



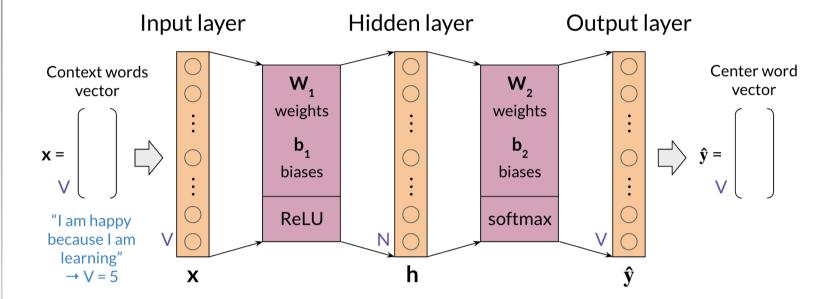
≔ Hide menu

Lecture: Word Embeddings

- Video: Week Introduction 1 min
- Video: Overview 2 min
- **Reading:** Overview
- Video: Basic Word Representations 3 min
- **Reading:** Basic Word Representations 5 min
- Video: Word Embeddings 3 min
- Reading: Word **Embeddings** 4 min
- (Video: How to Create Word Embeddings 3 min
- **Reading:** How to Create Word Embeddings? 4 min
- Video: Word Embedding Methods 3 min
- **Reading:** Word **Embedding Methods**
- Video: Continuous Bagof-Words Model 4 min
- **Reading:** Continuous Bag of Words Model 3 min
- Video: Cleaning and Tokenization 4 min
- Reading: Cleaning and

Architecture for the CBOW Model

The architecture for the CBOW model could be described as follows



You have an input, X, which is the average of all context vectors. You then multiply it by W^1 and add b^1 . The result goes through a ReLU function to give you your hidden layer. That layer is then multiplied by W^2 and you add b^2 . The result goes through a softmax which gives you a distribution over *V*, vocabulary words. You pick the vocabulary word that corresponds to the arg-max of the output.

Mark as completed





□ Dislike

☐ Report an issue



