NLP and Word Embeddings

Properties of word embeddings

deeplearning.ai

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}

Man → Woman   as   King → ? Queen

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

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

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

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

Analogies using word vectors
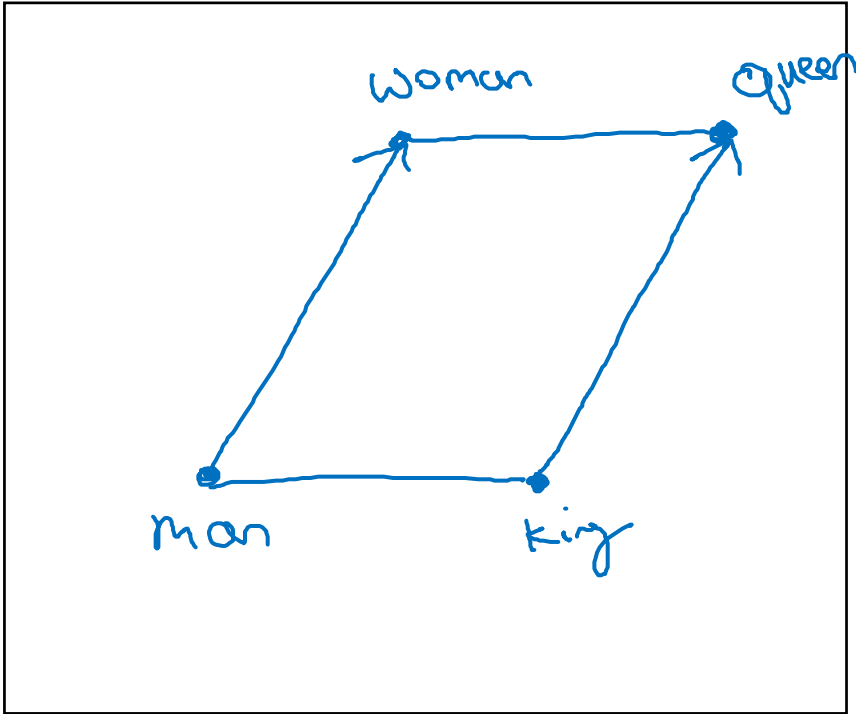
$3000 \rightarrow 2D$
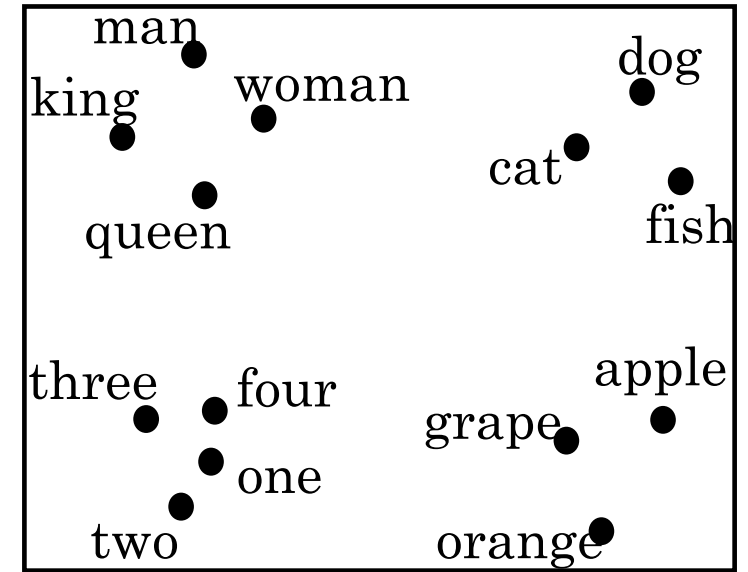
t-SNE

$$e_{man} - e_{woman} \approx e_{king} - e_?, \quad e_w$$

$300 D$

Find word w: $\arg\max\limits_{w} \text{sim}(e_w, e_{king} - e_{man} + e_{woman})$

$30 - 75\%$

Andrew Ng

Cosine similarity

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

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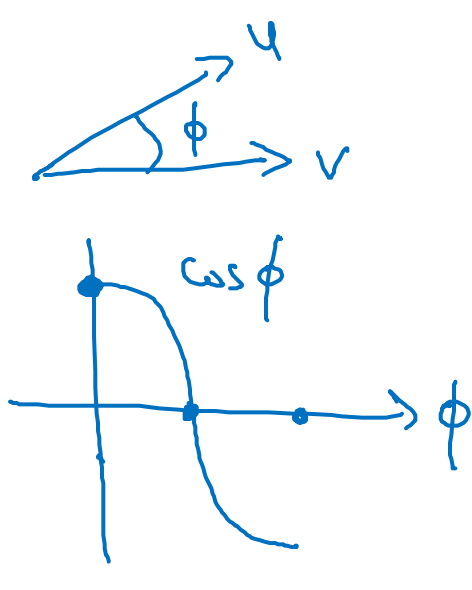
$\|u - v\|^2$

Man:Woman as Boy:Girl

Ottawa:Canada as Nairobi:Kenya

Big:Bigger as Tall:Taller

Yen:Japan as Ruble:Russia

Andrew Ng