



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

$X_{ij} = \# \text{ times } i \text{ appears in context of } j$

Diagram illustrating the GloVe matrix construction: The sentence "I want a glass of orange juice to go along with my cereal." is shown. The word "orange" is highlighted in red and labeled with i . The word "juice" is highlighted in red and labeled with j . The context words "c" (cereal) and "t" (to) are shown below "orange" and "juice" respectively, with arrows pointing to them. The matrix element X_{ij} is defined as the number of times word i appears in the context of word j .

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



[Pennington et. al., 2014. GloVe: Global vectors for word representation]

Andrew Ng

Model

网易云课堂

? minimize

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

weighting term

$f(x_{ij}) = 0$ at $x_{ij} = 0$. "0 log 0" = 0

this, is, at, a, ...
duration

θ_i, e_j are symmetric

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

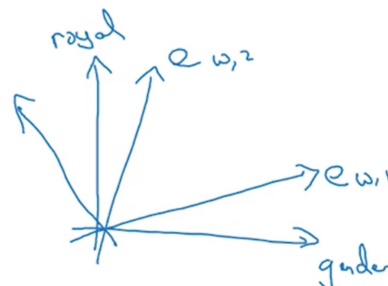


Andrew Ng

A note on the featurization view of word embeddings

网易云课堂

	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



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

$(A \theta_i)^T (A^T e_j) = \theta_i^T A^T A e_j$



Andrew Ng

