* Word 2 vec +>	
	I has a limited diamension vertex.
	y is reduced.
- Sym out	it meaning is maintained.
- deuse vertor is created in word to vec.	
- Juput for word 2 vec is simple iterable Corpus [-,-,-]	
word 2 ver our	
feature	Boy Girl King quen apple mango
Gerder	T-17 1 -0.92 +693 0 0.1
Poyal	0.01 0.02 0.95 0.96 -0.02 0.01
Age	
food	
:	
	This value assigned by The word 2 ver pretrained
300	The word2ver pretrained
	Neural Network
[according to there inbetween	
	Rejation ship]
Now this is the verter related to the	
	overed 'Boy'.
we don't know the must features over here	
(but we can define the Number of the	
fea	tures.
A Contineous Bag of Words:	
Ext	[Krish channel is related to Data Science]
suppose window size is = 5	
Then	

Independant feature 0/P Krish, channel, related, to IS Related chanvel, is, to, data Is, related, data, science TO In BOW How our feature look like : 7 = Vocabulary krish > 1,0,0,0,0,0,0,0 channel 0,1,0,0,0,0,0,0 0,0,1,0,0,0,0,0 live this all this fed to fully connected Neural Network. activation = Softmax word1 -0 +> knish[01,02,0.3,0.4,0.5] 0 - channel 0+35[word2 outer training the Neural Ne twonk Words -9 each node will Represent each word No. Neuron = Window vertor size- windo size = 5 size output layer +x4 Newson No. of Neuron = 7 [Jupits] which will represent The one ward. ISRO 447! DOM5

Skipgram & - This is same process as CBOW but in severse output Erish, channel, related, to channel, is, to, data. Related is related that science. Now The CBOW model, the distributed representation of the Context (or Summaring words) are Combined to predict the word in the middle.

While in skin-and while in stip-gram model, the distributed representation of the input word is used to predict the Gutert. If when to use Stip-gram and when to use CBOW? OCBOW: Dif you have larger dataset and you're primarily Intrested in capturing the meaning of frequent words effectively.

Owhen you have limited computational resources or when training time is concern. ② SHp-gram : O if you have a smaller dataset or you want to capture more information from less frequent D when you want their context a courably, segardters of the training time required.