92586 Computational Linguistics

Lesson 7. Latent Semantic Analysis

Alberto Barrón-Cedeño

Alma Mater Studiorum-Università di Bologna
a.barron@unibo.it @_albarron_

31/03/2020

Previously

- BoW representation
- Rule-based vs Naïve Bayes classifiers (sentiment)
- *tf-idf* (+ Zipf's law)

Previously

- BoW representation
- Rule-based vs Naïve Bayes classifiers (sentiment)
- *tf-idf* (+ Zipf's law)
- Word Model $\rightarrow$ Topic Model
- Linear Discriminant Analysis

Table of Contents

Section 4.2 of Lane et al. (2019)

Introduction

Introduction

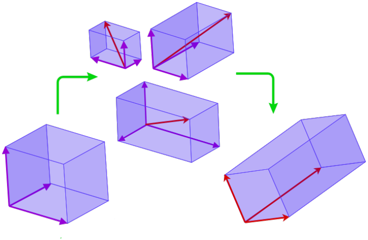
Latent semantic analysis (Lane et al., 2019, p. 112)

- A mathematical technique for finding the "best" way to linearly transform —**rotate and stretch**— any set of NLP vectors (e.g., TF-IDF, BoW)

Intuition (1)

1. Line up the axes (dimensions) in the new vectors with the greatest "spread" or variance in the word frequencies
2. Rotate the vectors so that the new dimensions (basis vectors) align with the maximum variance directions
3. Eliminate the dimensions in the new vector space that contribute the least to the variance in the vectors from document to document.

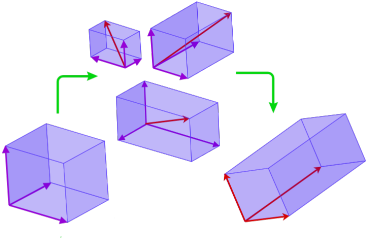
Intuition (2)



From Wikipedia: “Change of basis”

a Depart from a matrix (left)

Intuition (2)

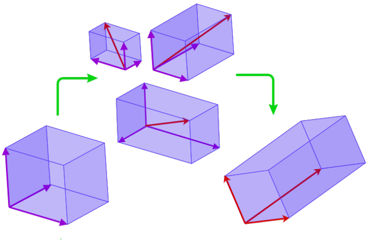


From Wikipedia: "Change of basis"

a Depart from a matrix (left)

b Decompose it into 3 simpler matrices

Intuition (2)



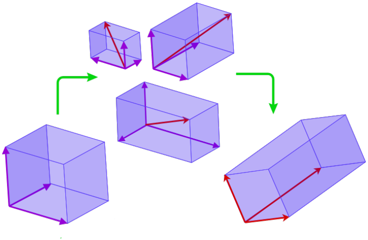
From Wikipedia: "Change of basis"

a Depart from a matrix (left)

b Decompose it into 3 simpler matrices

c Truncate the matrices

Intuition (2)



From Wikipedia: "Change of basis"

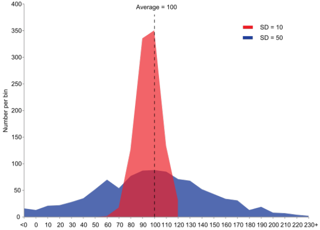
a Depart from a matrix (left)

b Decompose it into 3 simpler matrices

c Truncate the matrices

d Multiply them and produce a lower-dimensional matrix

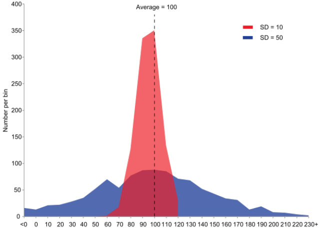
Intuition (3)



https://en.wikipedia.org/wiki/Variance

1. Line up the axes (dimensions) in the new vectors with the greatest variance in the word frequencies.

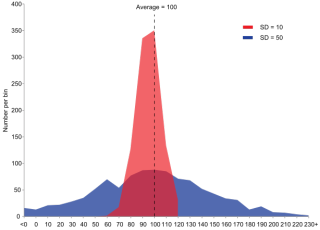
Intuition (3)



https://en.wikipedia.org/wiki/Variance

1. Line up the axes (dimensions) in the new vectors with the greatest variance in the word frequencies.
2. Rotate the vectors so that the new dimensions (basis vectors) align with the maximum variance word frequencies

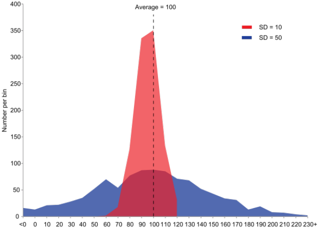
Intuition (3)



https://en.wikipedia.org/wiki/Variance

1. Line up the axes (dimensions) in the new vectors with the greatest variance in the word frequencies.
2. Rotate the vectors so that the new dimensions (basis vectors) align with the maximum variance word frequencies
3. Eliminate the dimensions that contribute the least to the variance in the vectors

Intuition (3)

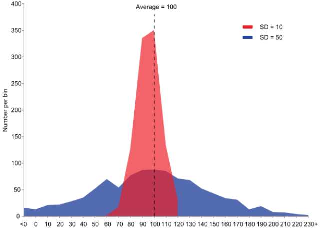


https://en.wikipedia.org/wiki/Variance

1. Line up the axes (dimensions) in the new vectors with the greatest variance in the word frequencies.
2. Rotate the vectors so that the new dimensions (basis vectors) align with the maximum variance word frequencies
3. Eliminate the dimensions that contribute the least to the variance in the vectors

- Each dimension (axis) becomes a **combination of word frequencies** rather than a single word frequency.

Intuition (3)



https://en.wikipedia.org/wiki/Variance

1. Line up the axes (dimensions) in the new vectors with the greatest variance in the word frequencies.
2. Rotate the vectors so that the new dimensions (basis vectors) align with the maximum variance word frequencies
3. Eliminate the dimensions that contribute the least to the variance in the vectors

- Each dimension (axis) becomes a **combination of word frequencies** rather than a single word frequency.
- They are weighted combinations of words that make up various "topics" in the corpus

Singular Value Decomposition

Singular Value Decomposition

- SVD finds co-occurring words by calculating the correlation between the terms of the term-document matrix
- SVD simultaneously finds the correlation of term use between documents and the correlation of documents with each other

Singular Value Decomposition

- SVD finds co-occurring words by calculating the correlation between the terms of the term-document matrix
- SVD simultaneously finds the correlation of term use between documents and the correlation of documents with each other

- With these two pieces of information SVD computes the linear combinations of terms that have the greatest variation across the corpus

Singular Value Decomposition

- SVD finds co-occurring words by calculating the correlation between the terms of the term-document matrix
- SVD simultaneously finds the correlation of term use between documents and the correlation of documents with each other

- With these two pieces of information SVD computes the linear combinations of terms that have the greatest variation across the corpus

These linear combinations of term frequencies will become topics

Considerations

- The machine does not *understand* what the combinations of words mean —just that they go together

Considerations

- The machine does not *understand* what the combinations of words mean —just that they go together
- `dog`, `cat`, and `love` appear together a lot $\rightarrow$ same topic

Considerations

- The machine does not *understand* what the combinations of words mean —just that they go together
- `dog`, `cat`, and `love` appear together a lot → same topic
- No idea this is (might be!) topic `pets`

Considerations

- The machine does not *understand* what the combinations of words mean —just that they go together
- `dog`, `cat`, and `love` appear together a lot $\rightarrow$ same topic
- No idea this is (might be!) topic `pets`
- It can include (near-)synonyms, but also antonyms

Considerations

- The machine does not *understand* what the combinations of words mean —just that they go together
- `dog`, `cat`, and `love` appear together a lot $\rightarrow$ same topic
- No idea this is (might be!) topic `pets`
- It can include (near-)synonyms, but also antonyms
- A human has to look at the words with a high weight to name the topic

Considerations

- The machine does not *understand* what the combinations of words mean —just that they go together
- `dog`, `cat`, and `love` appear together a lot $\rightarrow$ same topic
- No idea this is (might be!) topic `pets`
- It can include (near-)synonyms, but also antonyms
- A human has to look at the words with a high weight to name the topic
- **They can be used, even without a name**

Considerations

- The machine does not *understand* what the combinations of words mean —just that they go together
- `dog`, `cat`, and `love` appear together a lot $\rightarrow$ same topic
- No idea this is (might be!) topic `pets`
- It can include (near-)synonyms, but also antonyms
- A human has to look at the words with a high weight to name the topic
- **They can be used, even without a name**
- We can be add, subtract, compute similarities. . .

Behind SVD for NLP

Mathematical Formulation

$$W_{m\times n} \Rightarrow U_{m\times p} S_{p\times p} V^T_{p\times n}$$

where

- m is the size of the vocabulary

Behind SVD for NLP

Mathematical Formulation

$$W_{m \times n} \Rightarrow U_{m \times p} S_{p \times p} V^T_{p \times n}$$

where

- m is the size of the vocabulary
- n is the size of the corpus

Behind SVD for NLP

Mathematical Formulation

$$W_{m \times n} \Rightarrow U_{m \times p} S_{p \times p} V^{T}_{p \times n}$$

where

- m is the size of the vocabulary
- n is the size of the corpus
- p is the number of topics in the corpus (at time 0, $p = m$)

Behind SVD for NLP

Mathematical Formulation

$$W_{m\times n} \Rightarrow U_{m\times p} S_{p\times p} V^T_{p\times n}$$

where

- m is the size of the vocabulary
- n is the size of the corpus
- p is the number of topics in the corpus (at time 0, $p = m$)

We know what is W: BoW or TF-IDF matrix

Behind SVD for NLP

U–left singular vectors

$$W_{m \times n} \Rightarrow U_{m \times p} S_{p \times p} V^T_{p \times n}$$

- The **term-topic matrix**: "the company a word keeps"
- The cross-correlation between words and topics based on word co-occurrence in the same document.
- It is a square matrix

Behind SVD for NLP

S—singular vectors

$$W_{m \times n} \Rightarrow U_{m \times p} S_{p \times p} V^T_{p \times n}$$

- The **Sigma matrix**: the topic "singular values"

Behind SVD for NLP

S—singular vectors

$$W_{m \times n} \Rightarrow U_{m \times p} S_{p \times p} V^{T}_{p \times n}$$

- The **Sigma matrix**: the topic "singular values"
- A square diagonal matrix

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Behind SVD for NLP

S—singular vectors

$$W_{m\times n} \Rightarrow U_{m\times p} S_{p\times p} V^T_{p\times n}$$

- The **Sigma matrix**: the topic "singular values"
- A square diagonal matrix

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \rightarrow \begin{bmatrix} 0.6 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.05 \end{bmatrix}$$

- It tells you how much information is captured by each dimension in the new topic vector space.

Behind SVD for NLP

S—singular vectors

$$W_{m \times n} \Rightarrow U_{m \times p} S_{p \times p} V^T_{p \times n}$$

- The **Sigma matrix**: the topic “singular values”
- A square diagonal matrix

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \rightarrow \begin{bmatrix} 0.6 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.05 \end{bmatrix}$$

- It tells you how much information is captured by each dimension in the new topic vector space.
- The first dimension contains the most information (“explained variance”)

Behind SVD for NLP

V^T—right singular vectors

$$W_{m\times n} \Rightarrow U_{m\times p} S_{p\times p} V^T_{p\times n}$$

- The **document-document matrix**: the shared meaning between documents
- It measures how often documents use the same topics in the new model

Behind SVD for NLP

V^T—right singular vectors

$$W_{m\times n} \Rightarrow U_{m\times p} S_{p\times p} V^T_{p\times n}$$

- The **document-document matrix**: the shared meaning between documents
- It measures how often documents use the same topics in the new model
- A square matrix

SVD in Action

Let us see

Some Extra Pointers

Gensim Topic Modelling for Humans[1]

[1] https://radimrehurek.com/gensim/

Some Extra Pointers

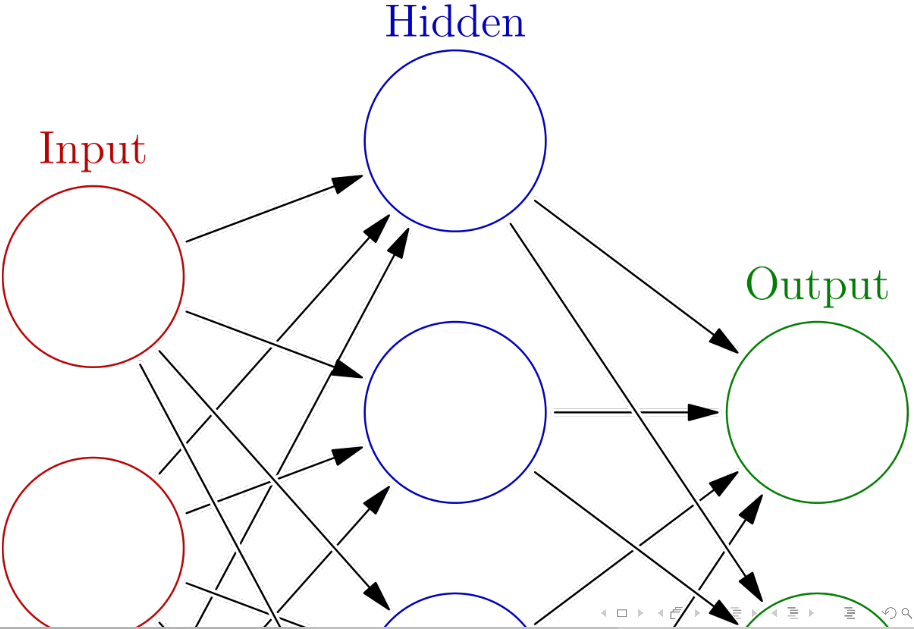
Gensim Topic Modelling for Humans[1]

(Literally) some random papers:

- Godin, et al. (2013). **Using Topic Models for Twitter Hashtag Recommendation**. WWW 2013 Companion.
- Rodriguez and Storer (2019). **A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data** JTHS.
- Seroussi, et al. (2014). **Authorship Attribution with Topic Models**. COLI

[1] `https://radimrehurek.com/gensim/`

Coming soon



References

Lane, H., C. Howard, and H. Hapkem
2019. *Natural Language Processing in Action*. Shelter Island, NY: Manning Publication Co.