Word 2 Vec

King → Power ↑, Rich ↑, Male, Palace

Man → Power, Rich, Male, Home

Horse → Power ↓, Rich ↓, Male, Home ↓, tail ↑

	Features / Properties			
Index	Health	Power	Gender	Tail
King	1.98	1.98	−1	0.0002
Queen	1.98	1.98	+1	0.0002
Man	0.98	0.98	−1	0.0002
Woman	0.98	0.98	+1	0.0002

Count Vectorizer and TFIDF ⟹ Sparse matrix

Word2Vec → Dense vector

0 ⟶ 0.0002

↑ ⟶ 1.98

Male → −1

Female → +1

Word2Vec → Simple neural network

Single hidden layer

Assumption of word2Vec:

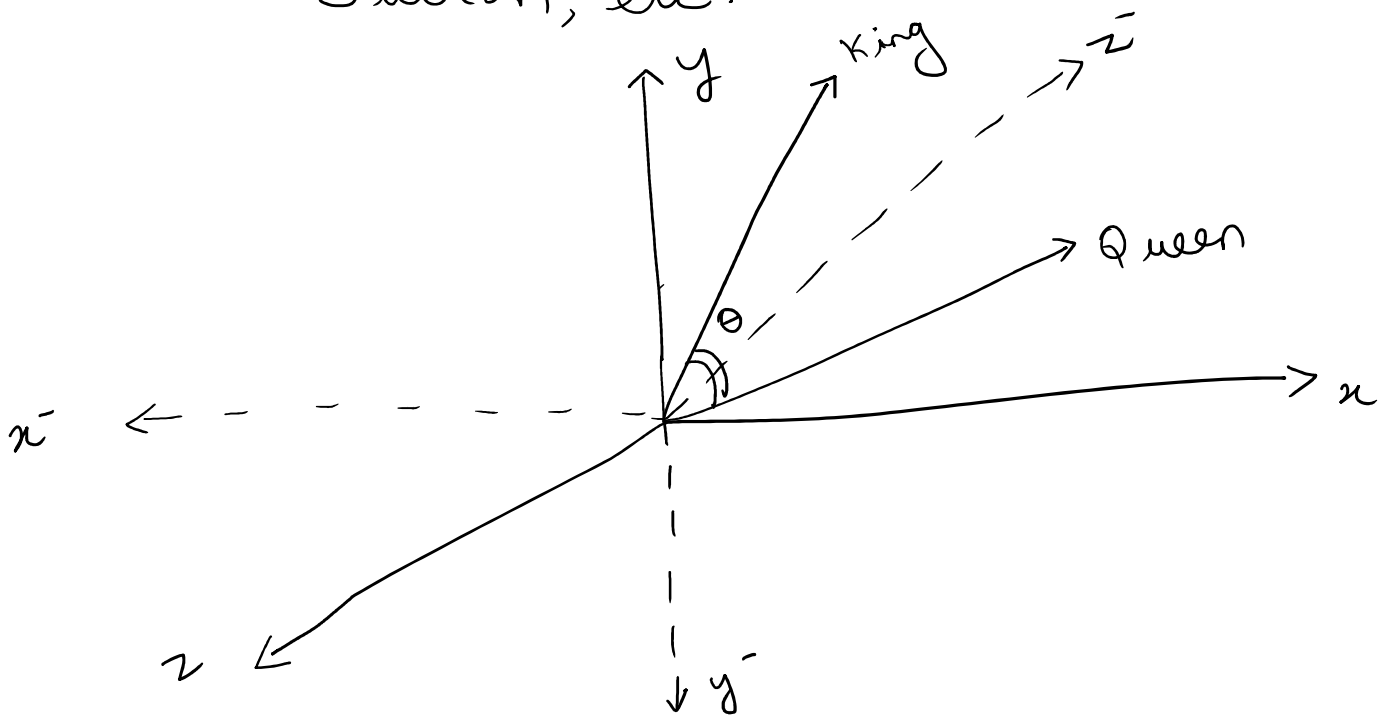
Words with similar context will have similar vector representation.

vector → Magnitude and direction

Similar context:

King → Prince, Princess, Queen, Ruler, Sultan, etc.

Sultan, etc.



$$\text{cosine similarity} = \frac{A \cdot B}{|A| \times |B|}$$

-1 to $+1$

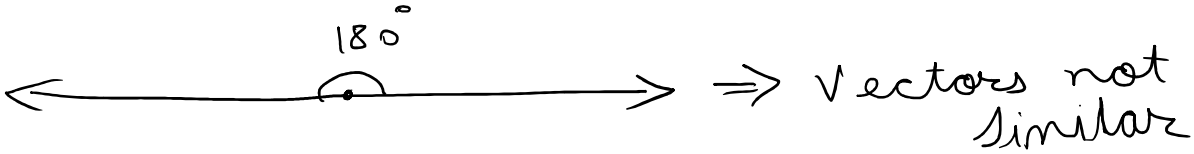
↑ Vectors are not similar at all

↑ Vectors are similar

$\cos 0 = 1$ $\quad \theta \uparrow \Rightarrow \cos\theta \downarrow$

$\cos 180 = -1$ $\quad \theta \downarrow \Rightarrow \cos\theta \uparrow$

0° ⟶ ⟶ $\Rightarrow$ Vectors similar


180°
⇒ Vectors not similar