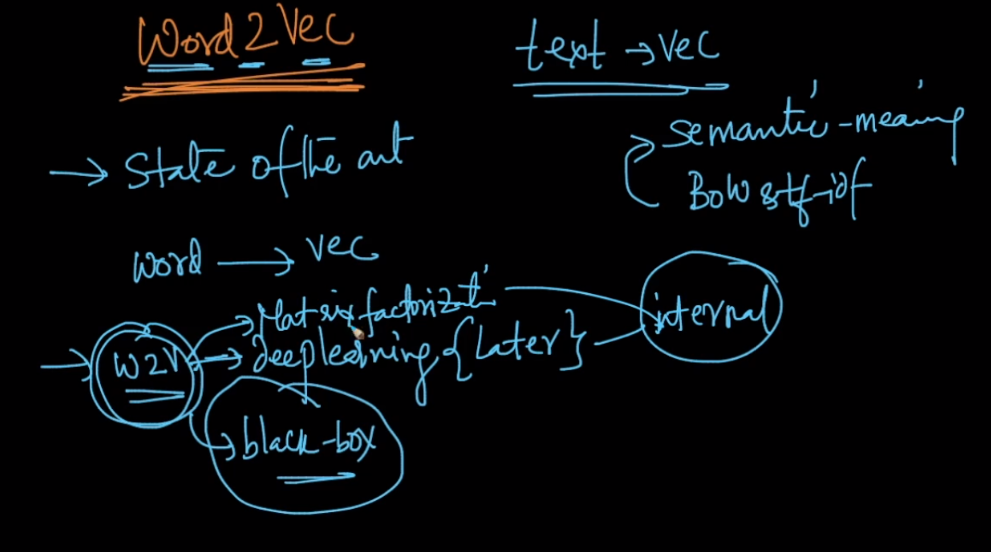
**Word2Vec.**

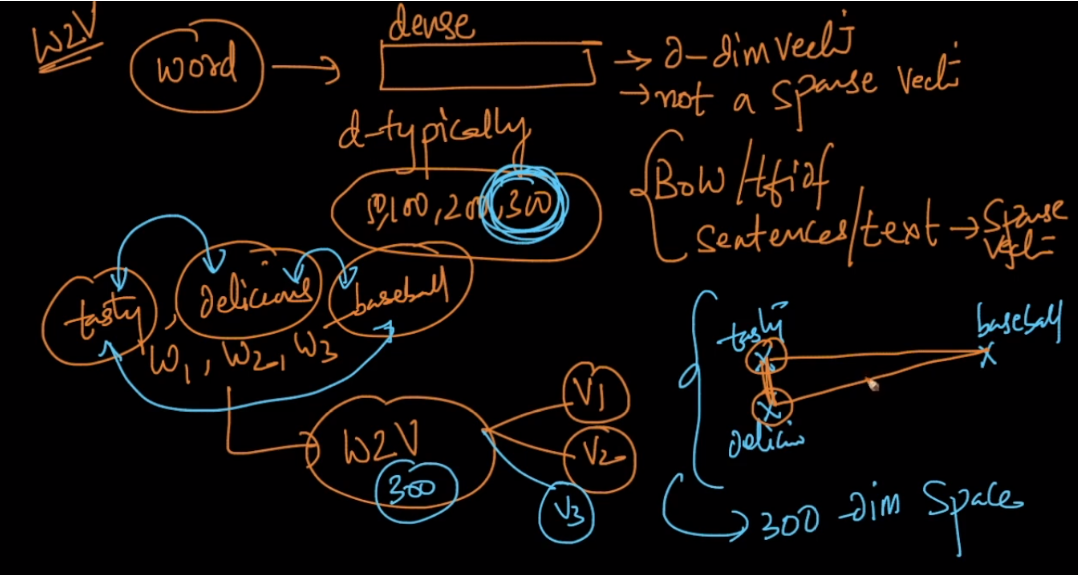
Word 2 vector is an state of the art to convert text to vector, it maintains the semantic meaning of text unlike BOW & TFIDF

We learn more about it in deep learning

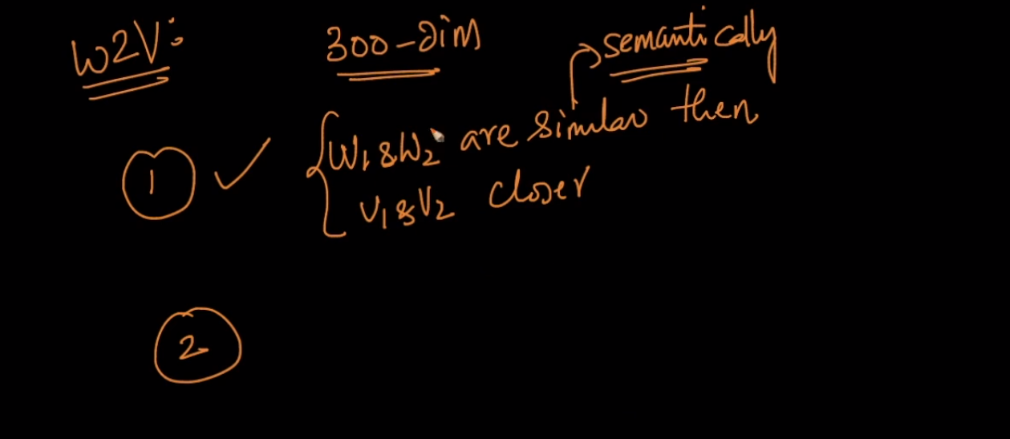


In this it converts word to vector which is d-dimension vector and it not a sparse vector i.e it is a dense vector.

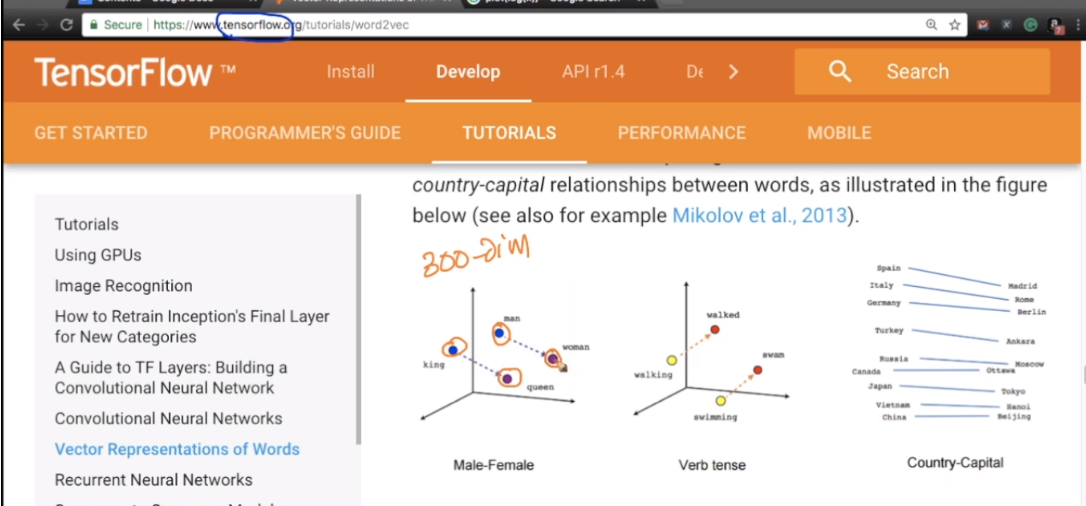
Suppose we give three words to W2V w1 : tasty, w2 : delicious, w3 : baseball and suppose W2v gives 300 dimension vector v1 for w1, v2 for w2, v3 for w3 and as w1 and w2 have same meaning therefore distance b/w vectors of w1 and w2 are smaller than w1 and w3 and w2 and w3.



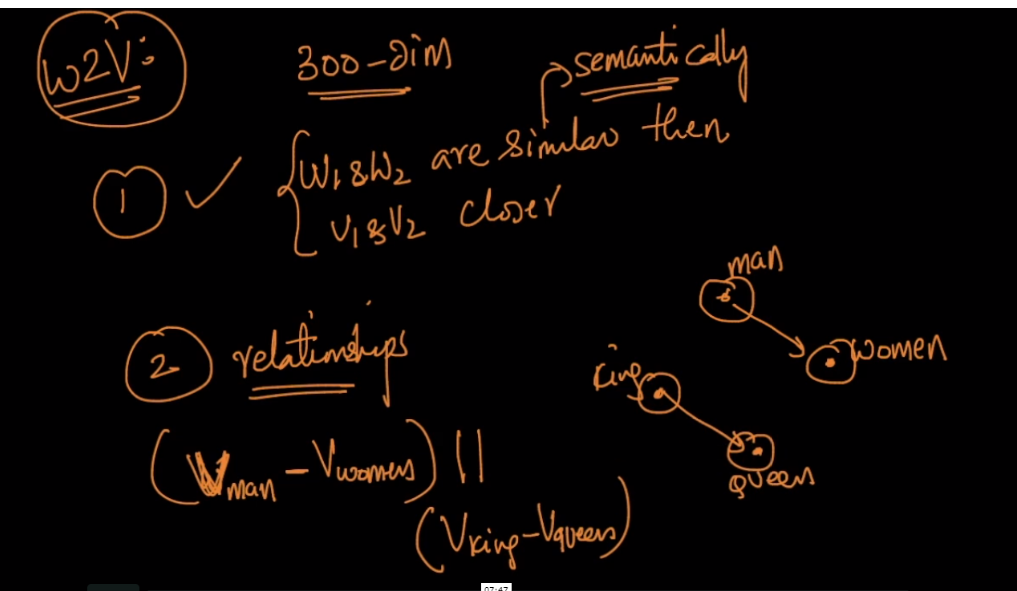
As shown below as w1 and w2 are similar then v1 and v2 closer.



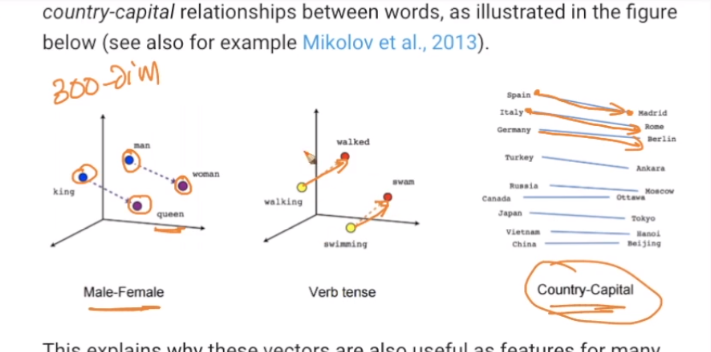
Below figure shows the relationship b/w man and woman , king and queen as this relationship is same b/w woman is a female gender of man and queen is a female gender of king therefore both vectors are parallel.



In 2nd point it shows relationship where Vman – Vwomen  gives direction of vector similar for king and queen.



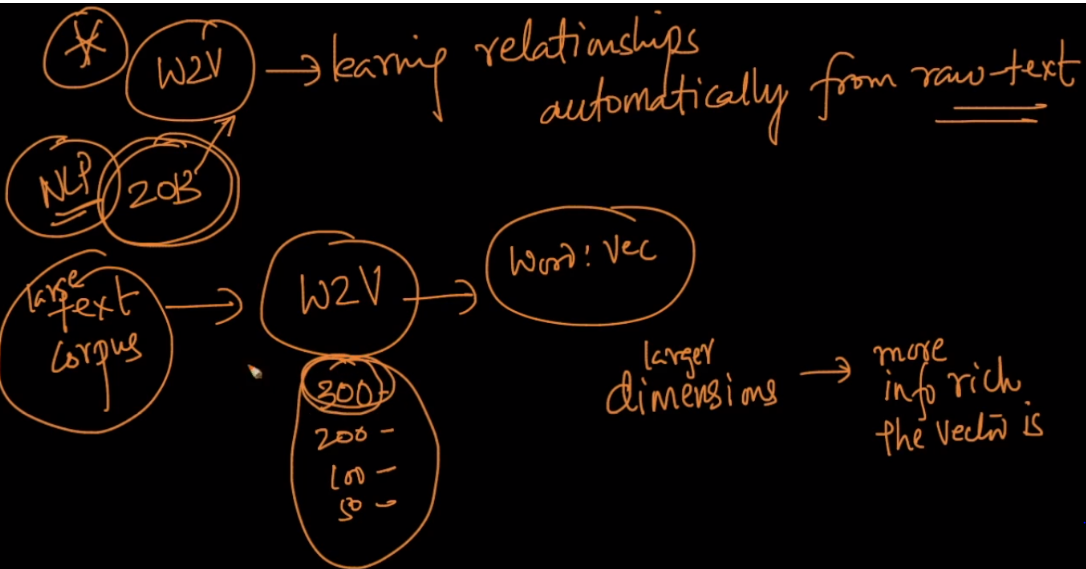
Similarly it shows relationship b/w country and capital , and verb and tense



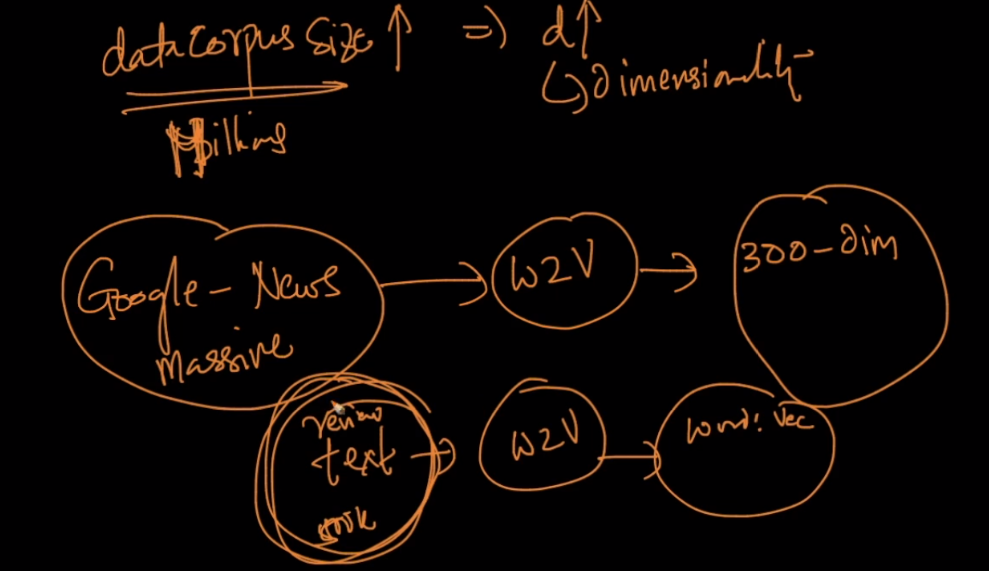
W2V learns this relationship automatically from raw-text.

And if there is larger text corpus given to W2V then it generates larger dimensions vector of each word.

And as larger dimensions of vector are there then more information rich the vector is.

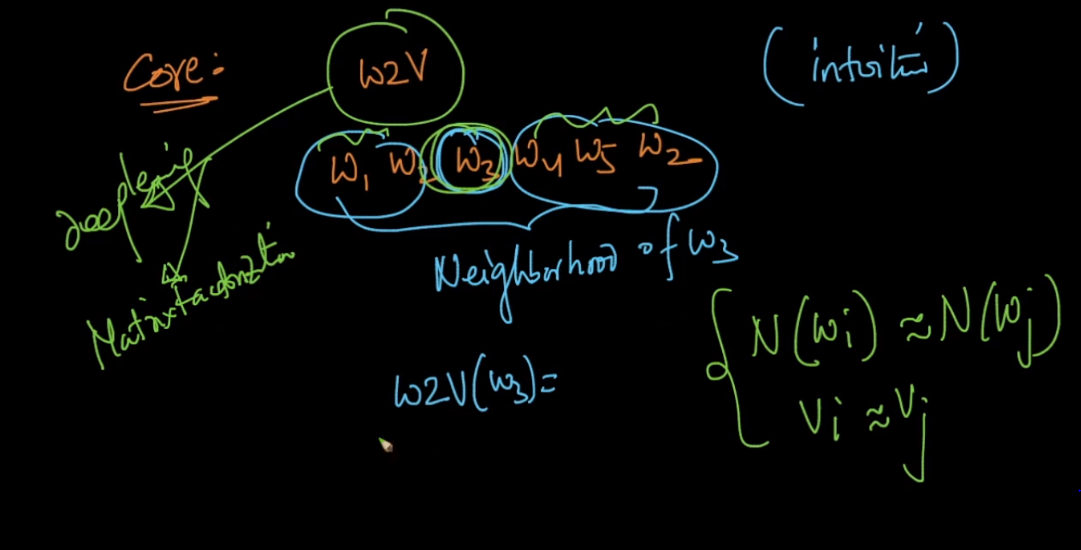


Below fig. shows as data corpus size increases, d dimensionality also increases and as d increases the more information vector have.



Below fig. shows that as words are neighbor to each other their vectors are almost similar or

Vectors are closer.



Notes:

Ques:

We have already trained model ie; glove which has numerous words and corresponding vectors.  
For a new word in my dataset, it first searches that word in its dictionary and returns the corresponding vector. Is it correct?  
If the word is not in glove dict, it is ignored. What if we give syno for a word present in dict, does the model consider this, if so how.

Ans:

Yes, the Word2Vec model gives the vector for a synonym also as it preserves the semantic meaning.   
For example, we have words like 'precious' and 'valuable' in our train data corpus. Now nly one vector will be created for these 2 words and they both are treated as a single feature. Now when in our test data if come across words like 'expensive', 'costly' and 'high-priced', then the same vector that was created for the words 'precious' and 'valuable' is given to these new words also. Word2Vec model takes the semantic meanings of the words into consideration.