

Can be anything you want!

- Entire contents of the sentence
- One word before and after
- Words in the same sentence
- Document it appears in
- Many other variations…

A bottle of tezguino is on the table.
Tezguino makes you drunk.
…
I had a fancy bottle of wine and got drunk last night!
The terrible wine is on the table.

What is a "Context"?

Can be anything you want!

- Entire contents of the sentence
  - Unlikely to occur again!
- One word before and after
- Words in the same sentence
- Document ID it appears in
- Many other variations…

C1 A bottle of tezguino is on the table.

C2 Tezguino makes you drunk.

…

C3 I had a fancy bottle of wine and got drunk last night!

C4 The terrible wine is on the table.

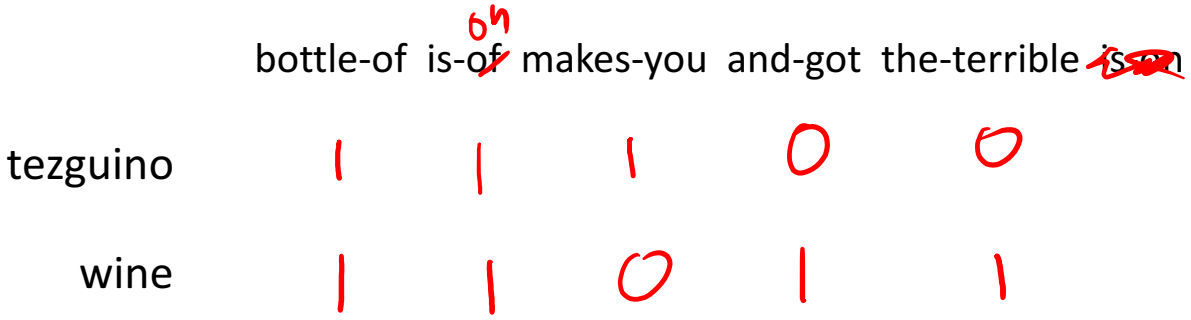
	C1	C2	C3	C4
tezguino	1	1	0	0
wine	0	0	1	1

What is a "Context"?

Can be anything you want!

- Entire contents of the sentence
- One word before and after
  - Or n-words
- Words in the same sentence
- Document it appears in
- Many other variations…

A bottle of tezguino is on the table.
Tezguino makes you drunk.
…
I had a fancy bottle of wine and got drunk last night!
The terrible wine is on the table.

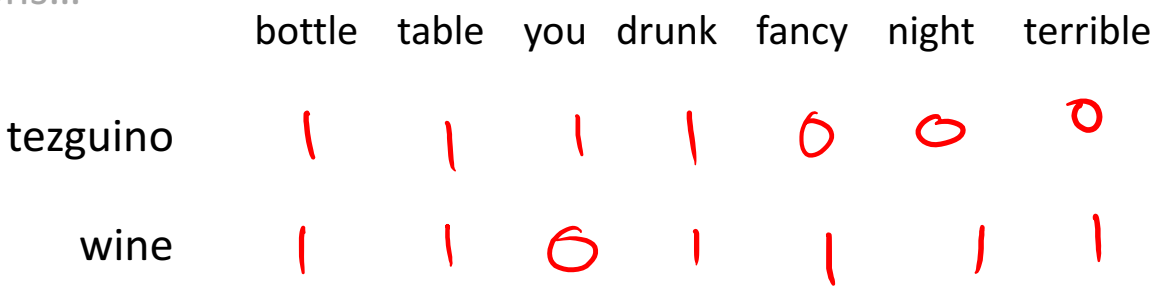


What is a "Context"?

Can be anything you want!

- Entire contents of the sentence
- One word before and after
- Words in the same sentence
  - Filter: nouns and verbs?
  - Bag of words in a window
- Document it appears in
- Many other variations…

A bottle of tezguino is on the table.
Tezguino makes you drunk.
…
I had a fancy bottle of wine and got drunk last night!
The terrible wine is on the table.

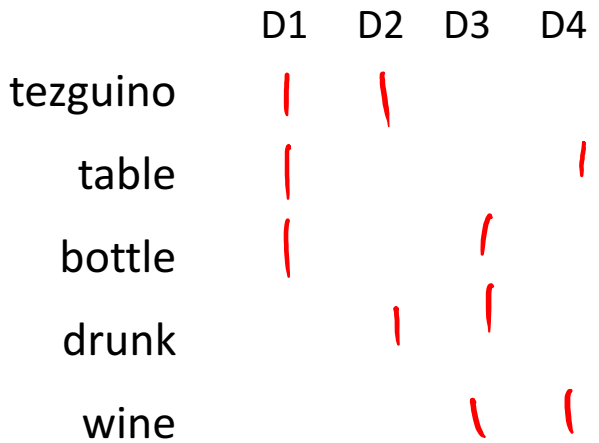


What is a "Context"?

Can be anything you want!

- Entire contents of the sentence
- One word before and after
- Words in the same sentence
- Document it appears in
  - Term-document matrix!
  - Latent Semantic Analysis
- Many other variations…

D1 A bottle of tezguino is on the table.
D2 Tezguino makes you drunk.
…
D3 I had a fancy bottle of wine and got drunk last night!
D4 The terrible wine is on the table.



Pointwise Mutual Information

Raw counts are not good

- Skewed towards common words/contexts
- Many of them are not *informative*
  - is, the, it, they, …

PMI(w,c)

- How much more likely is w to occur in c, than just randomly?

positive

$$PMI(w,c) = \log_2 \frac{P(w,c)}{P(w)P(c)}$$

$P(w,c)$ → $\frac{f_{w,c}}{N}$

$P(w)P(c)$ → $\frac{\sum_j f_{w,j}}{N} \cdot \frac{\sum_i f_{i,c}}{N}$

Outline

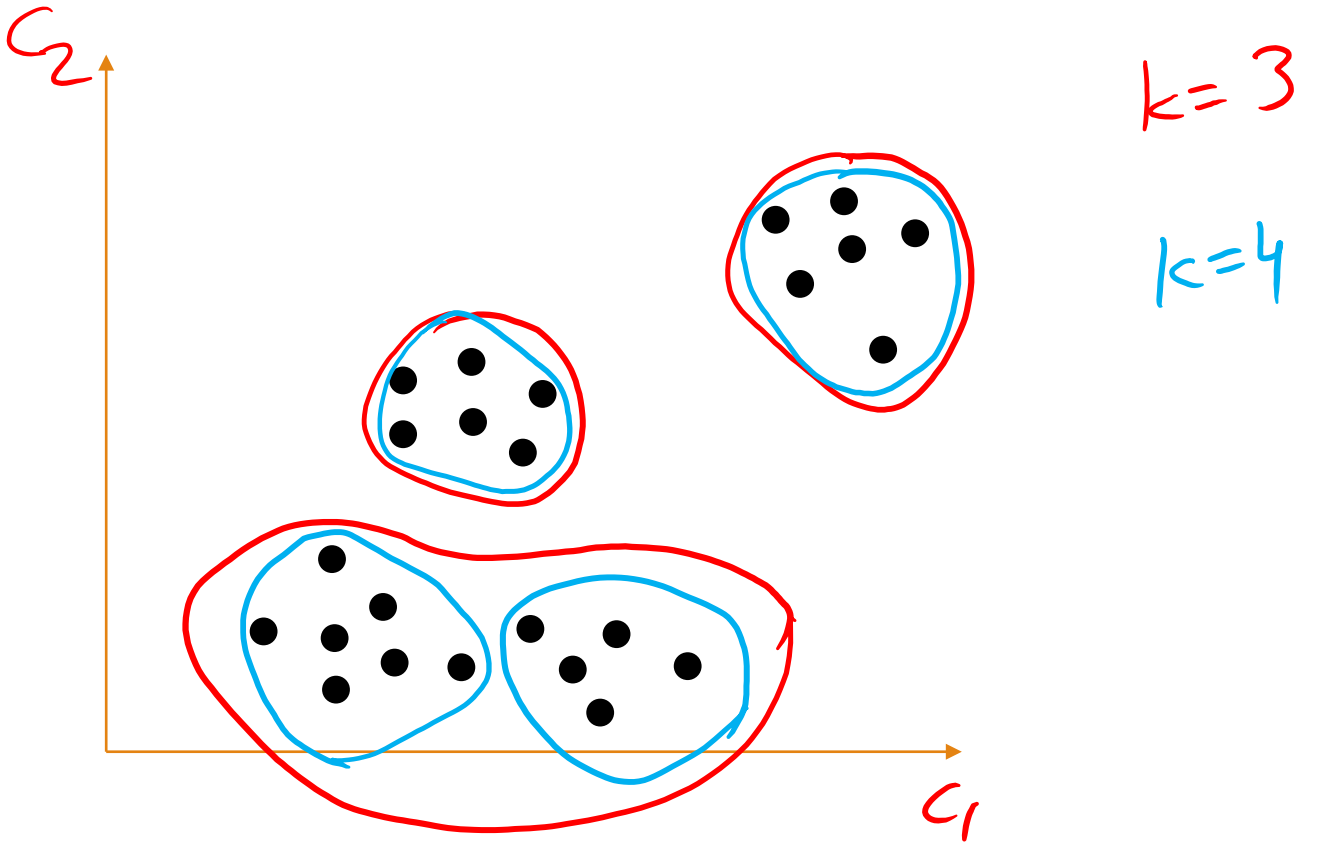
Latent Semantic Analysis

Vector Models for Words

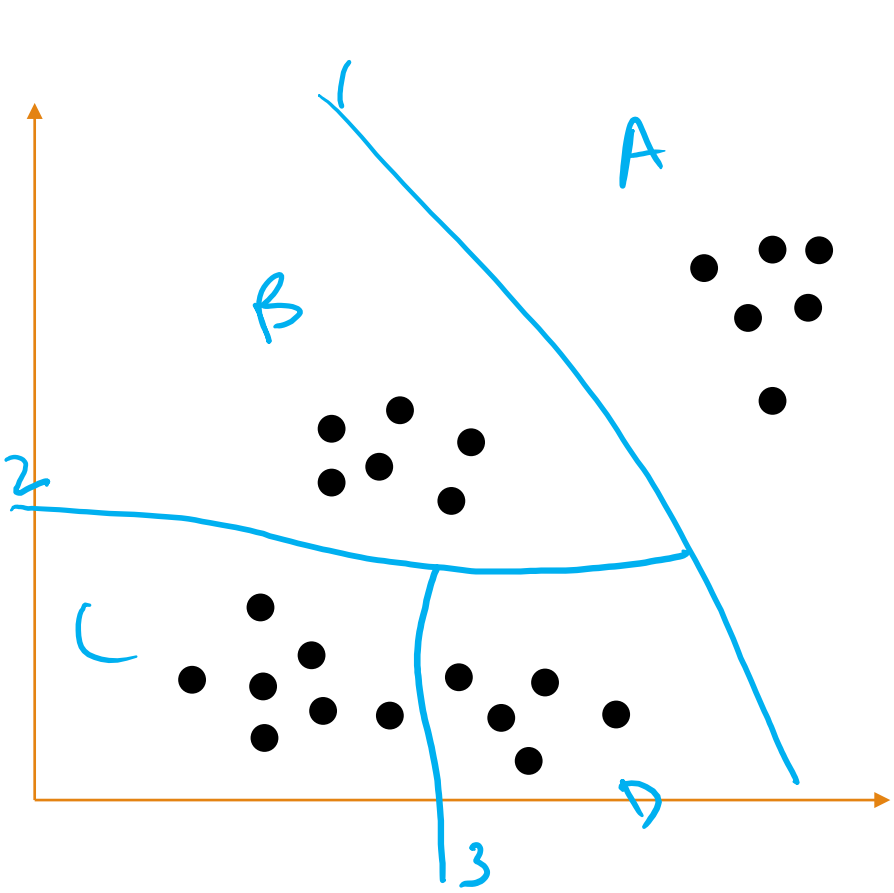
Reducing the Dimensions

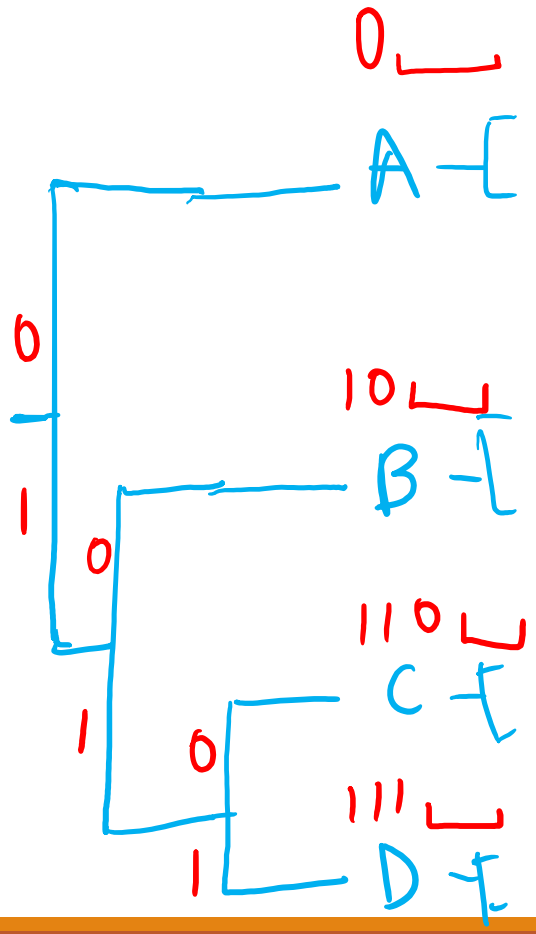
Direct Embeddings

Option 1: Revisiting Clustering

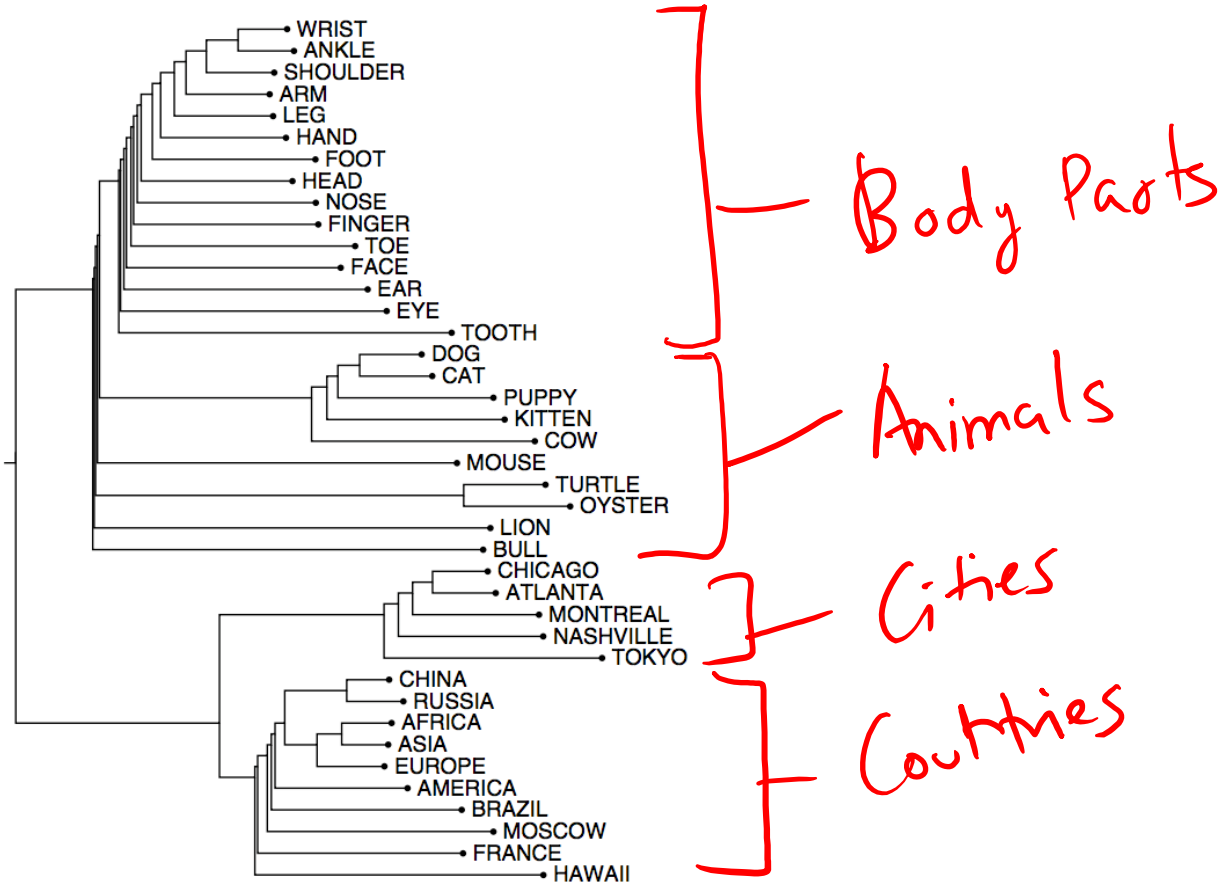


Hierarchical Clustering

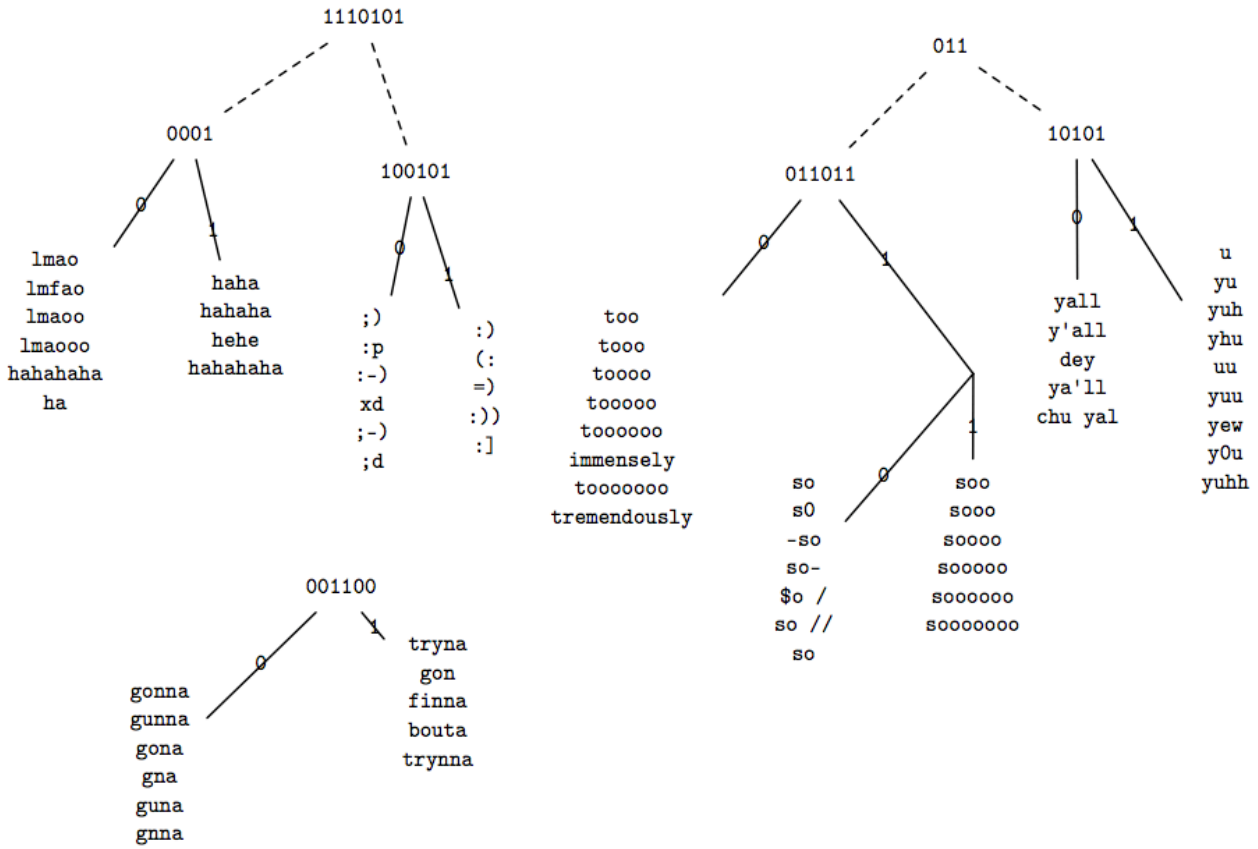




Example

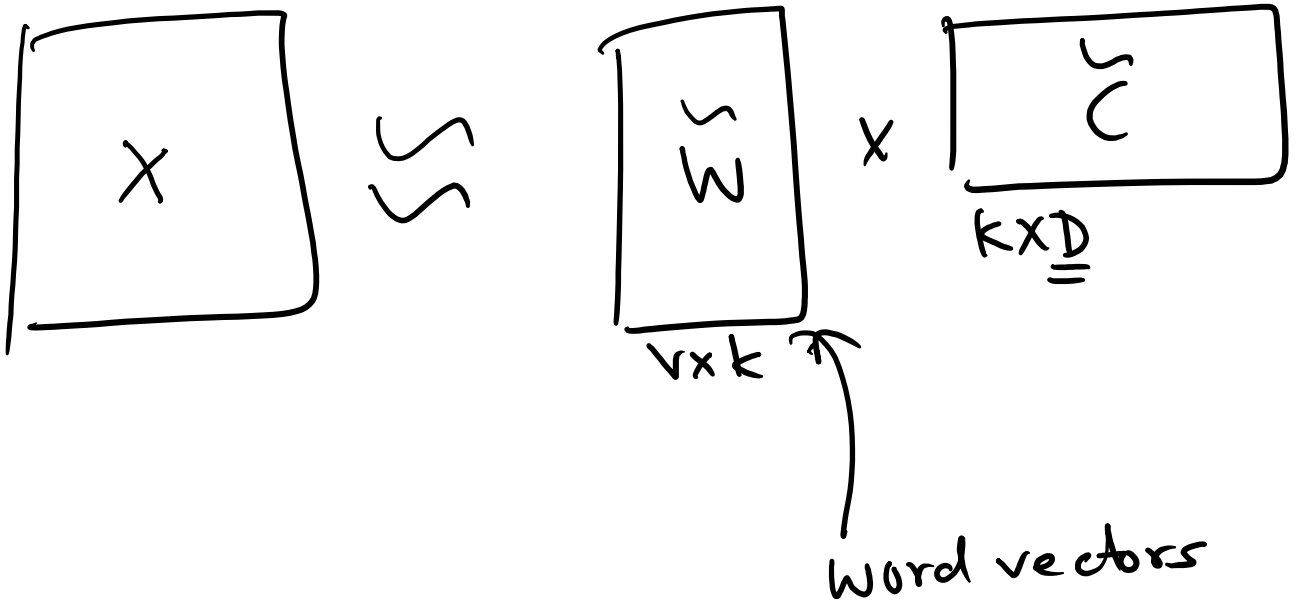

WRIST
ANKLE
SHOULDER
ARM
LEG
HAND
FOOT
HEAD
NOSE
FINGER
TOE
FACE
EAR
EYE
TOOTH
DOG
CAT
PUPPY
KITTEN
COW
MOUSE
TURTLE
OYSTER
LION
BULL
CHICAGO
ATLANTA
MONTREAL
NASHVILLE
TOKYO
CHINA
RUSSIA
AFRICA
ASIA
EUROPE
AMERICA
BRAZIL
MOSCOW
FRANCE
HAWAII
Body Parts
Animals
Cities
Countries

Brown Clusters for Twitter

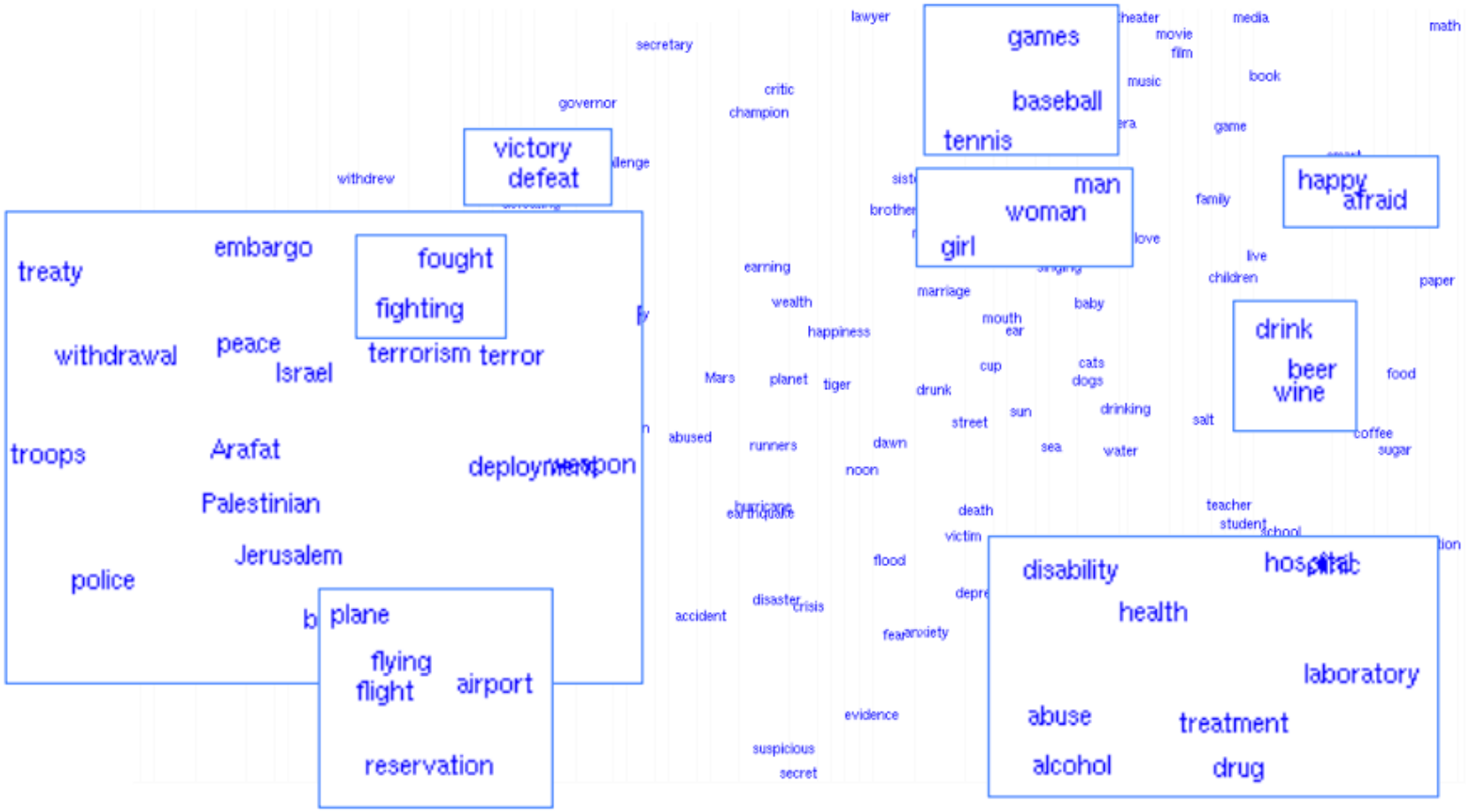


http://www.cs.cmu.edu/~ark/TweetNLP/cluster_viewer.html

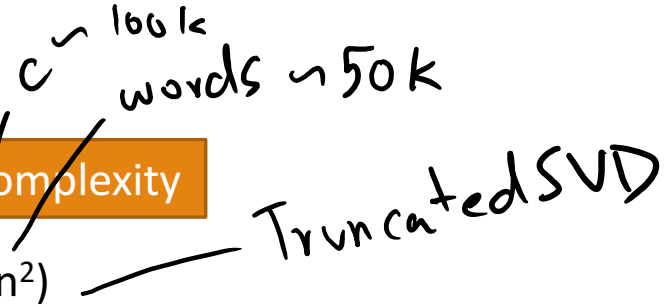
Option 2: SVD



Example Word Projection



Problem with SVD & Clustering



Computational Complexity

- SVD: $O(mn^2)$
- Clustering: $O(knm)$ per iteration, or $O(n^3)$
- But, n can be 100,000!

"One shot"

- Difficult to add new documents or words
- Cannot work with streaming data

Outline

Latent Semantic Analysis

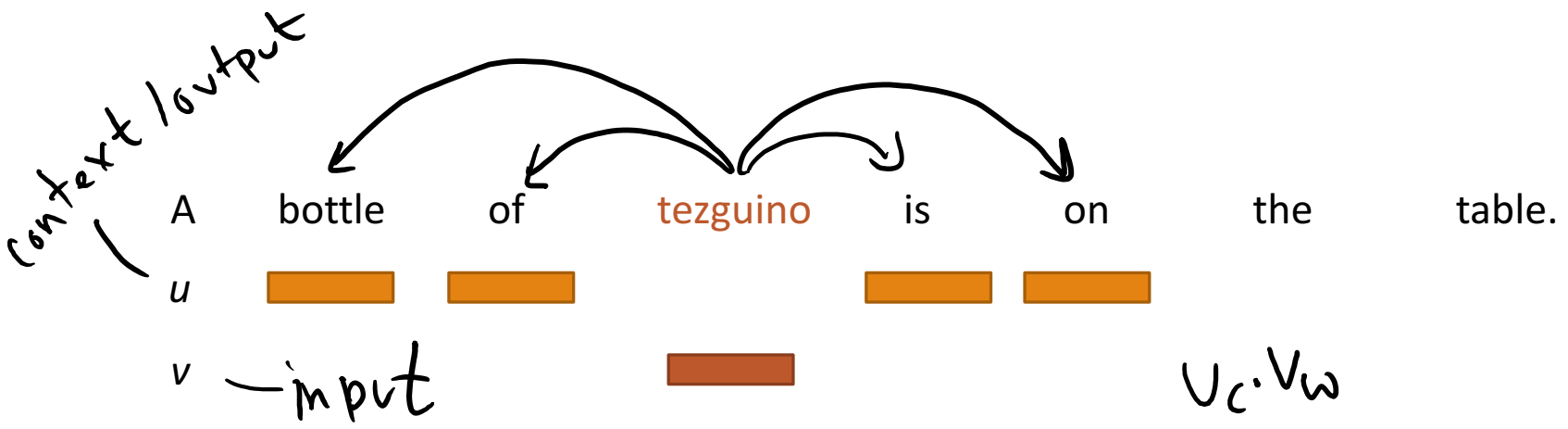
Vector Models for Words

Reducing the Dimensions

Direct Embeddings

Predict surrounding words

$\forall t \; \forall j, -m \dots m, j \neq 0 \quad p(w_{t+j} | w_t)$



$$p(c|w) = \frac{e^{u_c \cdot v_w}}{\sum_i e^{u_i \cdot v_w}}$$

Estimating the Word Vectors

$$\underset{U,V}{\text{argmax}} \sum_{t=1}^{T} \sum_{\substack{-m \le j \le m \\ j \neq 0}} \log P(w_{t+j} \mid w_t)$$

$$J(U,V)$$

$$P(w|c) = \frac{e^{u_w \cdot v_c}}{\sum_{i} e^{u_i \cdot v_c}}$$

(1) all words

$$\frac{\partial J}{\partial u_i}$$

$$\frac{\partial J}{\partial v_j}$$

$$\forall w, c \quad P(w|c) \neq 1$$

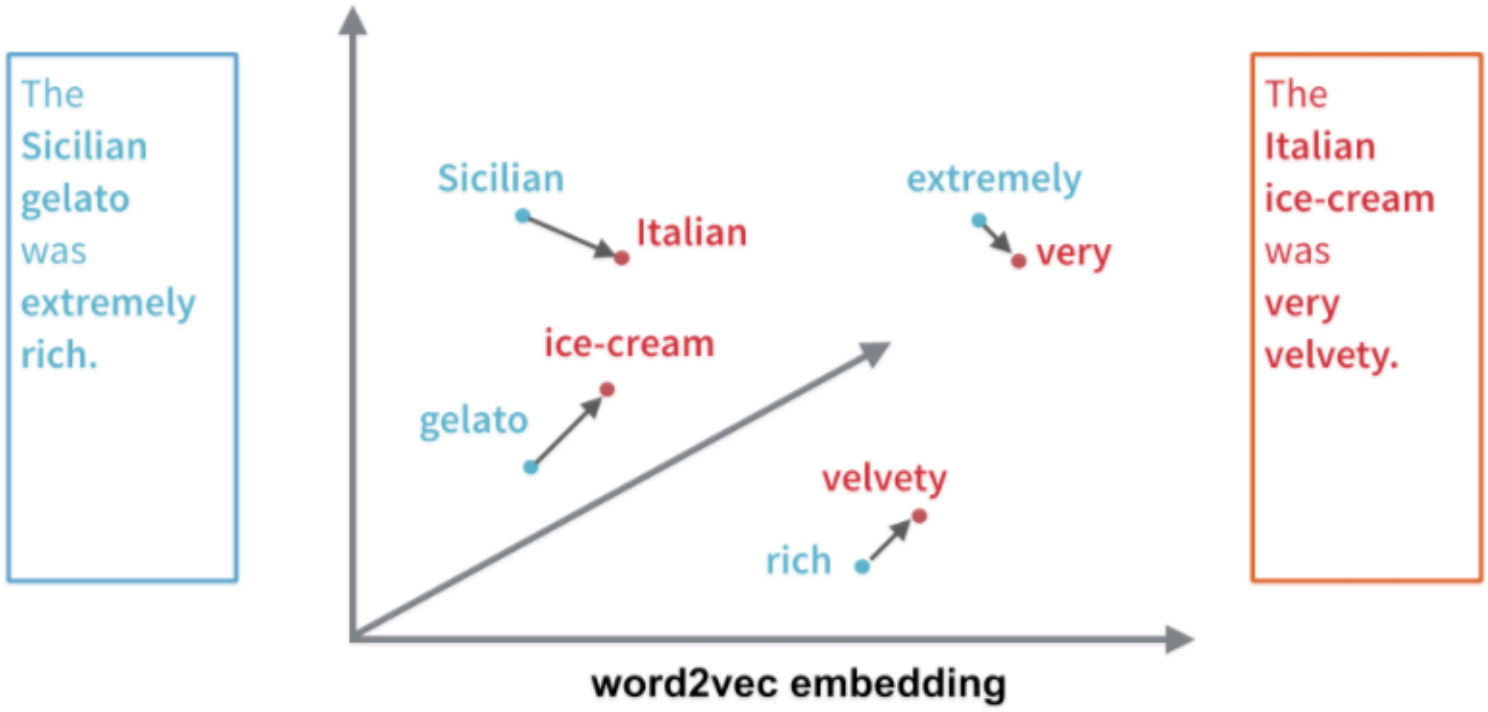
$$\sum_{w} P(w|c) = 1$$

Similar Meaning = Close

Target Word	BoW5	BoW2
batman	nightwing	superman
	aquaman	superboy
	catwoman	aquaman
	superman	catwoman
	manhunter	batgirl
hogwarts	dumbledore	evernight
	hallows	sunnydale
	half-blood	garderobe
	malfoy	blandings
	snape	collinwood
turing	nondeterministic	non-deterministic
	non-deterministic	finite-state
	computability	nondeterministic
	deterministic	buchi
	finite-state	primality

Target Word	BoW5	BoW2
florida	gainesville	fla
	fla	alabama
	jacksonville	gainesville
	tampa	tallahassee
	lauderdale	texas
object-oriented	aspect-oriented	aspect-oriented
	smalltalk	event-driven
	event-driven	objective-c
	prolog	dataflow
	domain-specific	4gl
dancing	singing	singing
	dance	dance
	dances	dances
	dancers	breakdancing
	tap-dancing	clowning

Similar Meaning = Close

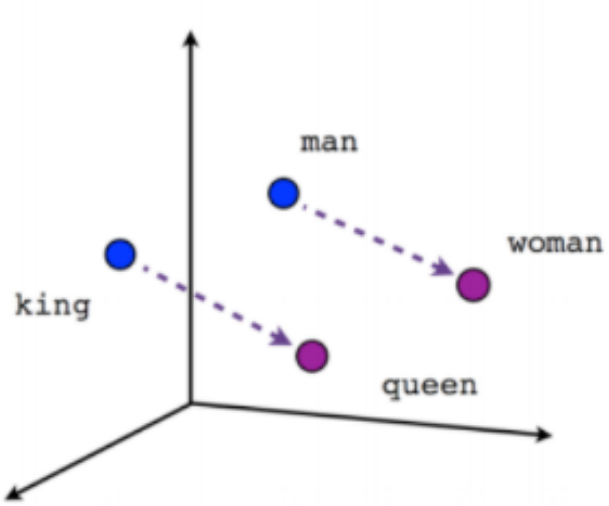


https://siddhant7.github.io/Vector-Representation-of-Words/

Vectors “know” Gender

male : female :: King : queen

King - male + female queen

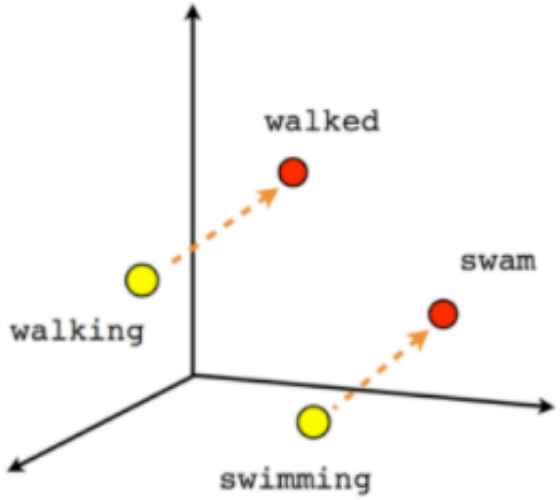


https://siddhant7.github.io/Vector-Representation-of-Words/

They “know” Tenses!

walking : walked :: swimming : swam

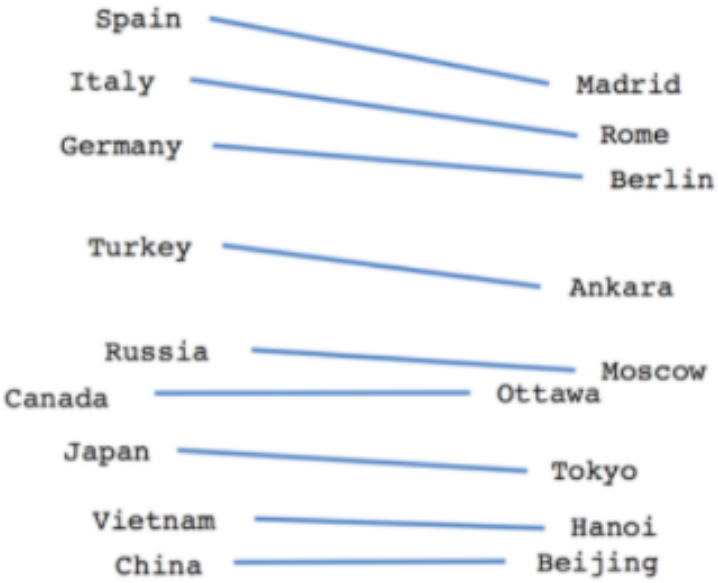
swimming – walking + walked swam



https://siddhant7.github.io/Vector-Representation-of-Words/

They “know” Facts!

Country – Capital + Spain Madrid



https://siddhant7.github.io/Vector-Representation-of-Words/

Upcoming...

Homework

- Homework 1 is up!
- No more material will be covered
- Due: January 26, 2017

Project

- Project pitch is due January 23, 2017!
- Start assembling teams now
- Tons of datasets on the “projects” page on website