NLP and Word Embeddings

deeplearning.ai

Properties of word embeddings

Analogies

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.00
Age	0.03	0.02	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97

e_{5391}
e_{man}

e_{woman}

$e_{man} - e_{woman} \approx \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

Man $\rightarrow$ Woman as King $\rightarrow$? Queen

$e_{king} - e_{queen} \approx \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$e_{man} - e_{woman} \approx e_{king} - e_{?}$
$\uparrow$ Queen

[Mikolov et. al., 2013, Linguistic regularities in continuous space word representations] Andrew Ng

Analogies using word vectors
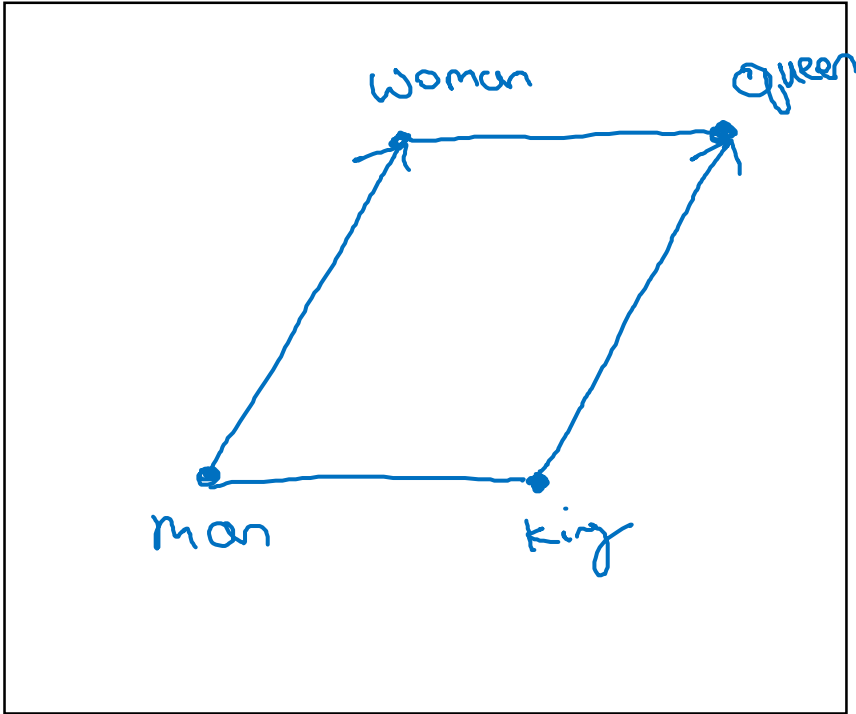
$$e_{man} - e_{woman} \approx e_{king} - e_?$$

300 D

Find word w: arg max $Sim(e_w, e_{king} - e_{man} + e_{woman})$
            w

30 - 75%

3000 → 2D

t-SNE

Andrew Ng
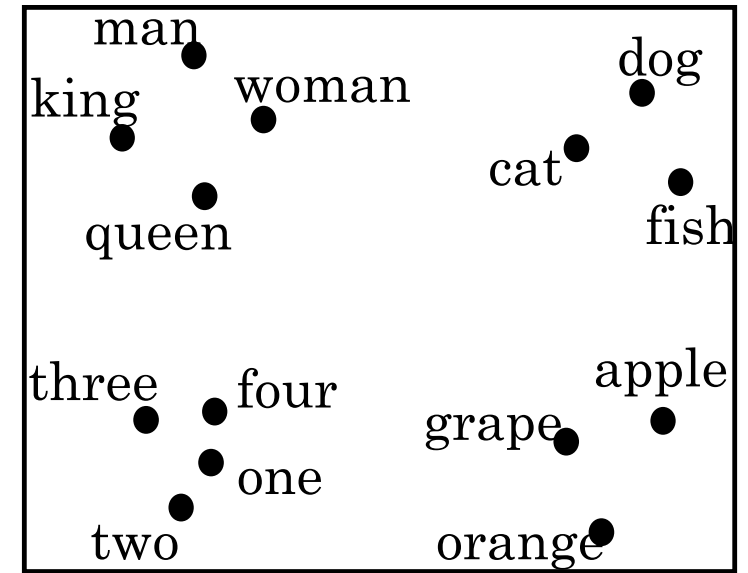
Cosine similarity

$$sim(e_w, e_{king} - e_{man} + e_{woman})$$

$$Sim(u,v) = \frac{u^T v}{||u||_2 ||v||_2}$$
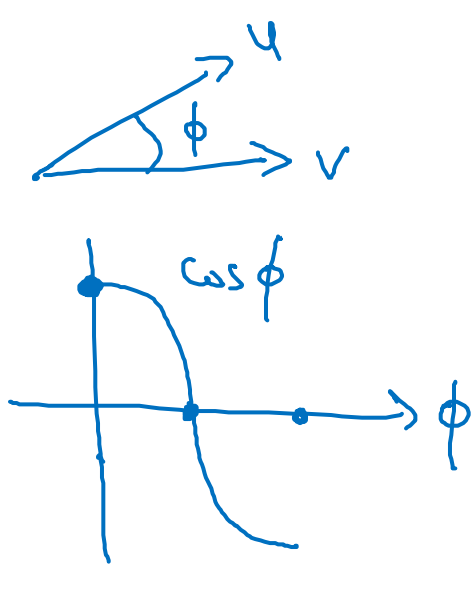
$$||u - v||^2$$

Man:Woman as Boy:Girl

Ottawa:Canada as Nairobi:Kenya

Big:Bigger as Tall:Taller

Yen:Japan as Ruble:Russia

Andrew Ng