Distributional word representations: dimensionality reduction

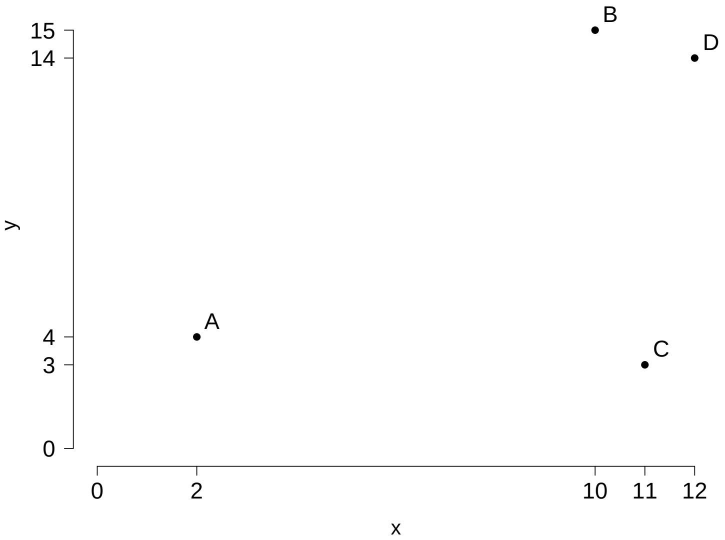
Chris Potts
Stanford Linguistics

CS 244U: Natural language understanding

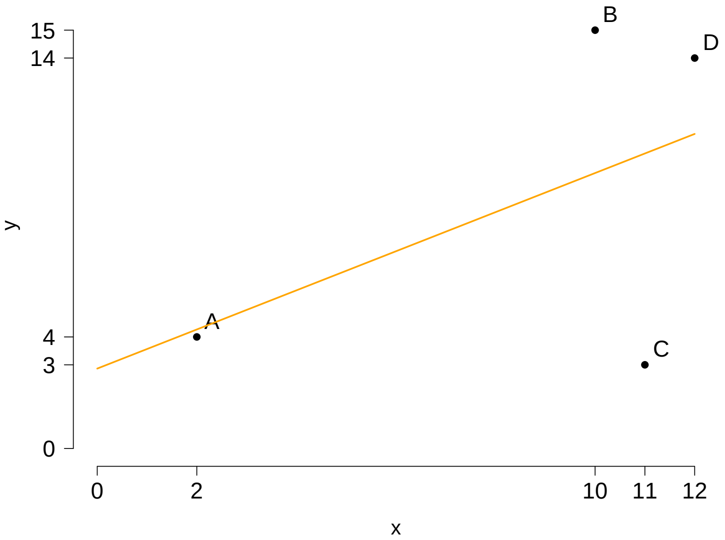
Goals

- Eliminate correlations
- Improve similarity measures

Guiding intuitions



Guiding intuitions



Latent Semantics Analayis (LSA)

Singular value decomposition

For any matrix of real numbers A of dimension $(m \times n)$ there exists a factorization into matrices T, S, D such that

$$A_{m \times n} = T_{m \times m} S_{m \times m} D^T_{n \times m}$$

Latent Semantics Analayis (LSA)

Singular value decomposition

For any matrix of real numbers A of dimension $(m \times n)$ there exists a factorization into matrices T, S, D such that

$$A_{m\times n} = T_{m\times m} S_{m\times m} D^T_{n\times m}$$

$$\underbrace{\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}}_{A_{3\times 4}} \underbrace{=}_{=} \underbrace{\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}}_{T_{3\times 3}} \underbrace{\begin{pmatrix} \cdot & & \\ & \cdot & \\ & & \cdot \end{pmatrix}}_{S_{3\times 3}} \underbrace{\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}^T}_{D^T_{4\times 3}}$$

Example

	d1	d2	d3	d4	d5	d6
gnarly	1	0	1	0	0	0
wicked	0	1	0	1	0	0
awesome	1	1	1	1	0	0
lame	0	0	0	0	1	1
terrible	0	0	0	0	0	1

$\Downarrow\Uparrow$

Example

	d1	d2	d3	d4	d5	d6
gnarly	1	0	1	0	0	0
wicked	0	1	0	1	0	0
awesome	1	1	1	1	0	0
lame	0	0	0	0	1	1
terrible	0	0	0	0	0	1

$\Downarrow\Uparrow$

Distance from *gnarly*
1. gnarly
2. awesome
3. terrible
4. wicked
5. lame

Example

	d1	d2	d3	d4	d5	d6
gnarly	1	0	1	0	0	0
wicked	0	1	0	1	0	0
awesome	1	1	1	1	0	0
lame	0	0	0	0	1	1
terrible	0	0	0	0	0	1

Distance from *gnarly*
1. gnarly
2. awesome
3. terrible
4. wicked
5. lame

$\Downarrow\Uparrow$

T(erm)

gnarly	0.41	0.00	0.71	0.00	-0.58
wicked	0.41	0.00	-0.71	0.00	-0.58
awesome	0.82	-0.00	-0.00	-0.00	0.58
lame	0.00	0.85	0.00	-0.53	0.00
terrible	0.00	0.53	0.00	0.85	0.00

$\times$

S(ingular values)

1	2.45	0.00	0.00	0.00	0.00
2	0.00	1.62	0.00	0.00	0.00
3	0.00	0.00	1.41	0.00	0.00
4	0.00	0.00	0.00	0.62	0.00
5	0.00	0.00	0.00	0.00	-0.00

$\times$

D(ocument)^T

d1	0.50	-0.00	0.50	0.00	-0.71
d2	0.50	0.00	-0.50	0.00	0.00
d3	0.50	-0.00	0.50	0.00	0.71
d4	0.50	-0.00	-0.50	-0.00	0.00
d5	-0.00	0.53	0.00	-0.85	0.00
d6	0.00	0.85	0.00	0.53	0.00

Example

	d1	d2	d3	d4	d5	d6
gnarly	1	0	1	0	0	0
wicked	0	1	0	1	0	0
awesome	1	1	1	1	0	0
lame	0	0	0	0	1	1
terrible	0	0	0	0	0	1

Distance from *gnarly*
1. gnarly
2. awesome
3. terrible
4. wicked
5. lame

$\Downarrow\Uparrow$

T(erm)					
gnarly	0.41	0.00	0.71	0.00	-0.58
wicked	0.41	0.00	-0.71	0.00	-0.58
awesome	0.82	-0.00	-0.00	-0.00	0.58
lame	0.00	0.85	0.00	-0.53	0.00
terrible	0.00	0.53	0.00	0.85	0.00

$\times$

S(ingular values)					
1	2.45	0.00	0.00	0.00	0.00
2	0.00	1.62	0.00	0.00	0.00
3	0.00	0.00	1.41	0.00	0.00
4	0.00	0.00	0.00	0.62	0.00
5	0.00	0.00	0.00	0.00	-0.00

$\times$

D(ocument)					
d1	0.50	-0.00	0.50	0.00	-0.71
d2	0.50	0.00	-0.50	0.00	0.00
d3	0.50	-0.00	0.50	0.00	0.71
d4	0.50	-0.00	-0.50	-0.00	0.00
d5	-0.00	0.53	0.00	-0.85	0.00
d6	0.00	0.85	0.00	0.53	0.00

T

gnarly	0.41	0.00
wicked	0.41	0.00
awesome	0.82	-0.00
lame	0.00	0.85
terrible	0.00	0.53

$\times$

2.45	0.00
0.00	1.62

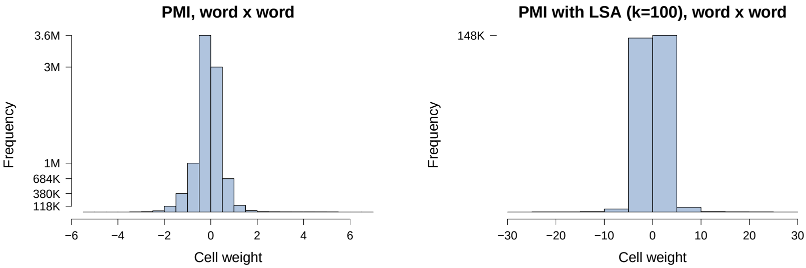
$=$

gnarly	1.00	0.00
wicked	1.00	0.00
awesome	2.00	0.00
lame	0.00	1.38
terrible	0.00	0.85

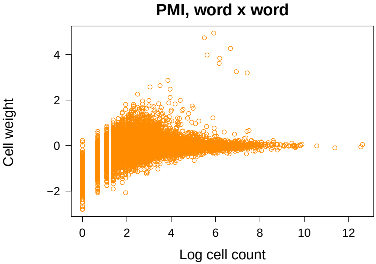
Distance from *gnarly*
1. gnarly
2. wicked
3. awesome
4. terrible
5. lame

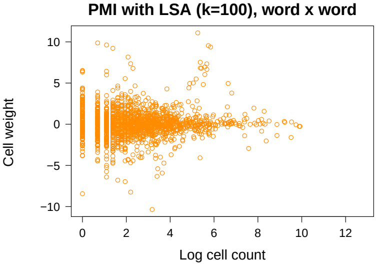
Comparisons before and after LSA with k=100

Comparisons before and after LSA with k=100

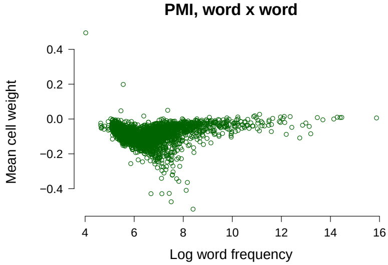


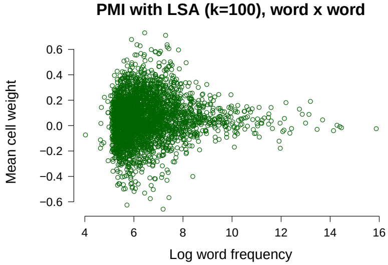
Comparisons before and after LSA with k=100





Comparisons before and after LSA with k=100





Other dimensionality reduction techniques

Other dimensionality reduction techniques

- Principal Components Analysis (PCA)

Other dimensionality reduction techniques

- Principal Components Analysis (PCA)
- Probabilistic LSA (PLSA)

Other dimensionality reduction techniques

- Principal Components Analysis (PCA)
- Probabilistic LSA (PLSA)
- Latent Dirichlet Allocation

Other dimensionality reduction techniques

- Principal Components Analysis (PCA)
- Probabilistic LSA (PLSA)
- Latent Dirichlet Allocation
- t-Distributed Stochastic Neighbor Embedding

Other dimensionality reduction techniques

- Principal Components Analysis (PCA)
- Probabilistic LSA (PLSA)
- Latent Dirichlet Allocation
- t-Distributed Stochastic Neighbor Embedding
- word2vec

Other dimensionality reduction techniques

- Principal Components Analysis (PCA)
- Probabilistic LSA (PLSA)
- Latent Dirichlet Allocation
- t-Distributed Stochastic Neighbor Embedding
- word2vec
- GloVe

Other dimensionality reduction techniques

- Principal Components Analysis (PCA)
- Probabilistic LSA (PLSA)
- Latent Dirichlet Allocation
- t-Distributed Stochastic Neighbor Embedding
- word2vec
- GloVe

For more: Turney and Pantel 2010, ‘From frequency to meaning’, p. 160.