



deeplearning.ai

NLP and Word Embeddings

GloVe word vectors

GloVe (global vectors for word representation)

I want a glass of orange juice to go along with my cereal.

c, t

X_{ij} = # times i appears in context of j .

$\begin{matrix} \uparrow & \uparrow \\ c & t \end{matrix}$ $\begin{matrix} \uparrow \\ t \end{matrix}$ $\begin{matrix} \uparrow \\ c \end{matrix}$

$X_{ij} = X_{ji} \leftarrow$

Model

minimize

$$\sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(x_{ij}) \left(\underbrace{\Theta_i^T e_j}_{\substack{t \quad c \\ \text{"}\Theta_t^T e_c\text{"}}} + b_i + b_j' - \log x_{ij} \right)^2 \quad \leftarrow$$

\downarrow \downarrow \downarrow \downarrow
 t c t c

weighting
term

$$f(x_{ij}) = 0 \text{ at } x_{ij} = 0$$

$$"0 \log 0" = 0$$

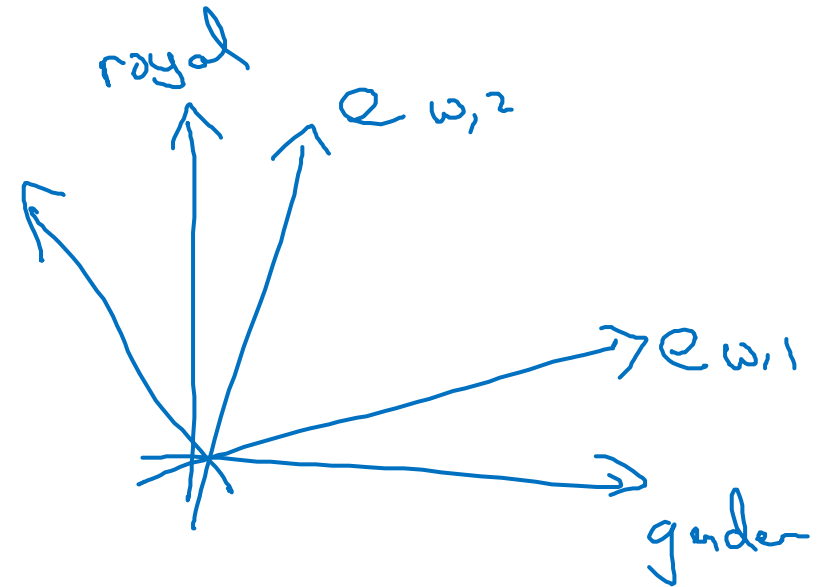
→ this, is, at, a, ...
 → derivation

Θ_i, e_j are symmetric

$$e_w^{(final)} = \frac{e_w + \Theta_w}{2}$$

A note on the featurization view of word embeddings

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



$$\text{minimize } \sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(X_{ij}) (\underbrace{\theta_i^T e_j}_{(A\theta_i)^T (A^T e_j)} + b_i - b'_j - \log X_{ij})^2$$

$$\underbrace{(A\theta_i)^T (A^T e_j)}_{\theta_i^T A^T A e_j}$$