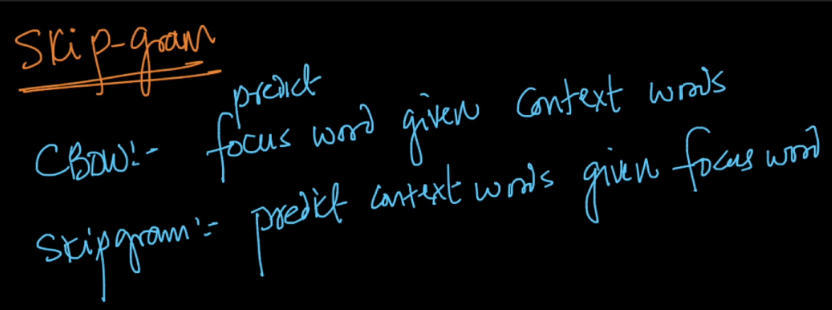
**Skip-gram:**

As in CBOW we were predicting focus word given context words, in **skipgram** we predict context words for given focus word.

It’s reverse of CBOW.

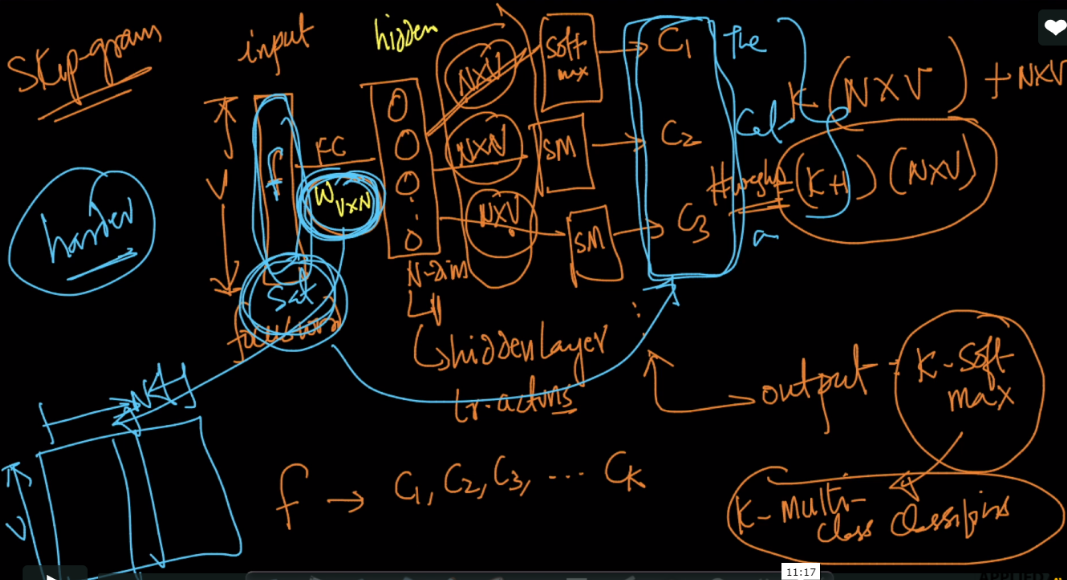


CBOW also have one hidden layer and one ip and op layer.

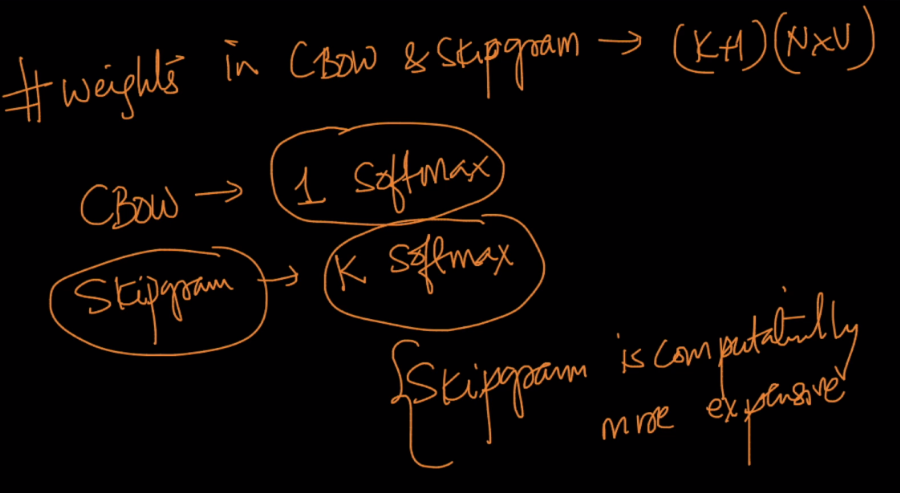
**Input layer:** It has one focus word as input having **v** dimensions.

**Hidden layer:** It has ‘n’ linear activation units and weights between input and hidden layer are used to find word vector or word embedding

**Output layer:** It has ‘k’ softmax classifier, for predicting ‘k’ context words.



If we see weights in CBOW and Skipgram both have (k+1)(n\*v) weights, but since skipgram have more than one softmax classifier so it will take more time that means it is more computationally expensive.



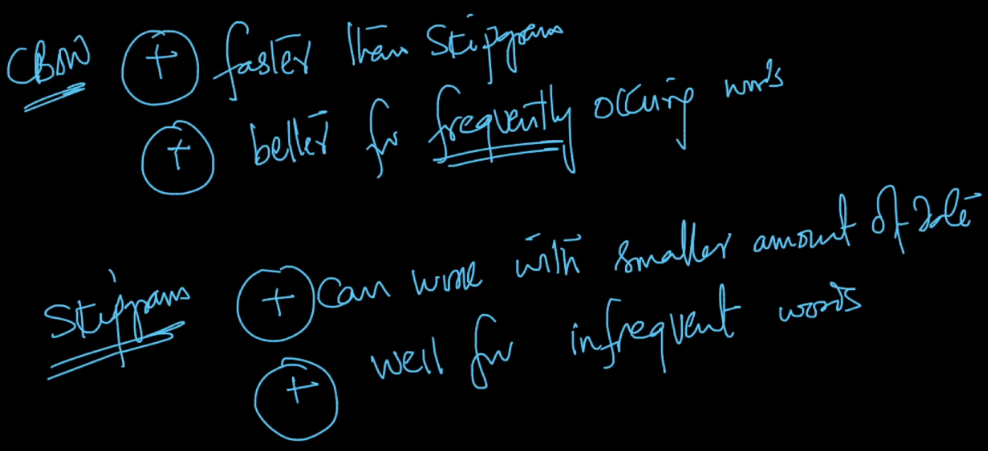
**Advantages:**

**CBOW:**

* Faster than skipgram
* Better for frequently occurring focus words, since focus word appear more and more time the network will be trained more for that word.

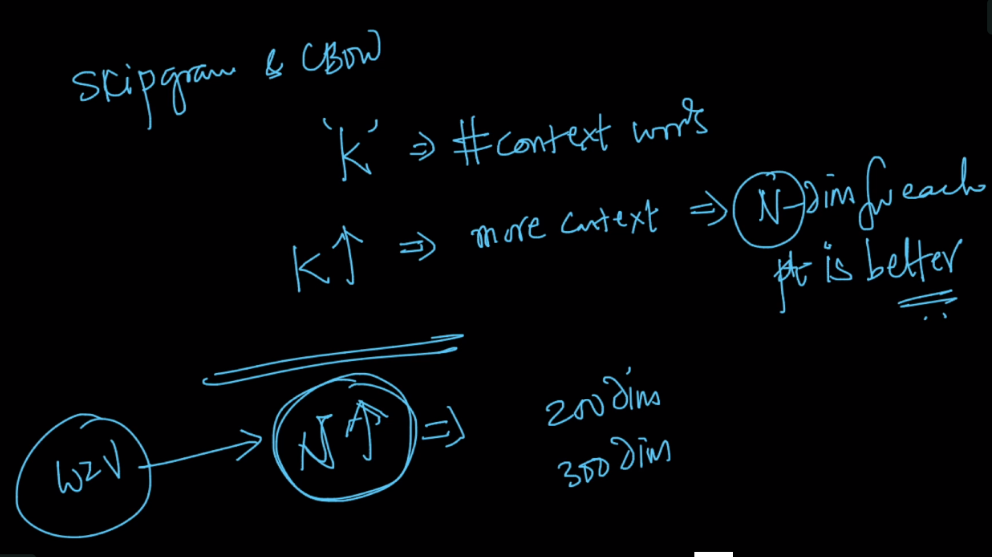
**Skipgram:**

* Can work well with smaller amount of data.
* Well for infrequent words.



In both skipgram and CBOW if we increase context words that is k, then N-dimensions or word vector become better for each poingt

Similarly if ‘n’ or no. of units in hidden layer increases, feature representation also increase.



But there is a drawback of both as we know that weights = (k+1)(n\*v), it becomes huge in terms of memory, so we need optimization for it.

