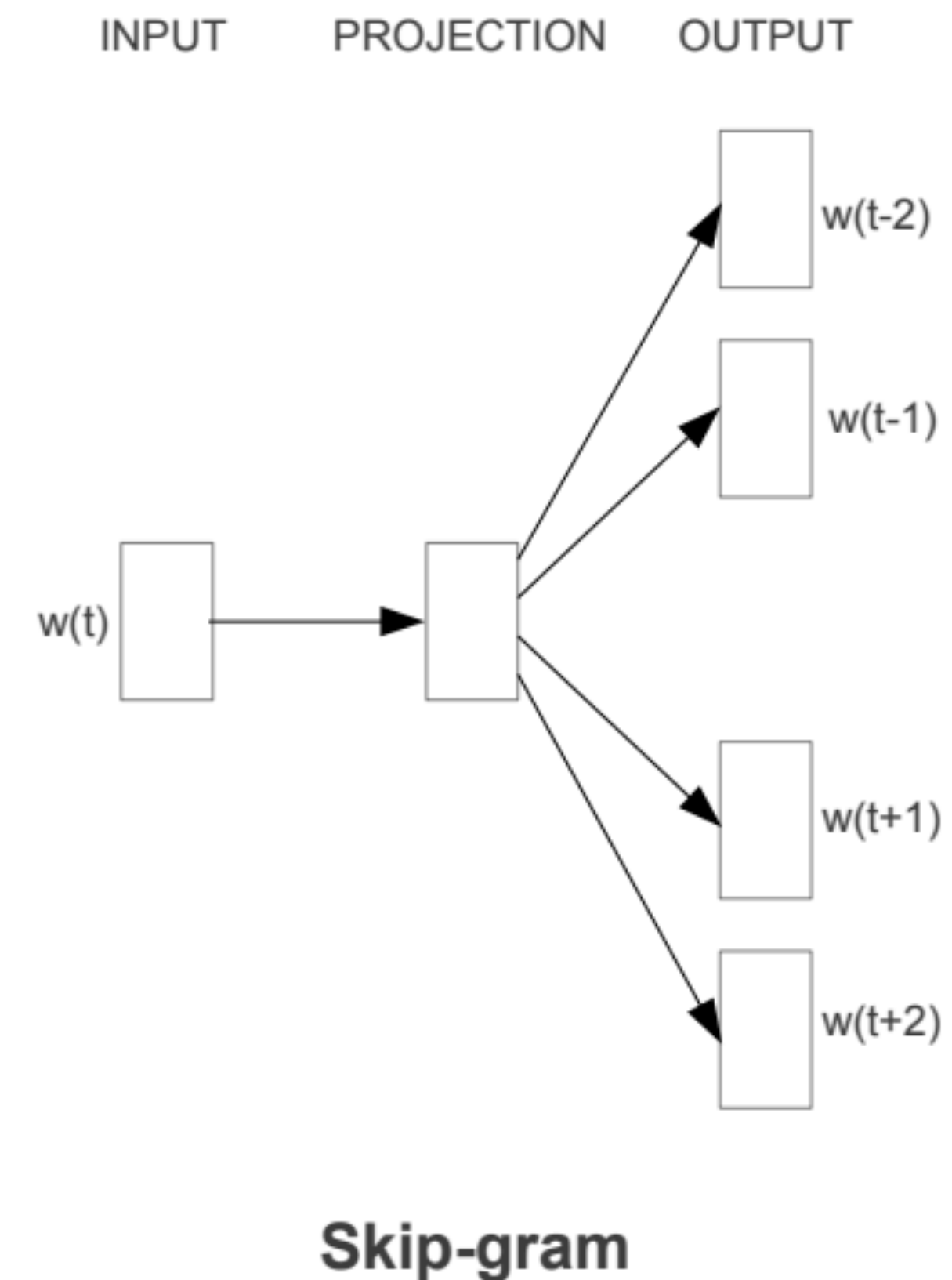
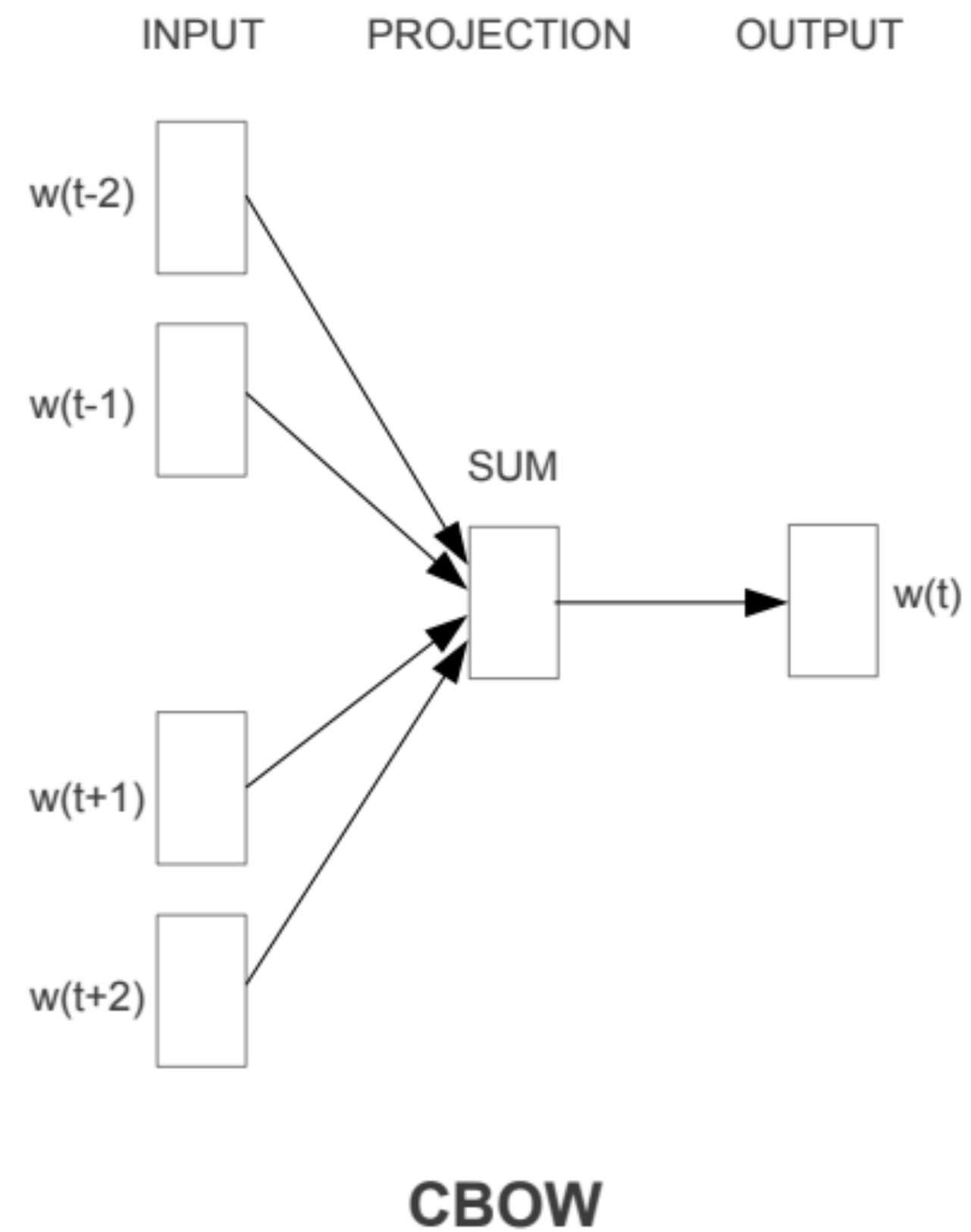


The distributional hypothesis

- States that the similarity between two words can be measured by the similarity of their contexts. This is called the *distributional hypothesis*.
- The natural question is then: how should we best measure similarity?
- Last lecture we learned about ways to compress BOW/TF-IDF features onto a dense sub-manifold, which allows us to measure distances reliably.
- There is fundamental problem with BOW though: it completely strip away contextual information in the text. The meaning of a word can be altered by the context in which it's used. Thus far we've successfully applied BOW to the areas of text classification, topic modeling, and information retrieval, but for other tasks this feature representation is the limiting factor. As we'll find out, moving away from the histogram approach will enable us to improve text classification performance significantly.

Continuous Bag of Words (CBOW) & Skip Gram modeling



- Taken from Mikolov et al., *Efficient Estimation of Word Representations in Vector Space* (2013)