

Word Embeddings  
Way of Representing words.  
Word representation:

$V = [a, aaron, \dots, zulu, \langle \text{UNK} \rangle]$

$|V| = 10,000$  words

1-hot representation

if MAN is 5894

$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$  — 5894

$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 0 \end{bmatrix}$  Queen  
05791

Same will be the case for other words such as WOMAN, QUEEN, APPLE

It treats each word independently and does not generalize the cross words

Example:

I want a glass of orange —

I want a glass of apple —.

juice is the likely word  
in both the sentences.

However, 1-hot representation  
does not show any representation  
between orange and Apple.

Any product between any two-different  
one-hot vector is zero.

Any product of two 1-hot vector  
is zero.

Euclidean distance between any  
pair of these vectors is also same.

so the algorithm does not know

the similarity between king and  
queen.

king and queen are more  
similar than king and orange.

## Featurized Representation: Word Embedding

	5175 Man	5283 Woman	King	Queen	Apple	Orange
Gender	-1	1	-0.95	0.97	0.0	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						
	$e_{5175}$	$e_{5283}$				
	300-dimension vector					

Learn featurized representation with each of these words.

Let's suppose there are 300-features

so we can use a 300 dimensional vector for each word.

- Low dimension
- Dense Matrix

Use Cosine Similarity b/w two vectors.



Analogy b/w Boy and Girl  
and King and Queen

Cosine Similarity can be used to  
compute distance b/w two Vectors

For two Vectors  $x(x_1, x_2, x_3, \dots, x_n)$   
and  $y(y_1, y_2, y_3, \dots, y_n)$

$$\text{Sim}(x, y) = \frac{x \cdot y}{|x| |y|}$$

$|x|$  is the Euclidean norm of Vector( $x$ ).

The closer the cosine value to 1,  
the smaller the angle and the  
greater the match.

Cosine value = 0; two vectors are at 90

Cosine value = 1; two vectors have 0 angle.

$$x = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0) \quad y = (3, 0, 2, 0, 1, 1, 0, 1, 0, 1)$$

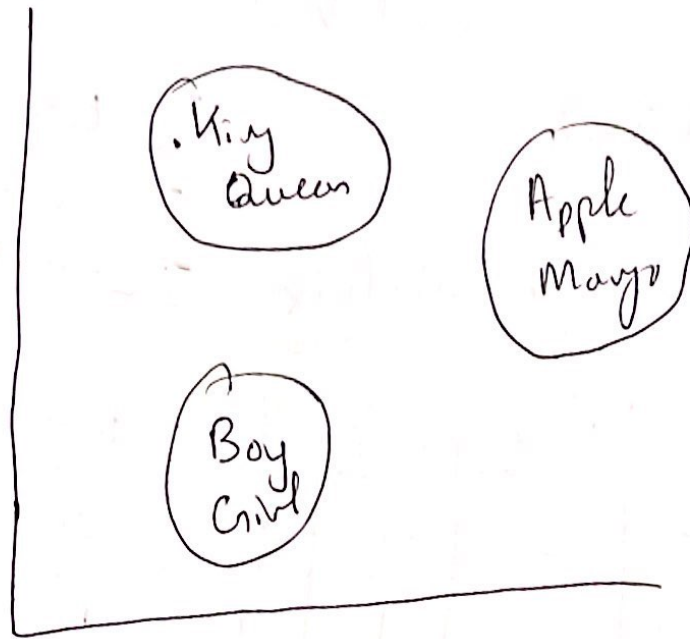
$$x \cdot y = 5 \times 3 + 0 \times 0 + 3 \times 2 + 0 \times 0 + 2 \times 1 + 0 \times 1 \\ + 0 \times 0 + 0 \times 1 = 25$$

$$|x| = \sqrt{5^2 + 0^2 + 3^2 + 0^2 + 2^2 + 0^2 + 0^2 + 2^2 + 0^2}$$

$$|y| = \sqrt{3^2 + 0^2 + 2^2 + 0^2 + 1^2 + 1^2 + 0^2 + 1^2 + 0^2}$$

$$\text{Sim}(x, y) = 0.94$$

Modify the 300-dimension into 2-dim.



Dimensionality reduction.