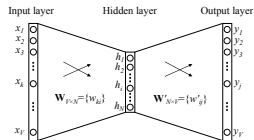
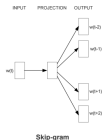


Word Embeddings

- Map each word to a dense vector.
- In practice 300–2000 dimensions are used.
 - The dimensions have no clear interpretation.
- Embeddings are trained for each particular task.
 - NNs: The matrix that maps 1-hot input to the first layer.
- The famous word2vec (Mikolov et al., 2013):
 - CBOW: Predict the word from its four neighbours.
 - Skip-gram: Predict likely neighbours given the word.



Right: CBOW with just a single-word context (<http://www-personal.umich.edu/~ronxin/pdf/w2vexp.pdf>)