

# Drawing word2vec

Kai Sasaki



# Training Algorithm

- \* CBOW
- \* Skip-gram
- \* Hierarchical Softmax
- \* Omit Negative Sampling this time :(

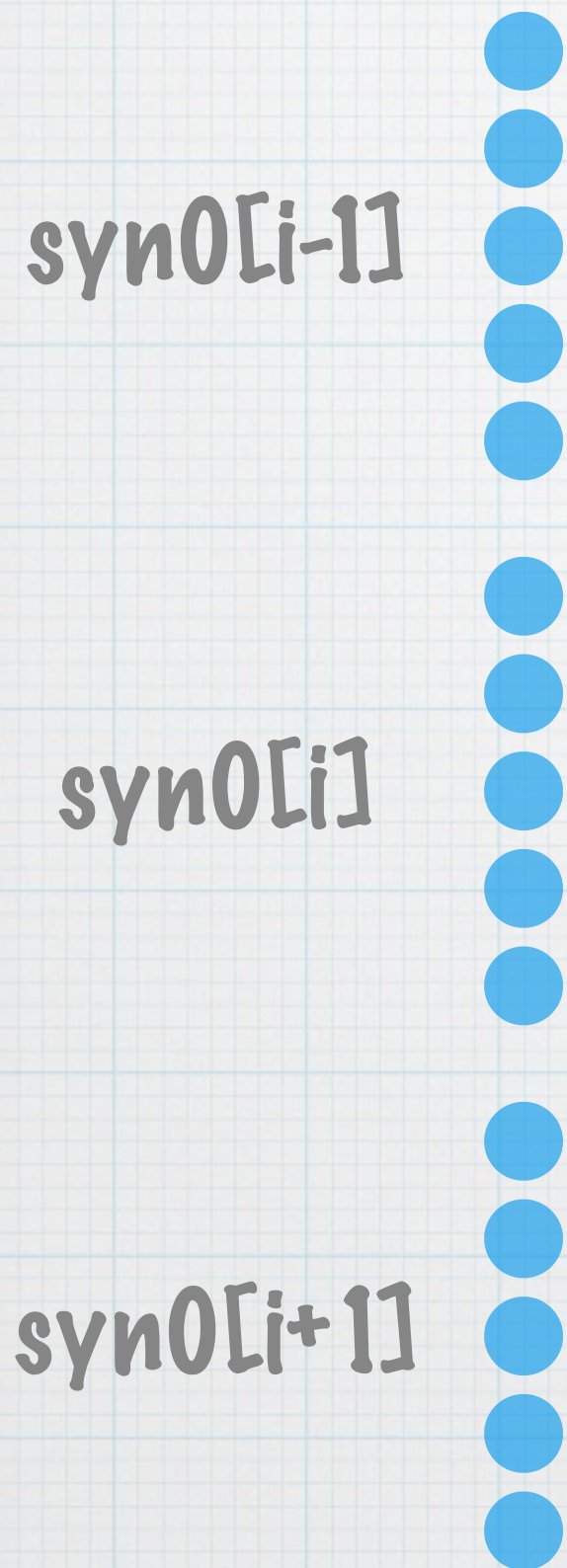


# CBOW

- \* Continuous Bag of Words
- \* Disregard grammar and word order
- \* Share the weight of each words
- \* Training around words

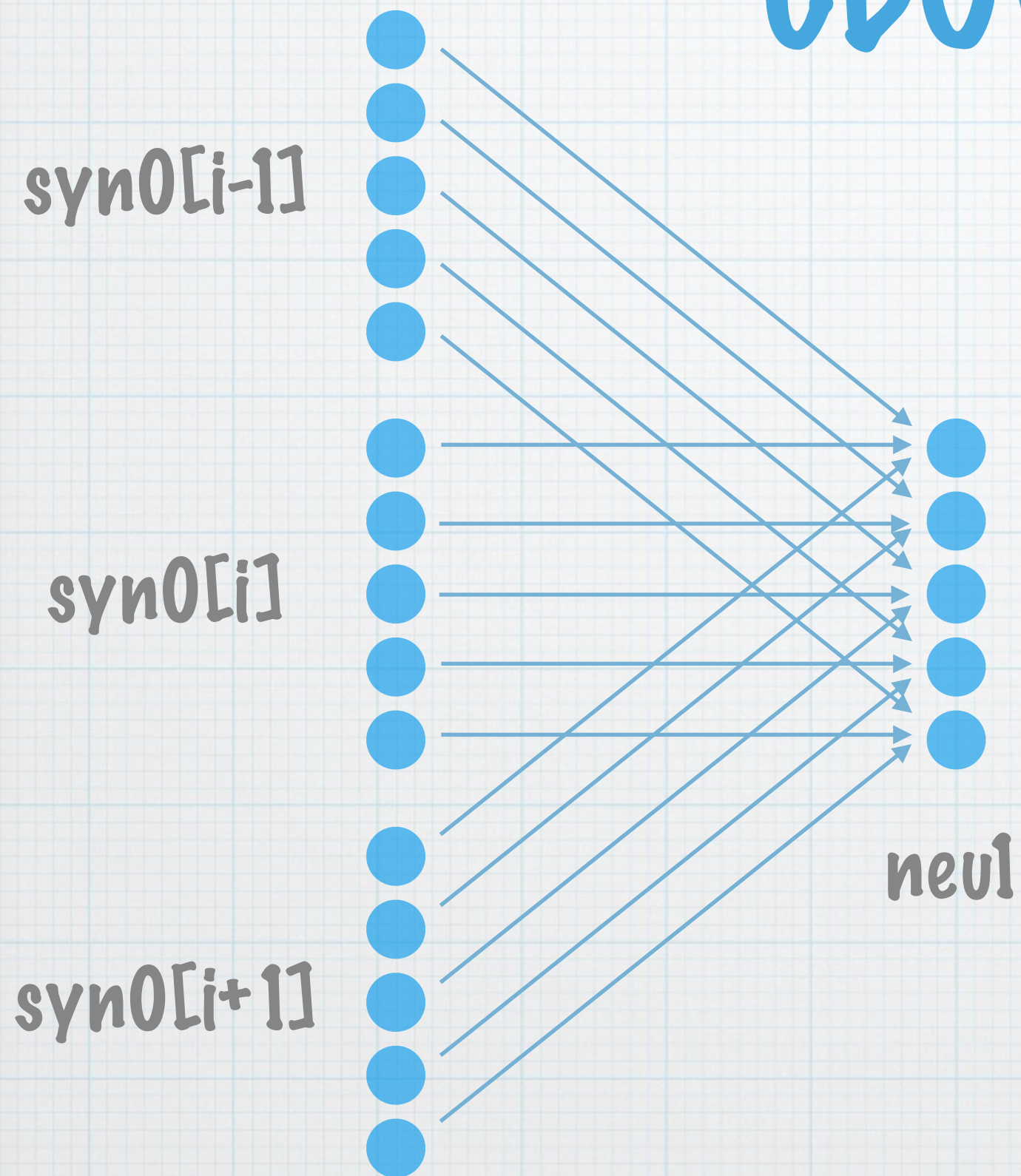


# CBOW



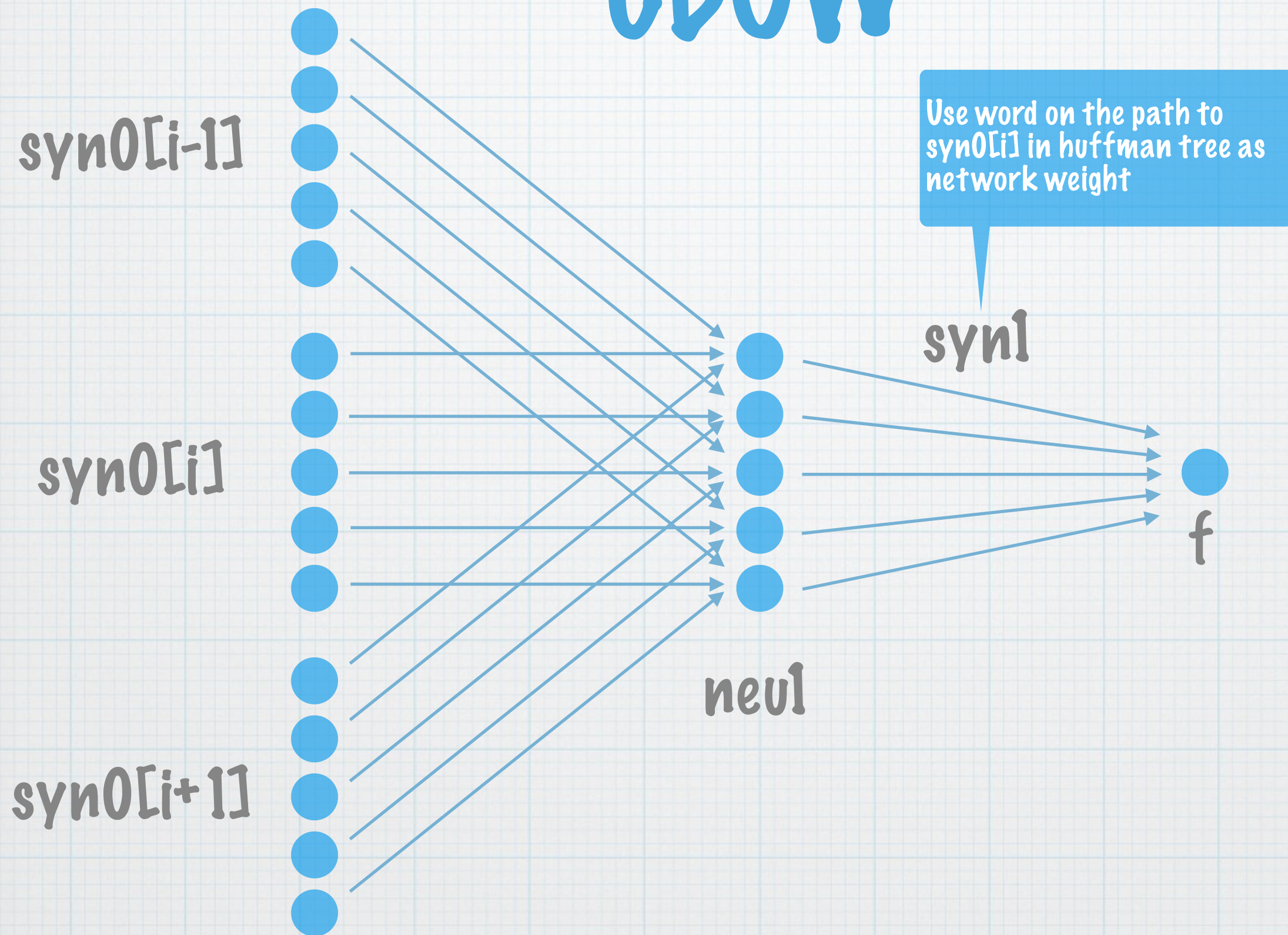


# CBOW



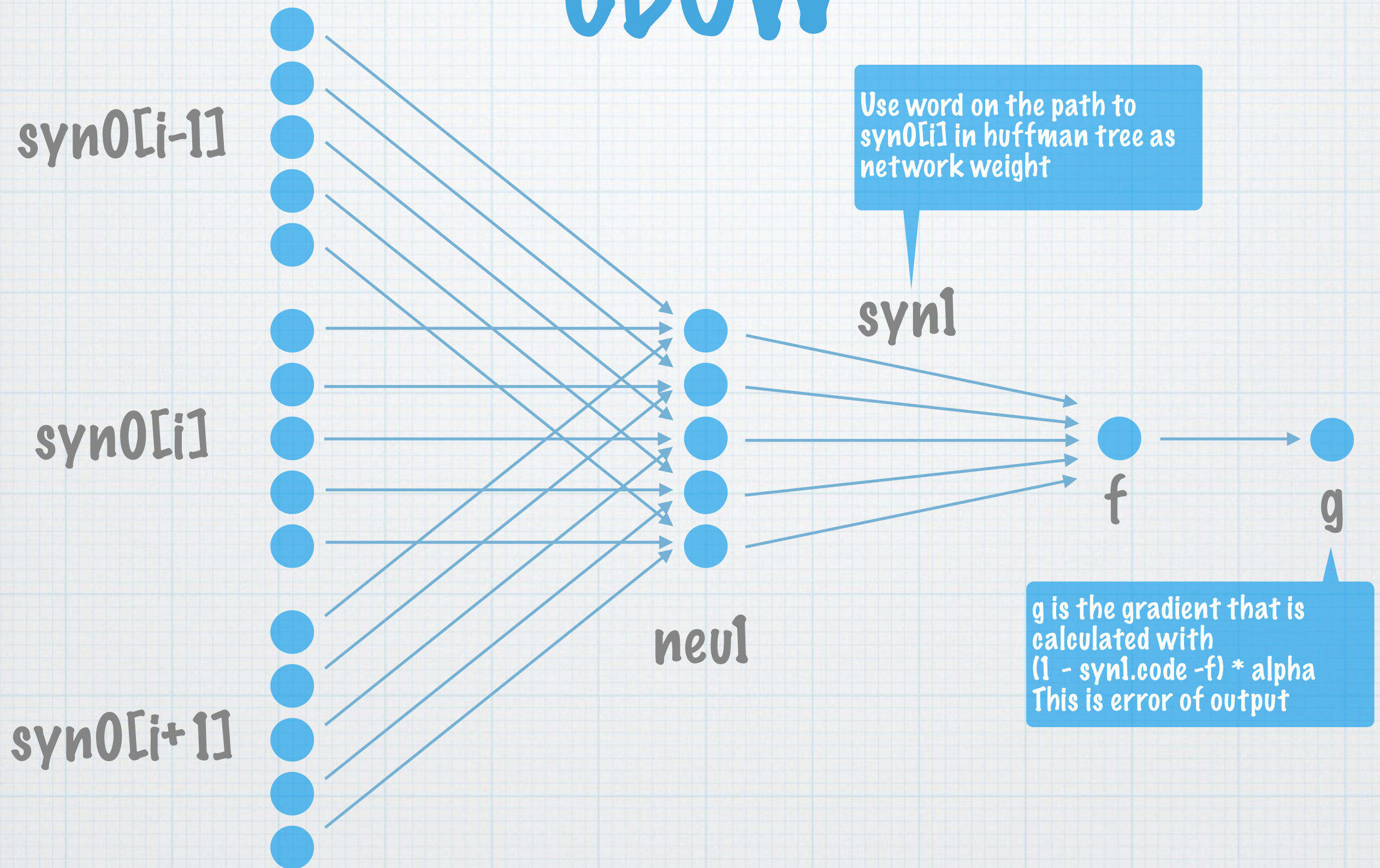


# CBOW



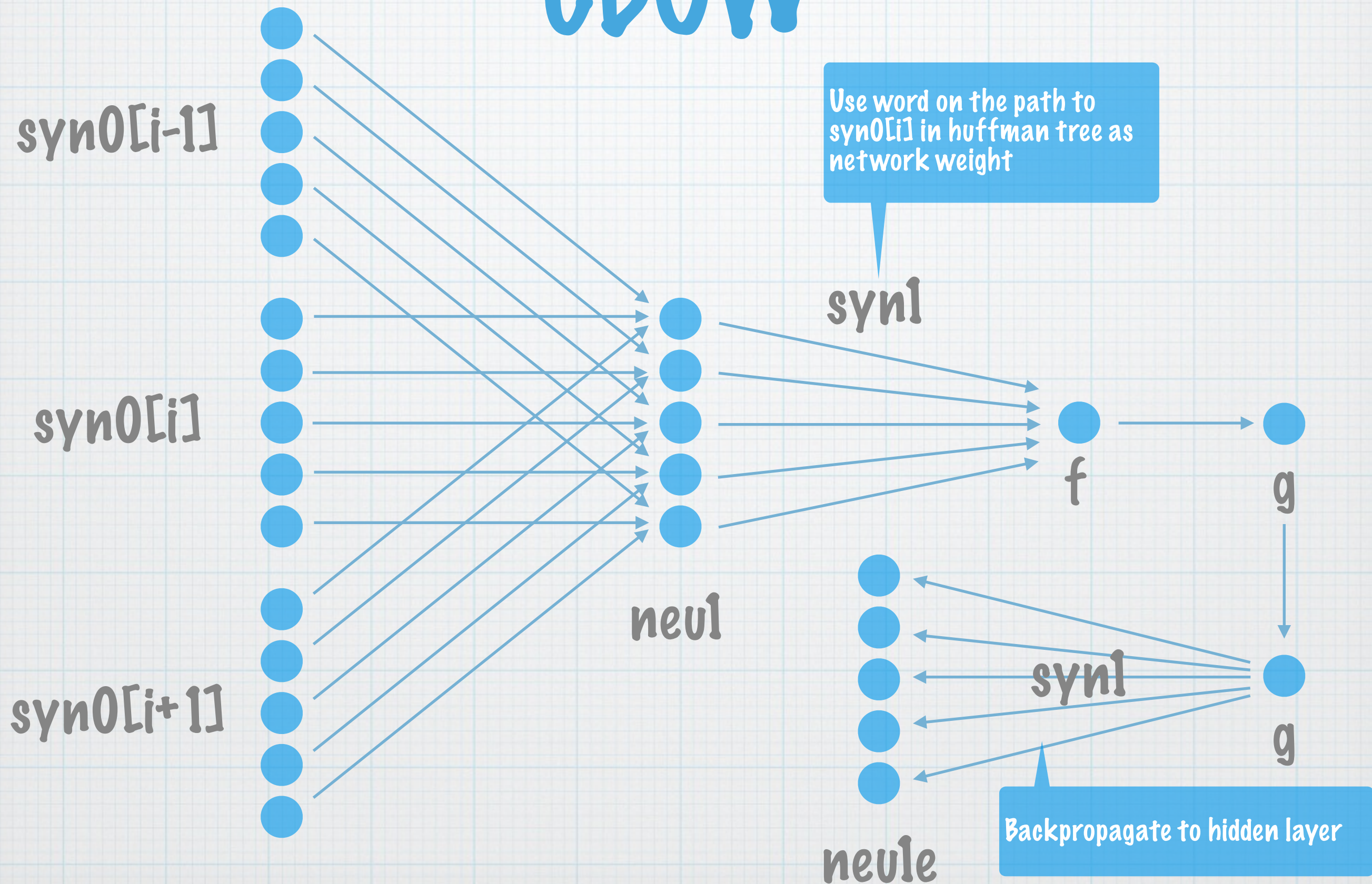


# CBOW



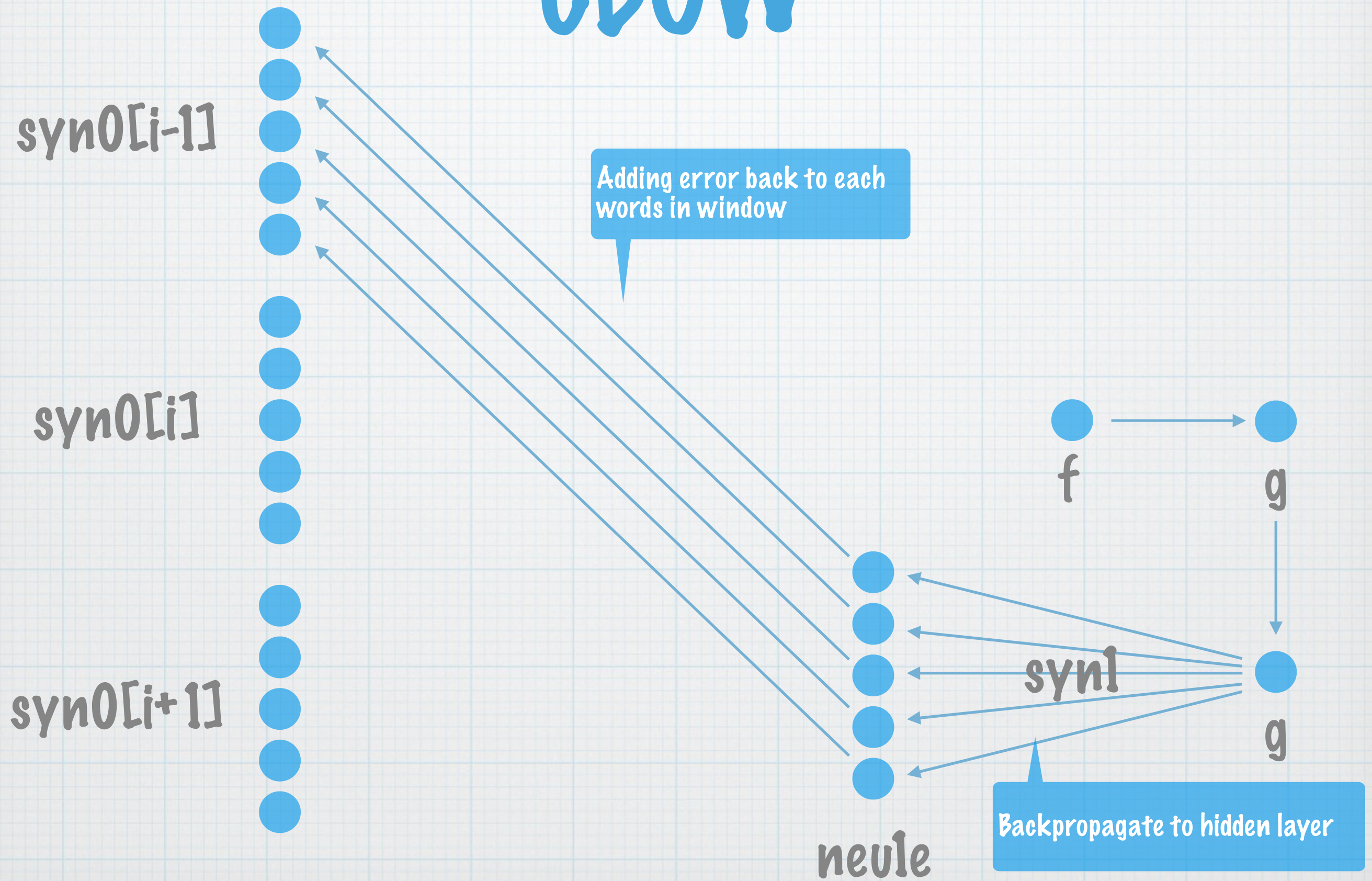


# CBOW



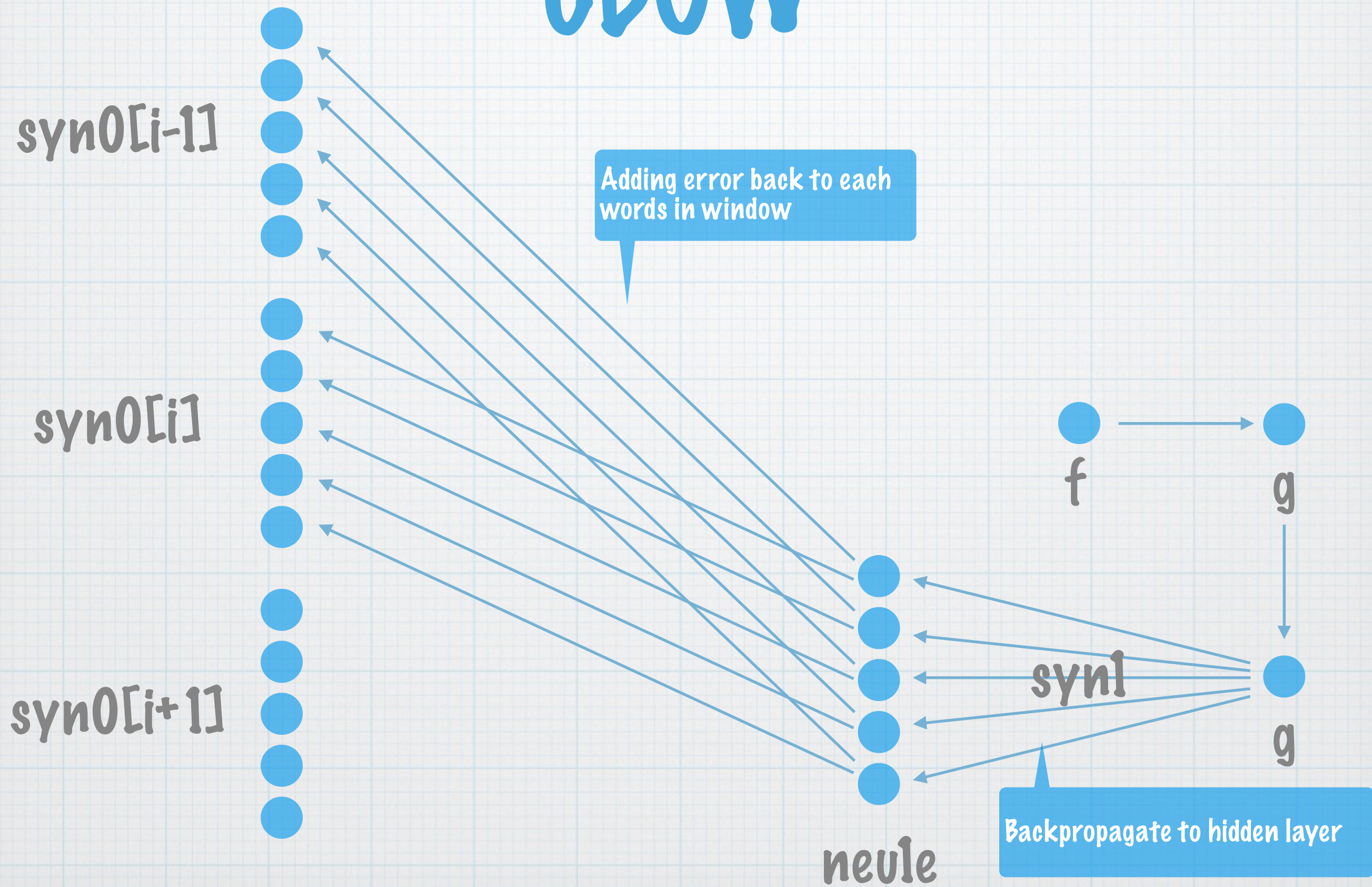


# CBOW



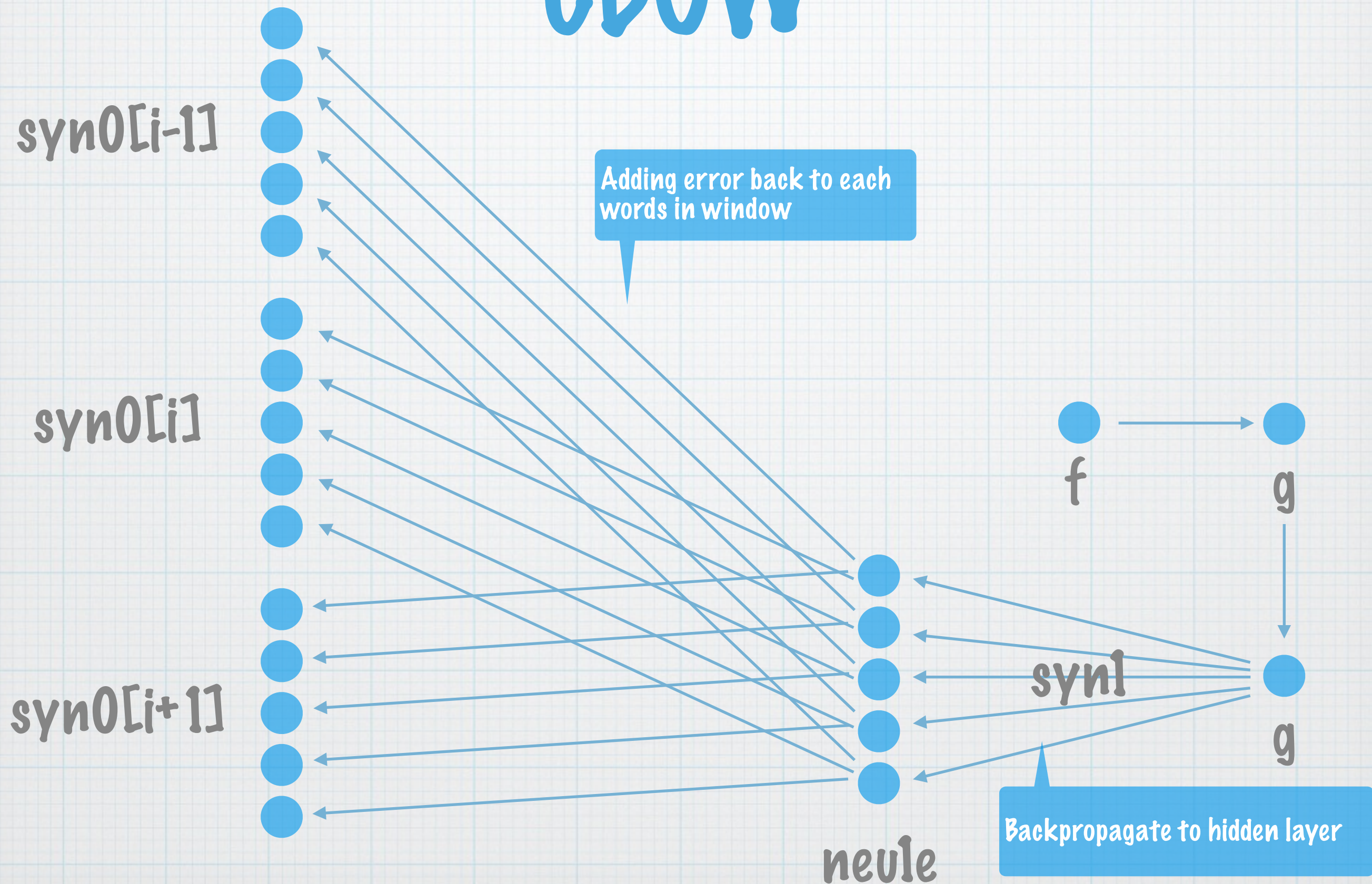


# CBOW



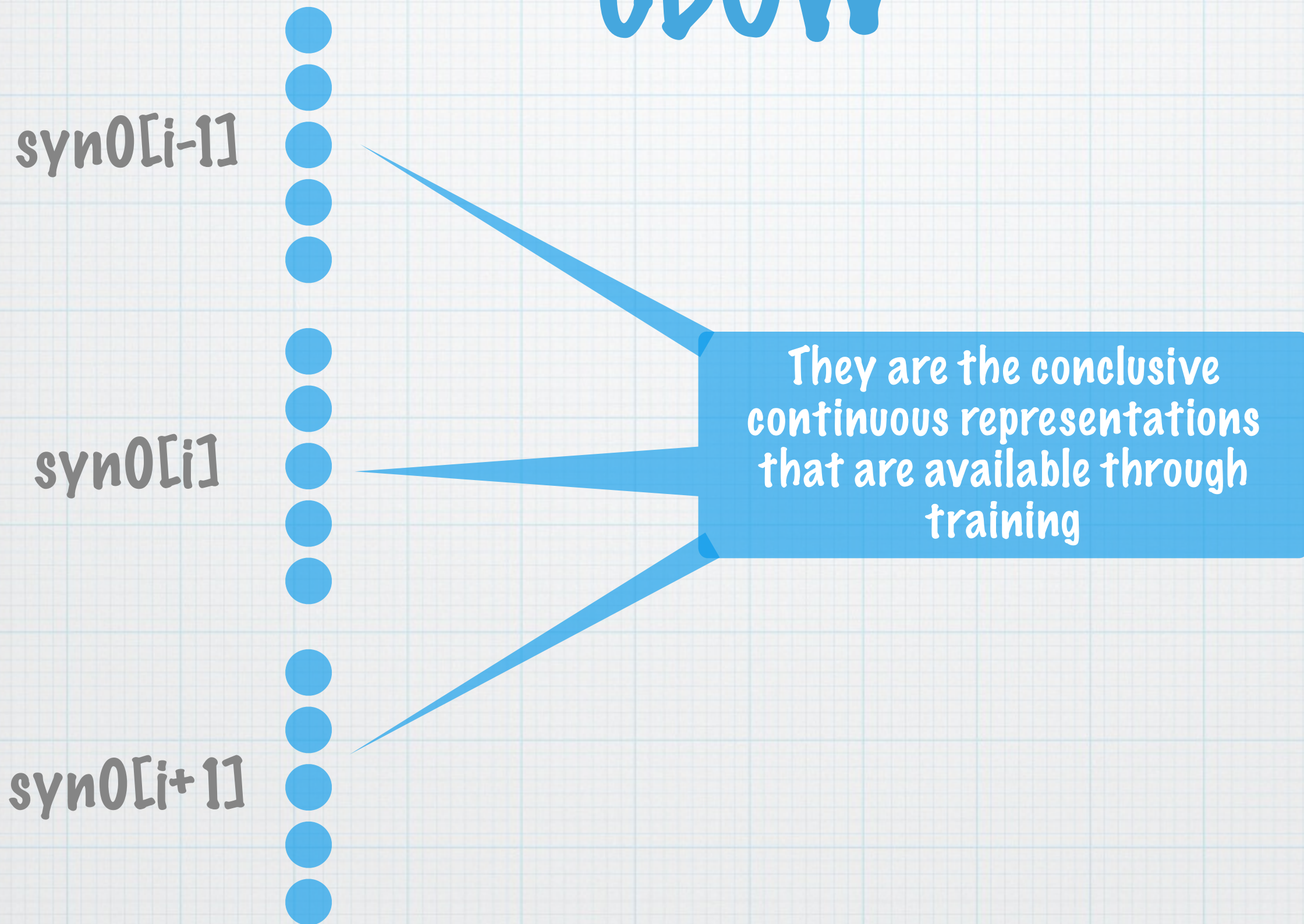


# CBOW





# CBOW





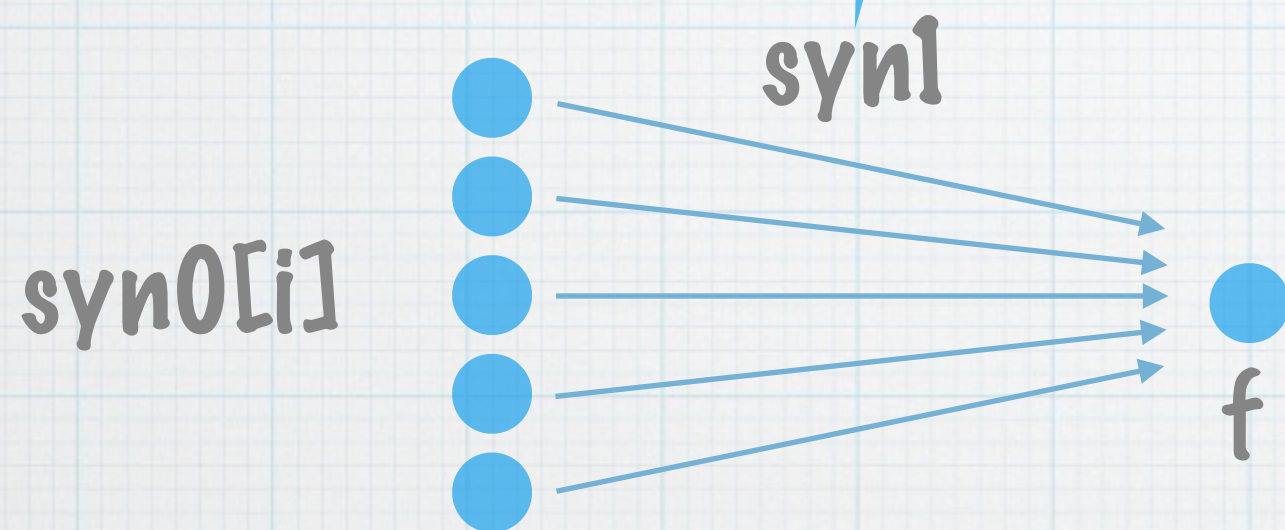
# Skip-gram

- \* Reverse format of CBOW
- \* Predict representations of word that is put around the target words



# Skip-gram

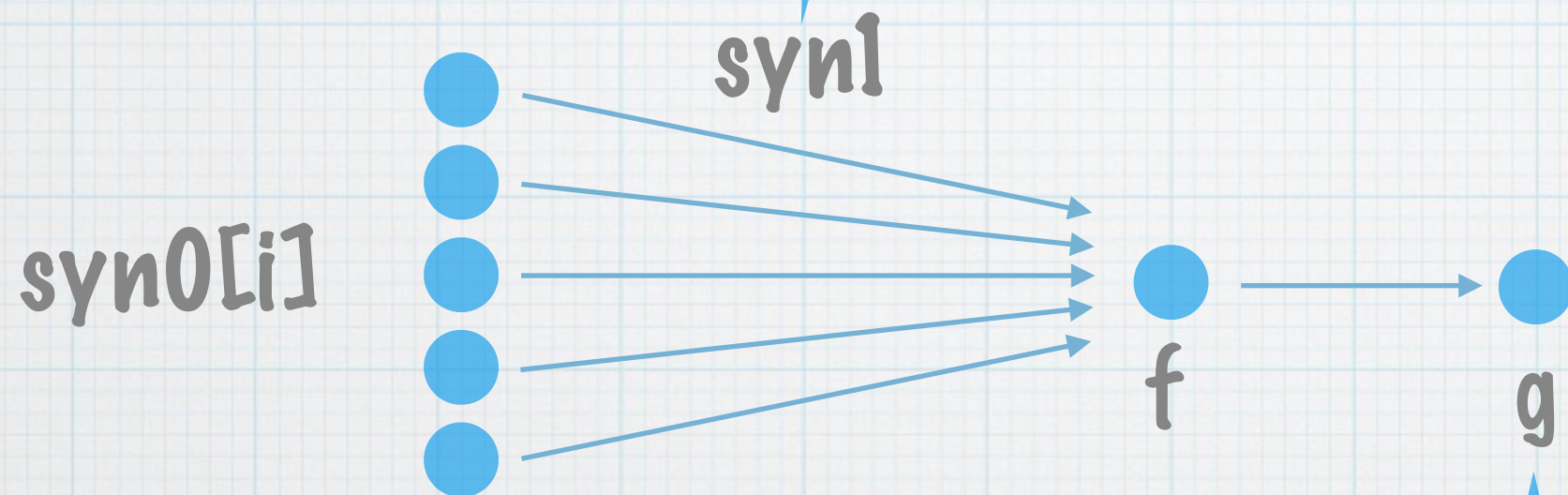
Use word on the path to  $\text{syn0}[i]$  in huffman tree as network weight





# Skip-gram

Use word on the path to  $\text{syn0}[i]$  in huffman tree as network weight

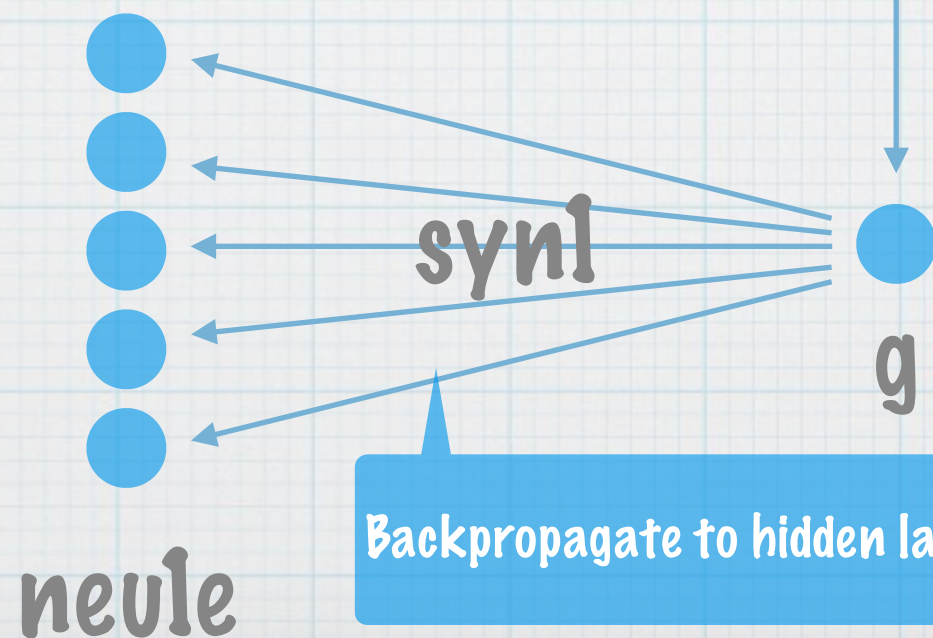
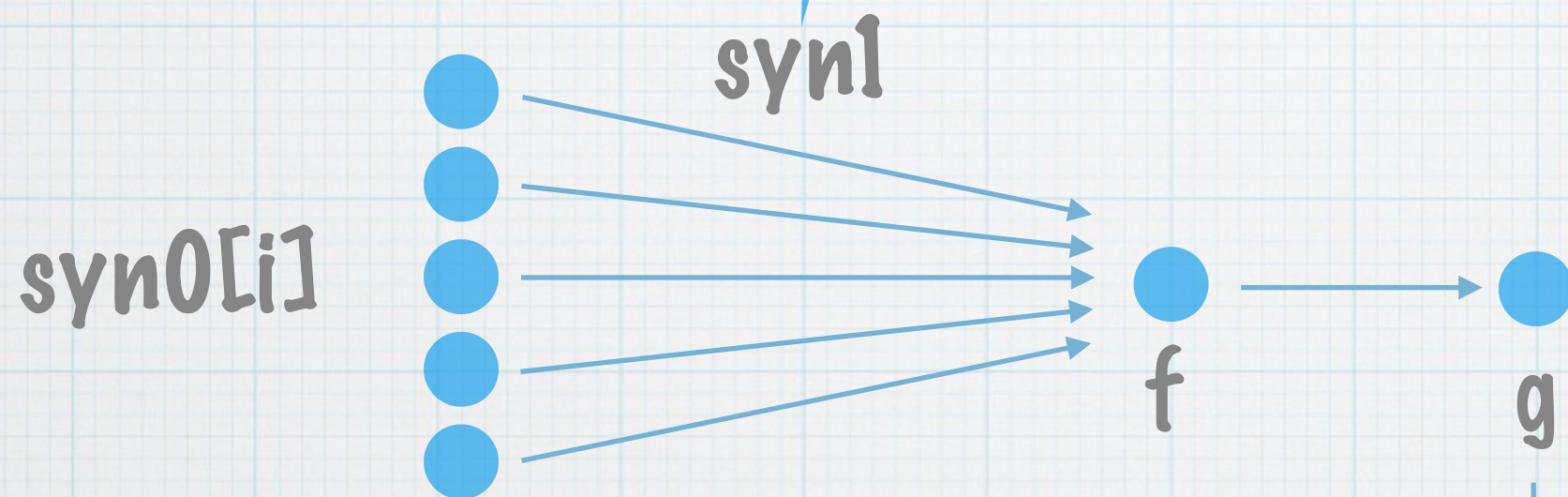


$g$  is the gradient that is calculated with  $(1 - \text{syn1.code} - f) * \alpha$   
This is error of output



# Skip-gram

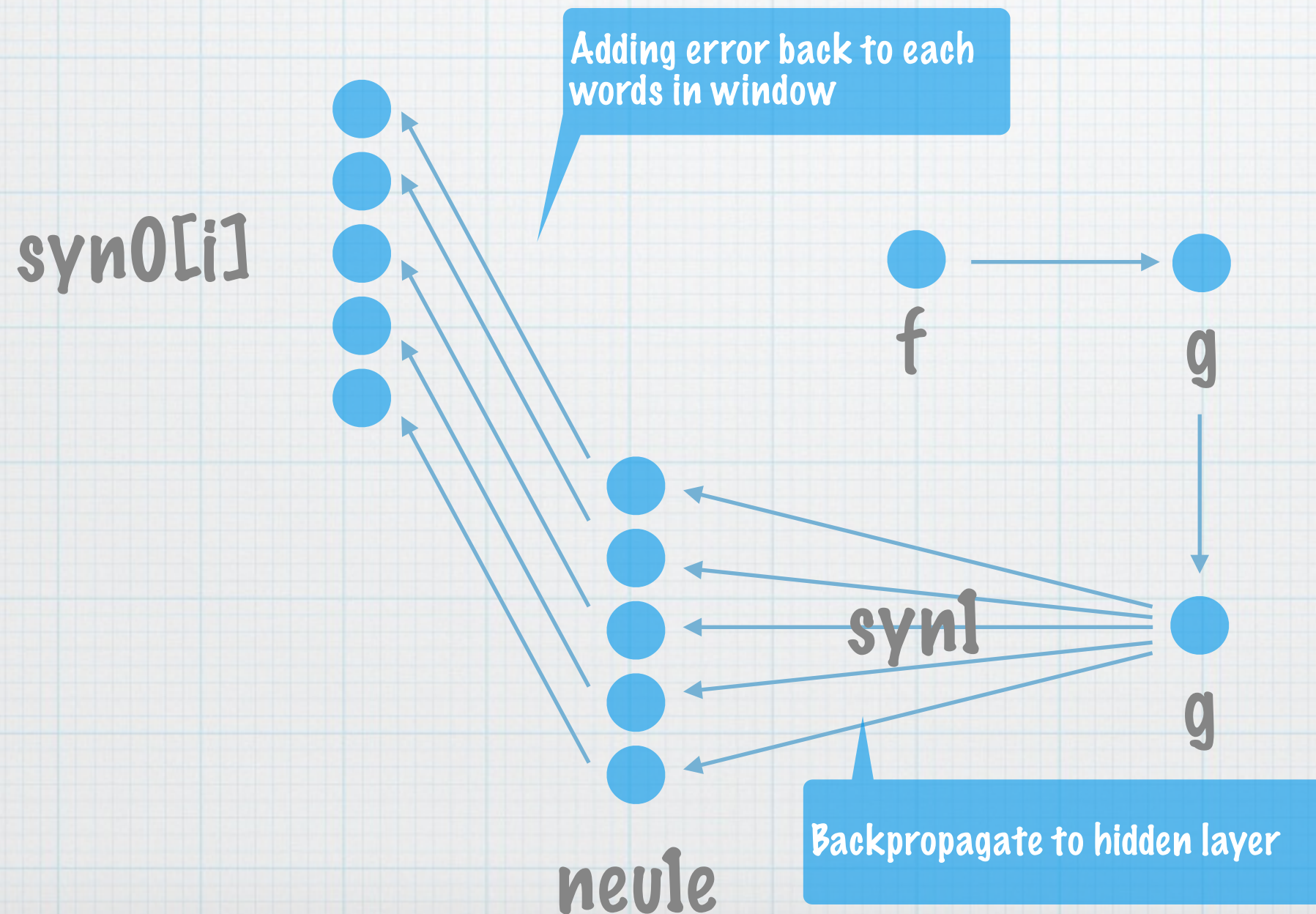
Use word on the path to  $\text{syn0}[i]$  in huffman tree as network weight



Backpropagate to hidden layer



# Skip-gram





# Reference

\* <https://code.google.com/p/word2vec/>