DISTRIBUTIONAL SEMANTICS AND COMPOSITIONALITY

Corina Dima

April 23rd, 2019
COURSE LOGISTICS

➤ Who?
  ➤ Corina Dima
  ➤ office: 1.05, Wilhelmstr. 19
  ➤ email: corina.dima@uni-tuebingen.de
  ➤ office hours: Tuesdays, 14-15 (please email me first)

➤ When?
  ➤ Tuesdays, 8:30-10 (DS)
  ➤ Thursdays, 8:30-10 (Comp)

➤ Where?
  ➤ Room 1.13, Wilhelmstr. 19

➤ What?
  ➤ Course webpage: https://dscomp2019.github.io/
Word representations (word embeddings) based on distributional information are a key ingredient for state-of-the-art natural language processing applications.

They represent similar words like ‘cappuccino’ and ‘espresso’ as similar vectors in vector space. Dissimilar vectors - like ‘cat’ - are far away.
What makes \( p \) most similar to \( w \)?

\[
f(\begin{bmatrix} 0.3 & 0.1 & 0.7 \\ \end{bmatrix} u, \begin{bmatrix} 0.5 & 0.9 & 0.1 \\ \end{bmatrix} v) = \begin{bmatrix} \_ & \_ & \_ & \_ \\ \end{bmatrix} \text{ p}
\]

**Compositionality**

- **Composition models** for distributional semantics extend the vector spaces by learning how to create representations for complex words (e.g. ‘apple tree’) and phrases (e.g. ‘black car’) from the representations of individual words.

- The course will cover several approaches for creating and composing distributional word representations.
COURSE PREREQUISITES

➤ Prerequisites

➤ **linear algebra** (matrix-vector multiplications, dot product, Hadamard product, vector norm, unit vectors, cosine similarity, cosine distance, matrix decomposition, orthogonal and diagonal matrices, tensor, scalar)

➤ **programming** (Java III), **computational linguistics** (Statistical NLP) - ISCL-BA-08 or equivalent; programming in Python (+numpy, Tensorflow/PyTorch) for the project

➤ **machine learning** (regression, classification, optimization objective, dropout, recurrent neural networks, autoencoders, convolutions)
GRADING

- For 6 CP
  - Active participation in class (30%)
  - Presenting a paper (70%)

- For 9 CP
  - Active participation in class (30%)
  - Doing a project (paper(s)-related) & writing a paper (70%)

- **Strict deadline** for the project: end of lecture time (27.07.2019)

- Both presentations and projects are individual
REGISTRATION

➤ register using your GitHub account until 29.04.2019

➤ Info

➤ Last name(s)
➤ First name(s)
➤ Email address
➤ Native language(s)
➤ Other natural languages
➤ Programming languages
➤ Student ID (Matrikelnr.)
➤ Degree program, semester (e.g. ISCL BA, 5th semester)
➤ Chosen variant of the course: 6CP/9CP
EXAMPLE PROJECTS (1)

➤ Implement a PMI-based tool for the automatic discovery of English noun-noun compounds in a corpus. The tool should be able to discover both two-part as well as multi-part compounds.

➤ References:

➤ Church & Hanks (1990) - Word Association Norms, Mutual Information and Lexicography

➤ Mikolov et al. (2013) - Distributed Representations of Words and Phrases and their Compositionality
EXAMPLE PROJECTS (2)

➢ Implement a recursive composition model that uses subword representations.


➢ recursively compose each two ngrams, each time replacing the two composed ngrams with the composed representation

➢ References:


➢ Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher Manning, Andrew Ng, Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.
NEXT WEEK

➤ Tuesday, 30.04 (DS)


➤ Thursday, 2.05 (COMP)

➤ Jeff Mitchell and Mirella Lapata. 2010. Composition in Distributional Models of Semantics (Corina)
IN TWO WEEKS

- **Tuesday, 7.05 (DS)**
  - Kenneth Church and Patrick Hanks. 1990. *Word Association Norms, Mutual Information and Lexicography* (?)

- **Thursday, 9.05 (COMP)**
  - Emiliano Guevara. 2010. *A Regression Model of Adjective-Noun Compositionality in Distributional Semantics* (?)
  - Marco Baroni and Roberto Zamparelli. 2010. *Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space* (?)
HOW TO WRITE A RESEARCH PAPER

➤ Jason Eisner’s blog post Write the paper first (https://www.cs.jhu.edu/~jason/advice/write-the-paper-first.html)

➤ “Writing is the best use of limited time”

➤ “If you run out of time, it is better to have a great story with incomplete experiments than a sloppy draft with complete experiments”

➤ “Writing is a form of thinking and planning. Writing is therefore part of the research process—just as it is part of the software engineering process. When you write a research paper, or when you document code, you are not just explaining the work to other people: you are thinking it through for yourself.”
HOW TO READ A RESEARCH PAPER

  ➤ multi-pass reading (skim first, more thorough second pass)
  ➤ write as you read (low-level notes, high-level notes)
  ➤ start early!

➤ Michael Nielsen’s blog post *Augmenting Long-Term Memory* (http://augmentingcognition.com/ltm.html)
  ➤ Using Anki to thoroughly read research papers (++)remember)
EBBINHAUS’S FORGETTING CURVE
LEARNING HOW TO LEARN

➤ Barbara Oakely & Terrence Sejnowski’s *Learning how to learn* course on Coursera (https://www.coursera.org/learn/learning-how-to-learn)

➤ Main points:

➤ learning doesn’t happen overnight - you need several passes through some material to really understand it

➤ re-reading/highlighting materials can give you the illusion of learning - avoid it by practicing active recall (testing yourself)

➤ spaced repetition can help you learn & remember forever-ish
THE EFFECTS OF SPACED REPETITION ON THE FORGETTING CURVE

- Percentage of materials retained
- Interval for repeating the materials

Line graph showing the percentage of materials retained over different intervals after a class.
HELPFUL POINTERS

➤ Khan Academy’s Linear Algebra course (https://www.khanacademy.org/math/linear-algebra)

➤ Dan Jurafsky and James H. Martin. *Speech and Language Processing*. 3rd edition draft (https://web.stanford.edu/~jurafsky/slp3/), esp. Ch. 6, Vector Semantics
What does a word mean?

cappuccino  | ˌkæpəˈtʃiːnə |

noun (plural cappuccinos)

a type of coffee made with espresso and milk that has been frothed up with pressurized steam.

ORIGIN

from Italian, literally ‘Capuchin’, because its colour resembles that of a Capuchin’s habit.
INTRO TO DISTRIBUTIONAL SEMANTICS

➤ How can the meaning of a word be represented on a computer?

➤ One-hot vectors

➤ each word is represented by a 1 in a particular dimension of the vector, with the other elements of the vector being 0

➤ local representation: no interaction between the different dimensions

[1, 0, 0]

[0, 1, 0]

[0, 0, 1]
Local representations, problem 1: word similarity does not correspond to vector similarity

‘cappuccino’ and ‘espresso’ are just as similar/dissimilar as ‘cappuccino’ and ‘cat’

one-hot vectors are orthogonal to each other
INTRO TO DISTRIBUTIONAL SEMANTICS

➤ measure **cosine similarity** in vector space

\[
\cos(u, v) = \frac{u \cdot v}{\|u\|_2 \|v\|_2} = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2} \sqrt{\sum_{i=1}^{n} v_i^2}}
\]

\[
\cos([1, 0, 0], [0, 1, 0]) = \frac{1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0}{\sqrt{1^2 + 0^2 + 0^2} \sqrt{0^2 + 1^2 + 0^2}} = \frac{0}{1} = 0
\]

\[
\cos([0, 0, 1]) = \frac{1 \cdot 0 + 0 \cdot 0 + 0 \cdot 1}{\sqrt{1^2 + 0^2 + 0^2} \sqrt{0^2 + 0^2 + 1^2}} = \frac{0}{1} = 0
\]

**cosine of 0 means angle of 90° between the vectors**

➤ **orthogonal vectors**
Local representations, problem 2: representing new words

- representing a new word involves expanding the vector, since the existing components are already “used up”
INTRO TO DISTRIBUTIONAL SEMANTICS

➤ Solution: distributed representations (Hinton, McClelland and Rumelhart, 1986)

➤ meaning is distributed over the different dimensions of the vector

➤ each word is represented by a configuration over the components of the vector representations

➤ each component contributes to the representation of every word in the vocabulary

[0.37, -0.93]

[0.45, -0.89]

[-0.92, 0.39]
INTRO TO DISTRIBUTIONAL SEMANTICS

[0.37, -0.93]
[0.45, -0.89]
[-0.92, 0.39]
Distributed representations solve problem 1: similar words can have similar vectors.

\[
\cos([0.37, -0.93], [0.45, -0.89]) = \frac{0.37 \cdot 0.45 + (-0.93) \cdot (-0.89)}{\sqrt{0.37^2 + (-0.93)^2} \sqrt{0.45^2 + (-0.89)^2}} \approx 0.9965
\]

\[
\cos([0.37, -0.93], [-0.92, 0.39]) = \frac{0.37 \cdot (-0.92) + (-0.93) \cdot 0.39}{\sqrt{0.37^2 + (-0.93)^2} \sqrt{(-0.92)^2 + 0.39^2}} \approx -0.7071
\]

\[y(x) = \cos(x)\]
INTRO TO DISTRIBUTIONAL SEMANTICS

- **similar vectors**
  - Angle is 0°
  - Cosine similarity is 1

- **orthogonal vectors**
  - Angle is 90°
  - Cosine similarity is 0

- **opposite vectors**
  - Angle is 180°
  - Cosine similarity is -1
Distributed representations solve problem 2: new words can be added to the vector space without changing the dimensions of the vectors.

- [0.37, -0.93]
- [0.45, -0.89]
- [-0.92, 0.39]
- [0.32, -0.95]
What information can be used to create the (local/distributed) word representations?

Distributional semantics

- **Harris (1954):** “Meaning as a function of distribution”
- **Firth (1957):** “You shall know a word by the company it keeps!”
If we consider *oculist* and *eye-doctor* we find that, as our corpus of actually-occurring utterances grows, these two occur in almost the same environments, except for such sentences as *An oculist is just an eye-doctor under a fancier name*, or *I told him Burns was an oculist, but since he didn’t know the professional titles, he didn’t realize that he could go to him to have his eyes examined*. If we ask informants for any words that may occupy the same place as *oculist* in sentences like the above (i.e. have these same environments), we will not in general obtain *eye-doctor*; but in almost any other sentence we would. In contrast, there are many sentence environments in which *oculist* occurs but *lawyer* does not: e.g. *I’ve had my eyes examined by the same oculist for twenty years*, or *Oculists often have their prescription blanks printed for them by opticians*.

- Zelling S. Harris (1954)
The *placing* of a text as a constituent in a context of situation contributes to the statement of meaning since situations are set up to recognize *use.* As Wittgenstein says, ‘the meaning of words lies in their use.’

The day to day practice of playing language games recognizes customs and rules. It follows that a text in such established usage may contain sentences such as ‘Don’t be such an ass!’; ‘You silly ass!’; ‘What an ass he is!’ In these examples, the word *ass* is in familiar and habitual company, commonly collocated with *you silly*, *he is a silly*, *don’t be such an*. You shall know a word by the company it keeps! One of the meanings of *ass* is its habitual collocation with such other words as those above quoted. Though Wittgenstein was dealing with another problem, he also recognizes the plain face-value, the physiognomy of words. They look at us! ‘The sentence is composed of the words and that is enough.’

-J.R. Firth (1957)
We found a cute, hairy wampimuk sleeping behind the tree.

Lazaridou et. al, 2014
We found a cute, hairy *wampimuk* sleeping behind the tree.

*Lazaridou et. al, 2014*
INTRO TO DISTRIBUTIONAL SEMANTICS

36 MAG the waitress in a neat black and white uniform. My cappuccino came with the correct amount of froth, sprinkled with chocolate.

37 NEWS ice cream in whimsical flavors like white pistachio and cappuccino chocolate truffles. The pair soon owned a string of stores, chocolate and a headless mannequin wearing a two-colored carpeting.

38 FIC 's an expensive wedding cake of a store, adorned with cappuccino cappuccino and a headless mannequin wearing a two-colored carpeting.

39 MAG from halfway around the world. Everyone agrees that the best cappuccino comes from the caf run by the French soldiers, but it.

40 ACAD added to create a chocolatey beverage. # Like today's cappuccino connoisseurs elite. Maya and Aztec cherished foam atop their.

41 FIC . # I did n't ask the natural next question. Cappuccino cream mustache on her upper lip, she volunteered. # "

42 NEWS you 're rewarded with a lush, velvety custard. The cappuccino creme brulee or cappuccino served in a coffee cup had.

43 NEWS for a custard moussage affair). Served in an oversized cappuccino cup the rich and velvety custard was topped with a twirl.

44 SPOK OK. Ms. RAY: she serves it in little espresso cups or cappuccino cups. So here 's all you do. Here, you.

45 MAG be gratis. COFFEE. With no more office pot, cappuccino doses run $ a day -- that 's up to $1,460 a.

46 MAG Scottie 's after school as the waiter slammed two mugs of cappuccino down in front of Meghan and me. She had been after.

47 FIC blackmail me into doing a job. # Tommy breaks and cappuccino cappuccino flies Hawk half-heartedly tries to lick up with his fingers. For all. Everyone except Malovich lights up strong European.

48 NEWS On this early morn, it 's all for one and cappuccino cappuccino flies Hawk half-heartedly tries to lick up with his fingers. For all. Everyone except Malovich lights up strong European.

49 MAG bag. To avoid feeling deprived, sip a frothy skim milk.

50 MAG satisfying drinks-like root-beer floats for kids or an iced cappuccino cappuccino for grown-ups and you 're dishing the kind of bliss that your.

51 FIC have to. " # Lyric and I get ice decaf cappuccinos from the store next to the condo. We get an Orangina.

52 SPOK subtle movement, his body temperature began to rise. And Cappuccino gambled again using a cooling machine to lower Everett 's body.

53 NEWS From a Grande at Starbucks to a plain cup of black no-sugar at a.

54 SPOK Starbucks. To a plain cup of black no-sugar at a.

55 MAG made my next change: I gave up the sugary convenience-store cappuccinos 1 c de beer drinking several times a day for lower-calorie vanilla.

56 NEWS away from slides of pan, 25-30 minutes. 71915 # Cappuccino Icebox Cookies # You can keep a roll of this dough in.

57 MAG cozy fireplaces and live Andean music. We stop for a cappuccino in b corner bar that advertises 15 types of coffee. But.

58 NEWS Norte, his city's main newspaper, while sipping a cappuccino in anticipation of a shopping spree. " Settling down in McAllen.

59 SPOK guess, makers -- oh, look, oh, sue cappuccino in general with their faces in the foam. GIFORD: Oh.

60 MAG their A-list friends in trendy clubs, they prefer sipping soy cappuccinos in local cafes. Moder is an outdoorys guy who enjoys running.

61 SPOK theaters adapt and they now have multi-screen and now they have cappuccinos in some movie theaters. So they adapt, and I think

https://www.wordandphrase.info/, made by Mark Davies, BYU Corpus of Contemporary American English (COCA)
**INTRO TO DISTRIBUTIONAL SEMANTICS**

**CAPPUCCINO** *n* (RANK 17250, FREQ 595)

<table>
<thead>
<tr>
<th>CLICK BAR TO LIMIT</th>
<th>SPOKEN</th>
<th>FICTION</th>
<th>MAGAZINE</th>
<th>NEWSPAPER</th>
<th>ACADEMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>STORED</td>
<td>21</td>
<td>56</td>
<td>61</td>
<td>57</td>
<td>7</td>
</tr>
<tr>
<td>MORE</td>
<td>56</td>
<td>200</td>
<td>180</td>
<td>144</td>
<td>15</td>
</tr>
</tbody>
</table>

*adj* iced, double, orthopedic, frothy, hot, instant, steaming, tall, excellent, fat-free *noun* cup, machine, espresso, latte, bar, coffee, sip, cafe, maker, decaf *verb* sip, drink, serve, order, buy, sell, finish, enjoy
# INTRO TO DISTRIBUTIONAL SEMANTICS

## Co-occurrence Matrix

<table>
<thead>
<tr>
<th>Context Words</th>
<th>Iced</th>
<th>(to) Drink</th>
<th>Owner</th>
<th>In</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cappuccino</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Espresso</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Cat</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Latte</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Leaf</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>
- the pointwise mutual information (PMI) between a target word $t$ and a context word $c$ is defined as

$$PMI(t, c) = \log_2 \frac{P(t, c)}{P(t)P(c)}$$
- the pointwise mutual information (PMI) between a target word \( t \) and a context word \( c \) is defined as

\[
PMI(t, c) = \log_2 \frac{P(t, c)}{P(t)P(c)}
\]

how often are \( t \) and \( c \) are observed together
- the pointwise mutual information (PMI) between a target word \( t \) and a context word \( c \) is defined as

\[
PMI(t, c) = \log_2 \frac{P(t, c)}{P(t)P(c)}
\]

how often are \( t \) and \( c \) are observed together

how often would we expect \( t \) and \( c \) to co-occur
(assuming each occurs independently)
- the **pointwise mutual information (PMI)** between a target word $t$ and a context word $c$ is defined as

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

how often are $t$ and $c$ are observed **together**

the ratio is an estimate of how much more the two words co-occur than is expected by chance

how often would we expect $t$ and $c$ to co-occur (assuming each occurs independently)
INTRO TO DISTRIBUTIONAL SEMANTICS

- the PMI for ‘Humpty Dumpty’ is 22.5
- the pair (Humpty, Dumpty) occurs 6,000,000 (~$2^{22.5}$) times more than one would expect from the frequencies of Humpty and Dumpty - from Brown et al. (1992)
- order matters!
- $\text{PMI}(\text{Humpty, Dumpty}) \neq \text{PMI}(\text{Dumpty, Humpty})$
- positive point wise mutual information (PPMI) is used
# Intro to Distributional Semantics

<table>
<thead>
<tr>
<th></th>
<th>iced</th>
<th>(to) drink</th>
<th>owner</th>
<th>p(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cappuccino</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>espresso</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>cat</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>p(c)</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>

\[
P(t = \text{cappuccino}, c = \text{iced}) = \frac{6}{15} = 0.4
\]

\[
P(t = \text{cappuccino}) = \frac{8}{15} = 0.53
\]

\[
P(c = \text{iced}) = \frac{7}{15} = 0.47
\]

\[
PMI(t = \text{cappuccino}, c = \text{iced}) = \log_2 \frac{0.4}{0.53 \times 0.47} = \log_2 1.6 = 0.68
\]

\[
P(t = \text{cappuccino}, c = \text{iced}) = \log_2 \frac{0.4}{0.53 \times 0.47} = \log_2 1.6 = 0.68
\]
vocabularies contain typically 10,000-1,000,000 words

sparse vectors (most components are 0) - most words will co-occur with a small subset of other words in the vocabulary

use dimensionality reduction techniques to transform high-dimensional, sparse representations to low-dimensional, dense representations
INTRO TO DISTRIBUTIONAL SEMANTICS

➤ singular value decomposition (SVD)

\[
A = U \Sigma V^\top
\]

➤ where \( A \in \mathbb{R}^{m \times n} \)

➤ \( U \in \mathbb{R}^{m \times n} \) is a matrix with orthogonal columns

➤ \( \Sigma \in \mathbb{R}^{n \times n} \) is a diagonal matrix of singular values; the singular values are, by convention, ordered from the largest to the smallest

➤ \( V^\top \in \mathbb{R}^{n \times n} \) is an orthogonal matrix (\( V^{-1} = V^\top \))

➤ by taking only the top \( k \) singular values, \( k \ll n \), SVD obtains an approximation of \( A \), \( A_k \), such that the distance between the matrices (the 2-norm, \( \| A - A_k \|_2 \)) is minimized
where does the dimensionality reduction come from?

Singular value decomposition separates any matrix into simple pieces.

- \( m = 30,000; \ n = 10,000; \ k = 300 \)
- Size of initial \( A \): \( 30,000 \times 10,000 = 300,000,000 \) numbers

\[
A_k = U_k \Sigma_k V_k^T
\]
INTRO TO DISTRIBUTIONAL SEMANTICS

\[
\begin{align*}
A & = UV^T \\
& = \begin{bmatrix}
XX & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
\end{bmatrix} \\
U & = \begin{bmatrix}
XX & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
\end{bmatrix} \\
\Sigma & = \begin{bmatrix}
XX & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
\end{bmatrix} \\
V^T & = \begin{bmatrix}
XX & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
X & X & X & X & X \\
\end{bmatrix}
\end{align*}
\]

\(m = 5, \quad m = 3\)
INTRO TO DISTRIBUTIONAL SEMANTICS

\[
A = \Lambda \mathbf{u} \mathbf{v}_1^T + \Sigma \mathbf{v}_2^T + \mathbf{v}_3^T
\]

\(m = s_1, m = 3\)

\[
A = \Lambda \mathbf{u} \mathbf{v}_1^T + \Sigma \mathbf{v}_2^T + \mathbf{v}_3^T
\]
INTRO TO DISTRIBUTIONAL SEMANTICS

\[
\frac{u_2}{1 1 1 1 1 1} \begin{bmatrix} a & a & b & b & c & c \end{bmatrix} = \begin{bmatrix} a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \end{bmatrix}
\]

\[
6 \times 1 \rightarrow 6 \times 6
\]
INTRO TO DISTRIBUTIONAL SEMANTICS

\[ A = \sum \eta_u u_\lambda v_\lambda^T + \sum \eta_2 u_2 v_2^T + \sum \eta_3 u_3 v_3^T \]

\[ A \approx \sum \lambda u_\lambda v_\lambda^T + \sum \eta_2 u_2 v_2^T \]
m = 30,000; n = 10,000; k = 300

size of initial A: 30,000 x 10,000 = 300,000,000 numbers

\[
A_k = U_k \Sigma_k V_k^T
\]

size of \(A_k\): 30,000 x 300 (\(U\)) + 300 (\(\Sigma\)) + 300 x 10,000 (\(V^T\)) =
9,000,000 + 300 + 3,000,000 = 12,000,300 numbers

words can now be represented as reduced-dimensionality vectors

\[
W^{SVD}_{p} = U_k \cdot \Sigma^p_k
\]

\(W \in \mathbb{R}^{m \times k}\)

in our example \(W \in \mathbb{R}^{30,000 \times 300}\)

\(p = 0, W^{SVD} = U_k\)

\(p = \frac{1}{2}, W^{SVD} = U_k \cdot \sqrt{\Sigma_k}\)

\(p = 1, W^{SVD} = U_k \cdot \Sigma_k\)
after dimensionality reduction, a particular vector component no longer has an associated “meaning”

the information is “spread” over the dimensions

more difficult to interpret individual vector components
REFERENCES


Image credits

Creative Commons 4.0 BY-NC:
http://pngimg.com/download/49645
http://pngimg.com/download/50514
http://pngimg.com/download/27425
https://commons.wikimedia.org/wiki/File:Espresso_BW_1.jpg

CC BY-SA 2.5
https://commons.wikimedia.org/wiki/Category:Latte_macchiato?uselang=de#/media/File:Latte_macchiato_with_coffee_beans.jpg

Public domain