In this week we have learnt about word embedding, to translate the word English

Notebook:	Natural Language Processing		
Created:	02-08-2020 14:38	**Updated:**	02-08-2020 14:55
Author:	Zaid Hasib		
Tags:	Week3		
URL:	https://colab.research.google.com/drive/1jYZnUgRrxG9UYwmyTYhOsKccpf0BSXiR#scrol...		

In this week we have learnt about word embedding, to translate the word English to French.

We will try to implement basic idea, like if e have to find the capital of Russia, and we know the capital of USA, so we can find like USA-DC+RUSSIA.

Once we had embedded the word, then use KNN to find the nearest neighbor of the result the we have found.

2.2 Testing the translation

k-Nearest neighbors algorithm

k-Nearest neighbors algorithm

- k-NN is a method which takes a vector as input and finds the other vectors in the dataset that are closest to it.
- The 'k' is the number of "nearest neighbors" to find (e.g. k=2 finds the closest two neighbors).

Searching for the translation embedding

Since we're approximating the translation function from English to French embeddings by a linear transformation matrix $\mathbf{R}$, most of the time we won't get the exact embedding of a French word when we transform embedding $\mathbf{e}$ of some particular English word into the French embedding space.

- This is where k-NN becomes really useful! By using 1-NN with $\mathbf{eR}$ as input, we can search for an embedding $\mathbf{f}$ (as a row) in the matrix $\mathbf{Y}$ which is the closest to the transformed vector $\mathbf{eR}$

We have used LSH (Local Sensitivity Hashing) to store the value. We will store the value for the nearest neighbor in one hash key