Assignment 2: Solution

Denis Paperno
Loria
CNRS

TyLex, Voronovo, 07.09.2017

Task

	crowned	says	knighthood	delivery	child
woman	10	82	0	4	275
queen	85	5	4	0	8
king	237	20	4	1	9
man	11	181	1	34	138

				crowned	says	knighthood	delivery	child
woman	49	-6	*	1,1	1,9	0,025	0,3	2,5
queen	8,1	-0,28		11	0,17	0,24	1,3	-26
king	35	2,7						
man	81	2,5						

Question 1

- 1. What kind of information do the 2-dimensional word vectors encode ? Can you give interpretation to individual dimensions ?

woman	49	-6
queen	8,1	-0,28
king	35	2,7
man	81	2,5

*

crowned	says	knighthood	delivery	child
1,1	1,9	0,025	0,3	2,5
11	0,17	0,24	1,3	-26

Question 1

- Dimension 2 is associated with gender
  - Positive values for man and king
  - Negative values for woman and queen
- Dimension 1 correlates with word frequency
  - also with the category of 'royalty' but we have too few examples to discriminate it from frequency

			crowned	says	knighthood	delivery	child
woman	49	-6 *	1,1	1,9	0,025	0,3	2,5
queen	8,1	-0,28	11	0,17	0,24	1,3	-26
king	35	2,7					
man	81	2,5					

Question 2

- Reconstruct the 5-dimensional vector of *woman* from the decomposition. How accurate is the reconstruction ? Do you find the reconstruction

			crowned	says	knighthood	delivery	child
woman	49	-6 *	1,1	1,9	0,025	0,3	2,5
queen	8,1	-0,28	11	0,17	0,24	1,3	-26
king	35	2,7					
man	81	2,5					

- woman=

			crowned	says	knighthood	delivery	child
woman	49	-6 *	1,1	1,9	0,025	0,3	2,5
queen	8,1	-0,28	11	0,17	0,24	1,3	-26
king	35	2,7					
man	81	2,5					

Reconstructed vector for *woman*

- Multiply first line by the latent factor-context table:

woman	49	-6
queen	8,1	-0,28
king	35	2,7
man	81	2,5

*

crowned	says	knighthood	delivery	child
1,1	1,9	0,025	0,3	2,5
11	0,17	0,24	1,3	-26

49*1.1+-6*11	49*1.9+-6*0.17	49*0.025+-6*0.24	49*0.3+-6*1.3	49*2.5+-6*-26

=

-12.1	92.08	-0.215	6.9	278.5

10	82	0	4	285

How big is the error?

49*1.1+-6*11	49*1.9+-6*0.17	49*0.025+-6*0.24	49*0.3+-6*1.3	49*2.5+-6*-26

=

-12.1	92.08	-0.215	6.9	278.5

10	82	0	4	285

- Error is noticeable
- Mean square error: 11.32
- Cosine with the original woman vector: 0.91
- Two dimensions are not enough

Question 3

Assume a linear function that maps a vector of a semantically masculine noun to the corresponding feminine noun, e.g. *king* to *queen*. Based on the 2-dimensional vectors for *king*, *queen*, *man*, and *woman*, estimate the matrix that encodes that function.

Question 3

- *king* * M=*queen*
- *man* * M=*woman*

  $$M= \begin{matrix} x & x' \\ y & y' \end{matrix}$$

- *man>woman:* 81 x + 2.5 y=49 ; 81 x' + 2.5 y'=-6
- *king>queen:* 35x + 2.7y=8.1; 35 x' + 2.7y'=-0.28

Gender mapping matrix

- *man>woman:* 81 *x* + 2.5 *y*=49 ; 81 *x'* + 2.5 *y'*=-6

- *king>queen:* 35*x* + 2.7*y*=8.1; 35 *x'* + 2.7*y'*=-0.28

- Four linear equations with 4 variables, exact solution (values rounded):

      0.85 -0.12

- M = -8.1   1.4

  No exact solution guaranteed with more example pairs