



deeplearning.ai

NLP and Word Embeddings

Properties of word embeddings

Analogy

	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

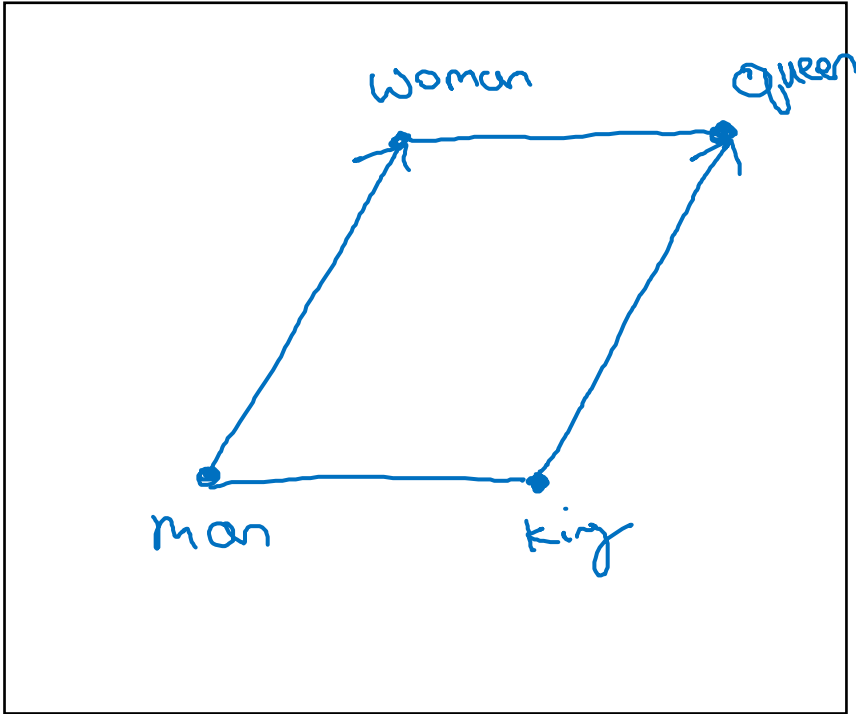
$\underbrace{e_{5391}}_{e_{\text{man}}} \rightarrow \underbrace{e_{9853}}_{e_{\text{woman}}} \quad \Leftrightarrow \quad \underbrace{e_{4914}}_{e_{\text{king}}} \rightarrow ? \quad \underbrace{e_{7157}}_{e_{\text{queen}}}$

$e_{\text{man}} - e_{\text{woman}} \approx e_{\text{king}} - e_{\text{?}}$

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

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

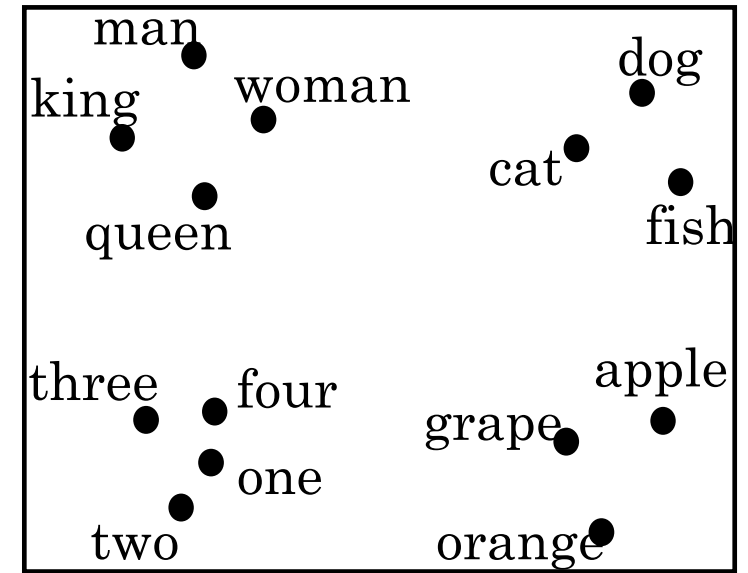
Analogies using word vectors



300 D

Find word w : $\arg \max_w$

3000 \rightarrow 20
↑



t-SNE

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

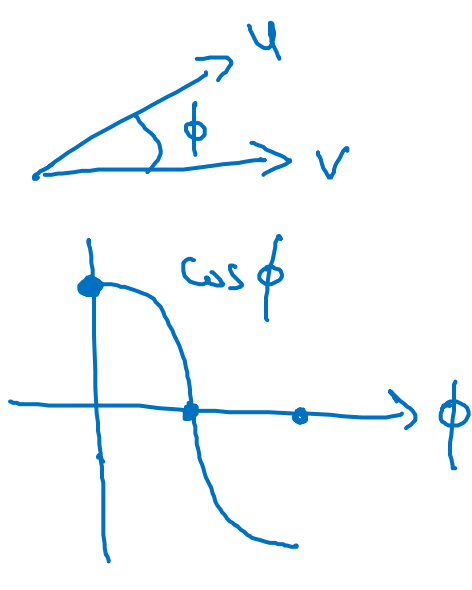
$$\text{Sim}(\underbrace{e_w}_{\uparrow}, \underbrace{e_{king} - e_{man} + e_{woman}}_{\text{30-75\%}})$$

30-75%

Cosine similarity

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

$$\text{sim}(u, v) = \frac{u^T v}{\|u\|_2 \|v\|_2}$$



$$\|u - v\|^2$$

Man:Woman as Boy:Girl

Ottawa:Canada as Nairobi:Kenya

Big:Bigger as Tall:Taller

Yen:Japan as Ruble:Russia