

Representation Learning (and Vector Semantics)

Chapter 6

Logistic Regression Review

Discriminative approach to supervised classification

Model representation:

$$\mathbf{WX} + b$$

Model Use:

pass the sum through a sigmoid function to generate a probability, then threshold to make a decision

Model Learning:

learned from a labeled training set via a loss function that must be minimized, e.g., by using (iterative) gradient descent to find the optimal weights, and regularization to avoid overfitting

An example

- I hated this movie.
 - Negative
- This movie was great, I would watch again.
 - Positive
- I found this movie to be a letdown.
 - Negative
-
- I found this movie to be a great failure.
 - Negative

What do words mean?

First thought: look in a dictionary

Words, Lemmas, Senses, Definitions

pepper, *n.*

lemma **sense** **definition**

Pronunciation: *Brit.* /ˈpeɪpə/, *U.S.* /ˈpeɪpər/

Forms: OE *peopor* (rare), OE *pipcer* (transmission error), OE *pipor*, OE *pipur* (rare)

Frequency (in current use):

Etymology: A borrowing from Latin. **Etymon:** Latin *piper*.
 < classical Latin *piper*, a loanword < Indo-Aryan (as is ancient Greek *πιπερι*); compare *Sa*

1. The spice or the plant.

1.

a. A hot pungent spice derived from the prepared fruits (peppercorns) of the pepper plant, *Piper nigrum* (see sense 2a), used from early times to season food, either whole or ground to powder (often in association with salt). Also (locally, chiefly with distinguishing word): a similar spice derived from the fruits of certain other species of the genus *Piper*; the fruits themselves.

The ground spice from *Piper nigrum* comes in two forms, the more pungent **black pepper**, produced from black peppercorns, and the milder **white pepper**, produced from white peppercorns: see **BLACK adj.** and **WHITE adj.** Special uses 5a, **reference n.** 1a, and **white adj.** and **n.** Special uses 7b(a).

2.

a. The plant *Piper nigrum* (family Piperaceae), a climbing shrub indigenous to South Asia and also cultivated elsewhere in the tropics, which has alternate stalked entire leaves, with pendulous spikes of small green flowers opposite the leaves, succeeded by small berries turning red when ripe. Also more widely: any plant of the genus *Piper* or the family Piperaceae.

b. *U.S.* with distinguishing word: any of numerous plants of other families having hot pungent fruits or leaves which resemble pepper (**1a**) in taste and in some cases are used as a substitute for it.

c. *U.S.* The California pepper tree, *Schinus molle*. Cf. **PEPPER TREE n.** 3.

3. Any of various forms of capsicum, esp. *Capsicum annuum* var. *annuum*. Originally (chiefly with distinguishing word): any variety of the *C. annuum* Longum group, with elongated fruits having a hot, pungent taste, the source of cayenne, chilli powder, paprika, etc., or of the perennial *C. frutescens*, the source of Tabasco sauce. Now frequently (more fully **sweet pepper**): any variety of the *C. annuum* Grossum group, with large, bell-shaped or apple-shaped, mild-flavoured fruits, usually ripening to red, orange, or yellow and eaten raw in salads or cooked as a vegetable. Also: the fruit of any of these capsicums.

Sweet peppers are often used in their green immature state (more fully **green pepper**), but some new varieties remain green when ripe.

Lemma pepper

Sense 1: spice from pepper plant

Sense 2: the pepper plant itself

Sense 3: another similar plant (Jamaican pepper)

Sense 4: another plant with peppercorns
(California pepper)

Sense 5: *capsicum* (i.e. chili, paprika, bell pepper, etc)

A sense or “concept” is the meaning component of a word

There are relations between senses

Relation: Synonymy

Synonyms have the same meaning in some or all contexts.

- filbert / hazelnut
- couch / sofa
- big / large
- automobile / car
- vomit / throw up
- Water / H₂O

Relation: Synonymy

Note that there are probably no examples of perfect synonymy.

- Even if many aspects of meaning are identical
- Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.

The Linguistic Principle of Contrast:

- Difference in form -> difference in meaning

Relation: Synonymy?

Water/H₂O

Big/large

Brave/courageous

Relation: Antonymy

Senses that are opposites with respect to one feature of meaning

Otherwise, they are very similar!

dark/light short/long fast/slow rise/fall
hot/cold up/down in/out

More formally: antonyms can

- define a binary opposition
 or be at opposite ends of a scale
- long/short, fast/slow
- Be *reversives*:
 - rise/fall, up/down

Relation: Similarity

Words with similar meanings. Not synonyms, but sharing some element of meaning

car, bicycle

cow, horse

Ask humans how similar 2 words are

word1	word2	similarity
vanish	disappear	9.8
behave	obey	7.3
belief	impression	5.95
muscle	bone	3.65
modest	flexible	0.98
hole	agreement	0.3

Relation: Word relatedness

Also called "word association"

Words be related in any way, perhaps via a semantic frame or field

- car, bicycle: **similar**
- car, gasoline: **related**, not similar

Semantic field

Words that

- cover a particular semantic domain
- bear structured relations with each other.

hospitals

surgeon, scalpel, nurse, anaesthetic, hospital

restaurants

waiter, menu, plate, food, menu, chef

houses

door, roof, kitchen, family, bed

Relation: Superordinate/ subordinate

One sense is a **subordinate** of another if the first sense is more specific, denoting a subclass of the other

- *car* is a subordinate of *vehicle*
- *mango* is a subordinate of *fruit*

Conversely **superordinate**

- *vehicle* is a superordinate of *car*
- *fruit* is a superordinate of *mango*

Superordinate	vehicle	fruit	furniture
Subordinate	car	mango	chair

These levels are not symmetric

One level of category is distinguished from the others

The "basic level"

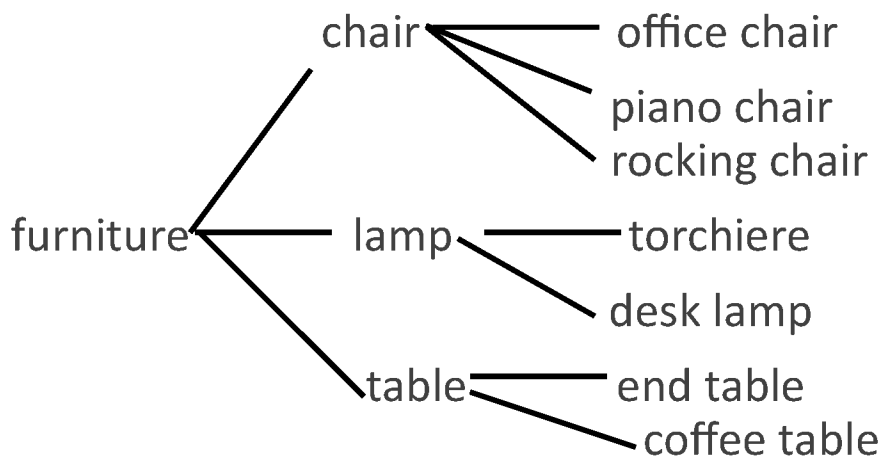
Name these items



Superordinate

Basic

Subordinate



Cluster of Interactional Properties

Basic level things are “human-sized”

Consider chairs

- We know how to interact with a chair (sitting)
- Not so clear for superordinate categories like furniture
 - “Imagine a furniture without thinking of a bed/table/chair/specific basic-level category”

The basic level

Is the level of distinctive actions

Is the level which is learned earliest and at which things are first named

It is the level at which names are shortest and used most frequently

Connotation

Words have **affective** meanings

positive connotations (*happy*)

negative connotations (*sad*)

positive evaluation (*great, love*)

negative evaluation (*terrible, hate*).

Concepts or word senses so far

Have a complex many-to-many association with **words**

- Homonymy
 - Lexemes share a form (phonological, orthographic) but have unrelated meanings
 - **Bank** (financial institution) versus **bank** (riverside)
 - Polysemy (multiple senses)
 - A single lexeme with multiple **related** meanings (**bank** the building, **bank** the financial institution)
 - The **bank** is constructed from red brick
 - I withdrew the money from the **bank**

Have relations with each other

- Synonymy
- Antonymy
- Similarity
- Relatedness
- Superordinate/subordinate
- Connotation

But how to define a concept?

Classical (“Aristotelian”) Theory of Concepts

The meaning of a word:

a concept defined by **necessary** and **sufficient** conditions

A **necessary** condition for being an X is a condition C that X must satisfy in order for it to be an X.

- If not C, then not X
- “Having four sides” is necessary to be a square.

A **sufficient** condition for being an X is condition such that if something satisfies condition C, then it must be an X.

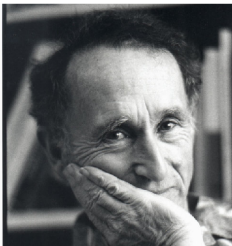
- If and only if C, then X
- The following necessary conditions, jointly, are sufficient to be a square
 - x has (exactly) four sides
 - each of x's sides is straight
 - x is a closed figure
 - x lies in a plane
 - each of x's sides is equal in length to each of the others
 - each of x's interior angles is equal to the others (right angles)
 - the sides of x are joined at their ends

Example
from
Norman
Swartz,
SFU

Problem 1: The features are complex and may be context-dependent

William Labov. 1975

Cup or bowl?

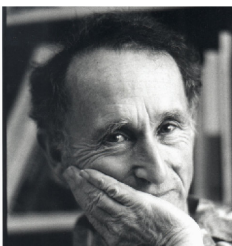


Problem 1: The features are complex and may be context-dependent

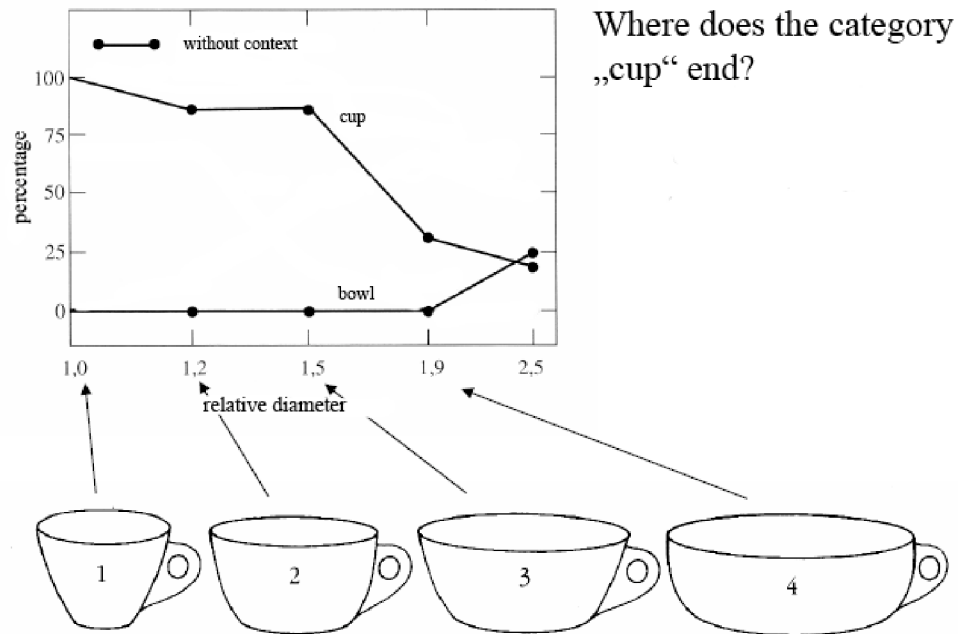
William Labov. 1975

What are these?

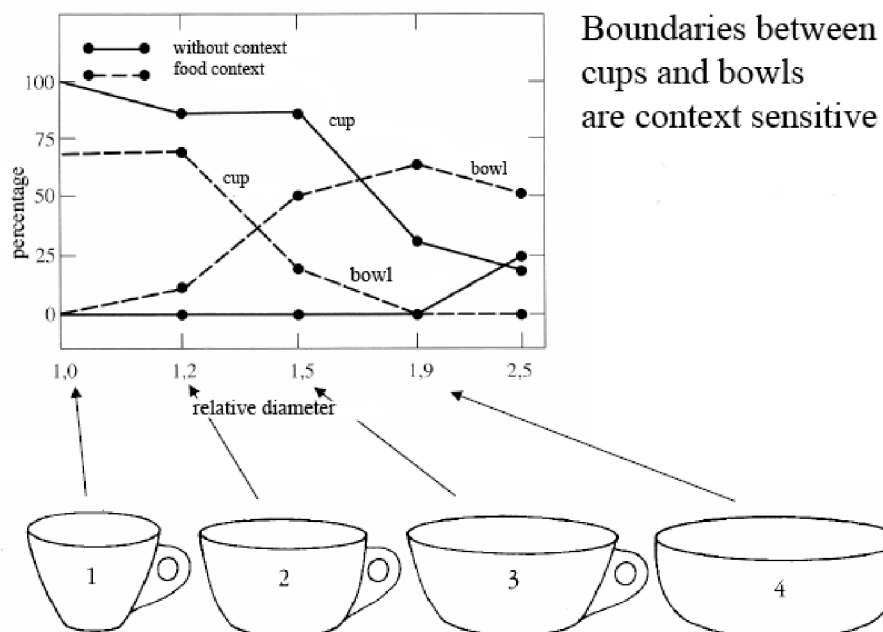
Cup or bowl?



The category depends on complex features of the object (diameter, etc)



The category depends on the context!
(If there is food in it, it's a bowl)



Labov's definition of cup

The term *cup* is used to denote round containers with a ratio of depth to width of $1 \pm r$ where $r \leq r_b$, and $r_b = \alpha_1 + \alpha_2 + \dots + \alpha_n$ and α_i is a positive quality when the feature i is present and 0 otherwise.

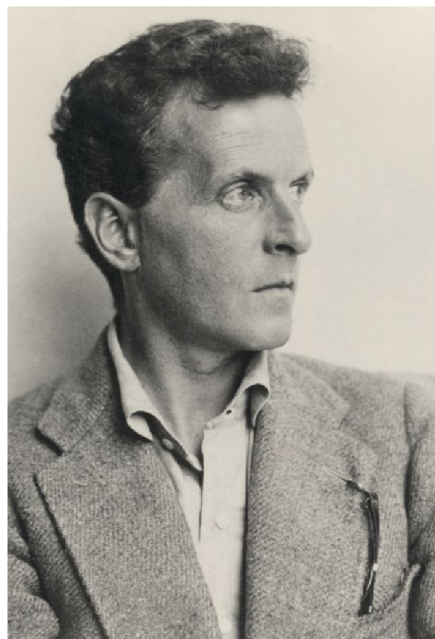
- feature
- 1 = with one handle
 - 2 = made of opaque vitreous material
 - 3 = used for consumption of food
 - 4 = used for the consumption of liquid food
 - 5 = used for consumption of hot liquid food
 - 6 = with a saucer
 - 7 = tapering
 - 8 = circular in cross-section

Cup is used variably to denote such containers with ratios width to depth $1 \pm r$ where $r_b \leq r \leq r_1$ with a probability of $r_1 - r/r_1 - r_b$. The quantity $1 \pm r_b$ expresses the distance from the modal value of width to height.

Ludwig Wittgenstein (1889-1951)

Philosopher of
language

In his late years, a
proponent of studying
"ordinary language"



Wittgenstein (1945)
*Philosophical
 Investigations.*
 Paragraphs 66,67

66. Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: "There *must* be something common, or they would not be called 'games'?"—but *look and see* whether there is anything common to all.—For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!—Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost.—Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! And we can go through the many, many other groups of games in the same way; can see how similarities crop up and disappear.

And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail.

67. I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way.—And I shall say: 'games' form a family.

And for instance the kinds of number form a family in the same way. Why do we call something a "number"? Well, perhaps because it has a—direct—relationship with several things that have hitherto been called number; and this can be said to give it an indirect relationship to other things we call the same name. And we extend our concept of number as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres.

But if someone wished to say: "There is something common to all these constructions—namely the disjunction of all their common properties"—I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread—namely the continuous overlapping of those fibres".

What is a game?

Wittgenstein's thought experiment on "What is a game":

PI #66:

"Don't say "there must be something common, or they would not be called `games'" —but *look and see* whether there is anything common to all"

- Is it amusing?
- Is there competition?
- Is there long-term strategy?
- Is skill required?
- Must luck play a role?
- Are there cards?
- Is there a ball?

Family Resemblance

Game 1	Game 2	Game 3	Game 4
ABC	BCD	ACD	ABD

"each item has at least one, and probably several, elements in common with one or more items, but no, or few, elements are common to all items" Rosch and Mervis

How about a radically different approach?

Ludwig Wittgenstein

PI #43:

"The meaning of a word is its use in the language"

Let's define words by their usages

In particular, words are defined by their environments (the words around them)

Zellig Harris (1954): **If A and B have almost identical environments we say that they are synonyms.**

What does ong choi mean?

Suppose you see these sentences:

- Ong choi is delicious **sautéed with garlic**.
- Ong choi is superb **over rice**
- Ong choi **leaves** with salty sauces

And you've also seen these:

- ...spinach **sautéed with garlic over rice**
- Chard stems and **leaves** are **delicious**
- Collard greens and other **salty** leafy greens

Conclusion:

- Ongchoi is a leafy green like spinach, chard, or collard greens

Ong choi: *Ipomoea aquatica* "Water Spinach"



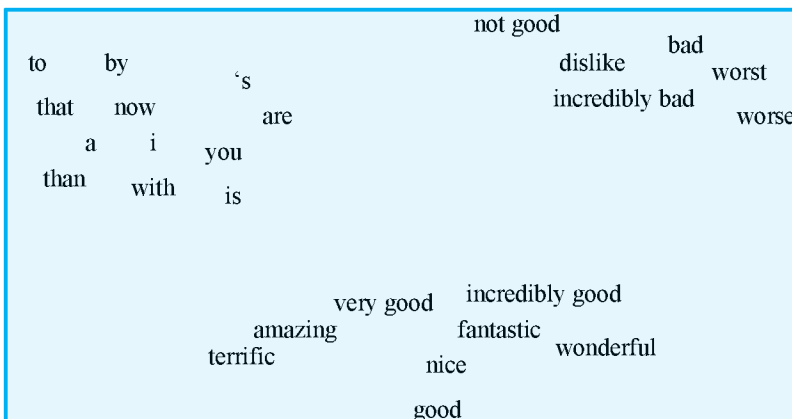
Yamaguchi, Wikimedia Commons, public domain

We'll build a new model of meaning focusing on similarity

Each word = a vector

- Not just "word" or word45.

Similar words are "nearby in space"



We define a word as a vector

Called an "embedding" because it's embedded into a space

The (fairly recent) standard way to represent meaning in NLP

Fine-grained model of meaning for similarity

- NLP tasks like sentiment analysis
 - With words, requires **same** word to be in training and test
 - With embeddings: ok if **similar** words occurred!!!

We'll introduce 2 kinds of embeddings

Tf-idf

- A common baseline model
- Sparse vectors
- Words are represented by a simple function of the counts of nearby words

Word2vec

- Dense vectors
- Representation is created by training a classifier to distinguish nearby and far-away words

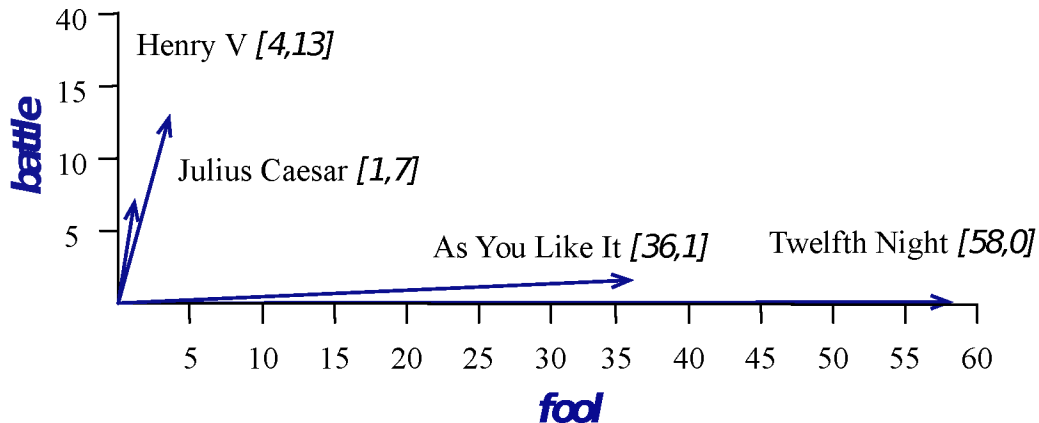
Words, vectors, and co-occurrence matrices

Term-document matrix

Each document is represented by a vector of words

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Visualizing document vectors



Vectors are the basis of information retrieval

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

- “Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).” *Stanford NLP*
- Vectors are similar for the two comedies and different than the history
- Comedies have more fools and wit and fewer battles.

Words can be vectors too

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

battle is "the kind of word that occurs in Julius Caesar and Henry V"

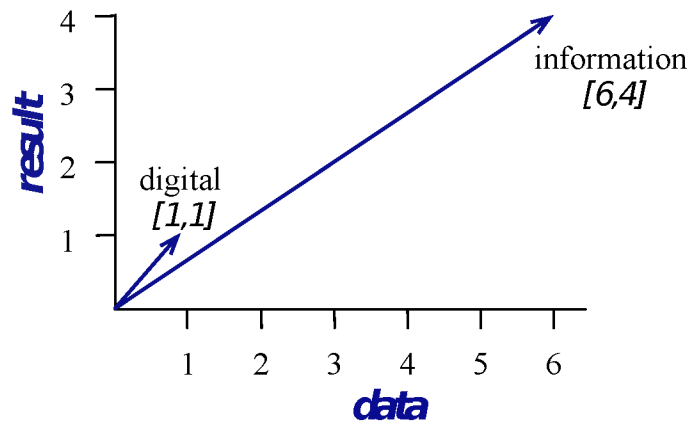
fool is "the kind of word that occurs in comedies, especially Twelfth Night"

More common: word-word matrix (or "term-context matrix")

Two **words** are similar in meaning if their context vectors are similar

sugar, a sliced lemon, a tablespoonful of their enjoyment. Cautiously she sampled her first well suited to programming on the digital for the purpose of gathering data and **apricot** **pineapple** **computer** **information** jam, a pinch each of, and another fruit whose taste she likened In finding the optimal R-stage policy from necessary for the study authorized in the

	aardvark	computer	data	pinch	result	sugar	...
apricot	0	0	0	1	0	1	
pineapple	0	0	0	1	0	1	
digital	0	2	1	0	1	0	
information	0	1	6	0	4	0	



Reminders from linear algebra

$$\text{dot-product}(\vec{v}, \vec{w}) = \vec{v} \cdot \vec{w} = \sum_{i=1}^N v_i w_i = v_1 w_1 + v_2 w_2 + \dots + v_N w_N$$

$$\text{vector length } |\vec{v}| = \sqrt{\sum_{i=1}^N v_i^2}$$

Cosine for computing similarity

Sec. 6.3

$$\text{cosine}(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} = \frac{\sum_{i=1}^N v_i w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}}$$

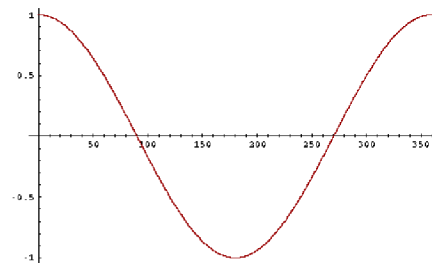
v_i is the count for word v in context i
 w_i is the count for word w in context i .

Cosine as a similarity metric

-1: vectors point in opposite directions

+1: vectors point in same directions

0: vectors are orthogonal



Frequency is non-negative, so cosine range 0-1

	large	data	computer
apricot	1	0	0
digital	0	1	2
information	1	6	1

$$\cos(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} = \frac{\sum_{i=1}^N v_i w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}}$$

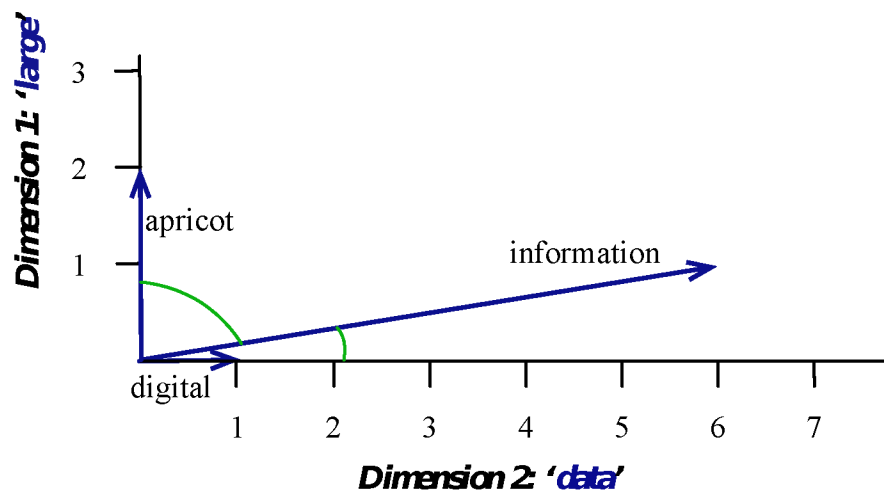
Which pair of words is more similar?

$\cosine(\text{apricot}, \text{information}) = \frac{1+0+0}{\sqrt{1+0+0} \sqrt{1+36+1}} = \frac{1}{\sqrt{38}} \approx .16$

$\cosine(\text{digital}, \text{information}) = \frac{0+6+2}{\sqrt{0+1+4} \sqrt{1+36+1}} = \frac{8}{\sqrt{38}\sqrt{5}} \approx .58$

$\cosine(\text{apricot}, \text{digital}) = \frac{0+0+0}{\sqrt{1+0+0} \sqrt{0+1+4}} = 0$

Visualizing cosines (well, angles)



But raw frequency is a bad representation

Frequency is clearly useful; if *sugar* appears a lot near *apricot*, that's useful information.

But overly frequent words like *the*, *it*, or *they* are not very informative about the context

Need a function that resolves this frequency paradox!

tf-idf: combine two factors

tf: term frequency. frequency count (usually log-transformed):

$$\text{tf}_{t,d} = \begin{cases} 1 + \log_{10} \text{count}(t,d) & \text{if } \text{count}(t,d) > 0 \\ 0 & \text{otherwise} \end{cases}$$

idf: inverse document frequency: tf-

$$\text{idf}_i = \log \left(\frac{N}{\text{df}_i} \right)$$

Total # of docs in collection

Words like "the" or "good" have very low idf

of docs that have word i

tf-idf value for word t in document d :

$$w_{t,d} = \text{tf}_{t,d} \times \text{idf}_t$$

Summary: tf-idf

Compare two words using tf-idf cosine to see if they are similar

An alternative to tf-idf

Ask whether a context word is **particularly informative** about the target word.

- Positive Pointwise Mutual Information (PPMI)

Pointwise Mutual Information

Pointwise mutual information:

Do events x and y co-occur more than if they were independent?

$$\text{PMI}(X, Y) = \log_2 \frac{P(x, y)}{P(x)P(y)}$$

PMI between two words: (Church & Hanks 1989)

Do words x and y co-occur more than if they were independent?

	Count(w, context)				
	computer	data	pinch	result	sugar
apricot	0	0	1	0	1
pineapple	0	0	1	0	1
digital	2	1	0	1	0
information	1	6	0	4	0

	PPMI(w, context)				
	computer	data	pinch	result	sugar
apricot	-	-	2.25	-	2.25
pineapple	-	-	2.25	-	2.25
digital	1.66	0.00	-	0.00	-
information	0.00	0.57	-	0.47	-

Summary

- Survey of Lexical Semantics
- Idea of Embeddings: Represent a word as a function of its distribution with other words
- Tf-idf
- Cosines
- PPMI (barely)