deeplearning.ai

NLP and Word Embeddings

Word representation

Word representation

V = [a, aaron, …, zulu, <UNK>]

|V| = 10,000

1-hot representation

Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
O_{5391}	O_{9853}				

I want a glass of orange _____.

I want a glass of apple _____.

Disadvantage for 1-hot representation is that: it doesn't reflect the correlation (similarity) between words.

Andrew Ng

Featurized representation: word embedding

	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	0.7	0.69	0.03	-0.02
Food	0.04	0.01	0.02	0.01	0.95	0.97
⋮	⋮	⋮				
Size						
Cost						
Alive						
Verb						

300

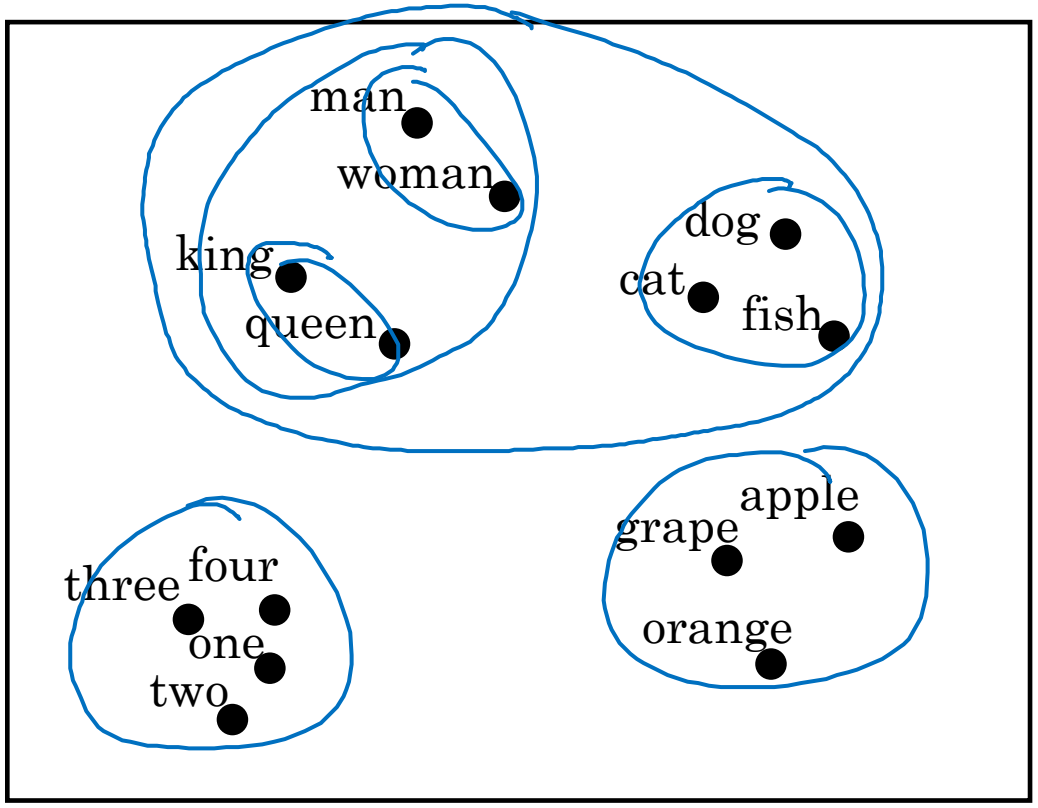
e_{5391} e_{9853}

I want a glass of orange juice.

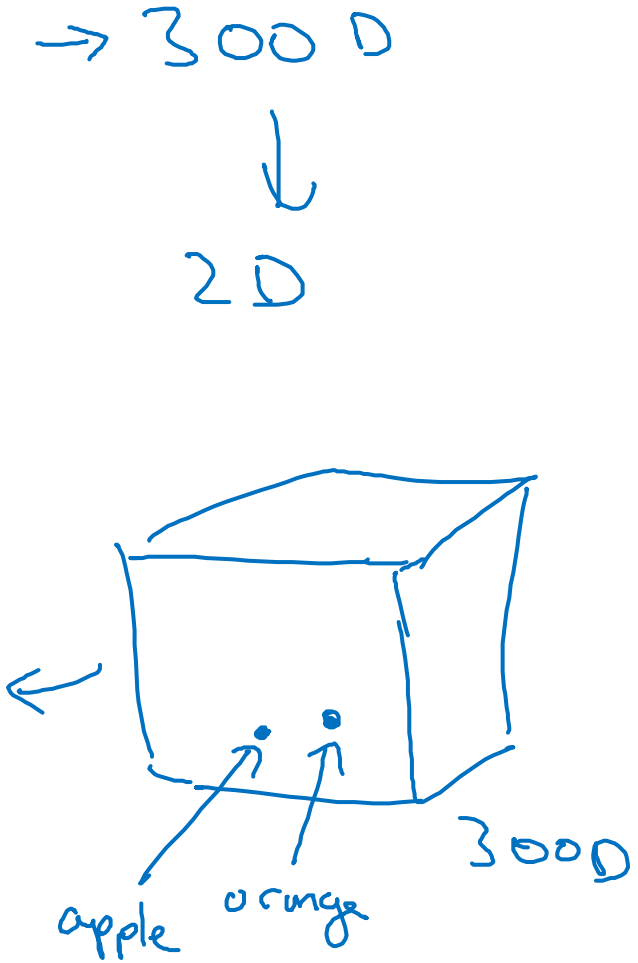
I want a glass of apple juice.

Andrew Ng

Visualizing word embeddings



t-SNE



[van der Maaten and Hinton., 2008. Visualizing data using t-SNE]