



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

$$\underbrace{e_{\text{man}}}_{5391} - \underbrace{e_{\text{woman}}}_{9853} \approx \underbrace{e_{\text{king}}}_{4914} - \underbrace{e_{\text{queen}}}_{7157}$$

$$\underbrace{e_{\text{man}}}_{5391} - \underbrace{e_{\text{woman}}}_{9853} \approx \underbrace{e_{\text{king}}}_{4914} - \underbrace{e_{\text{queen}}}_{7157}$$

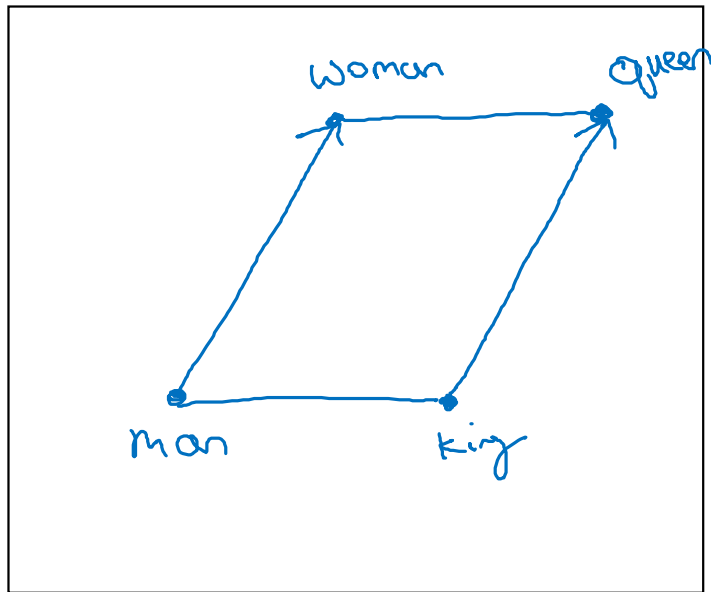
$$\underbrace{e_{\text{man}}}_{5391} - \underbrace{e_{\text{woman}}}_{9853} \approx \underbrace{e_{\text{king}}}_{4914} - \underbrace{e_{\text{queen}}}_{7157}$$

$$\underbrace{e_{\text{man}}}_{5391} - \underbrace{e_{\text{woman}}}_{9853} \approx \underbrace{e_{\text{king}}}_{4914} - \underbrace{e_{\text{queen}}}_{7157}$$

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

Andrew Ng

# Analogies using word vectors



300 D

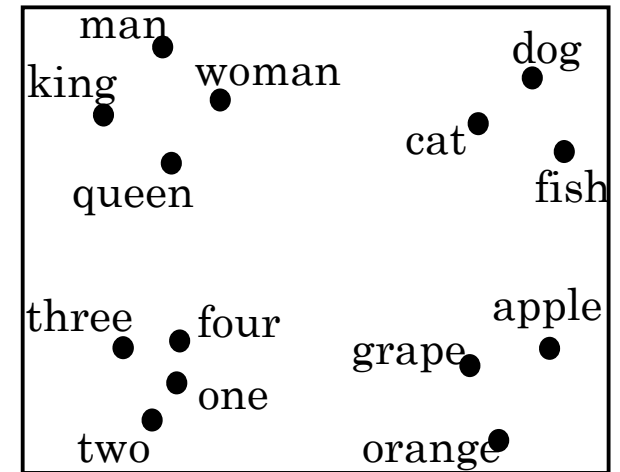
Find word  $w$ :  $\arg \max_w$

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

$$\text{Sim}(\underline{e_w}, \underline{e_{\text{king}} - e_{\text{man}} + e_{\text{woman}}})$$

30 - 75%

300D → 2D  
↑

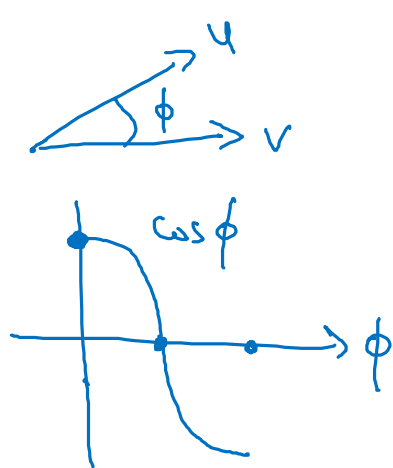


t-SNE

# Cosine similarity

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

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



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

Man:Woman as Boy:Girl

Ottawa:Canada as Nairobi:Kenya

Big:Bigger as Tall:Taller

Yen:Japan as Ruble:Russia