

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

4 min
- ▶

Video: Continuous Bag-of-Words Model

4 min

Week 4 Training a CBOW Model: Backpropagation and Gradient Descent

PreviousNext

Training a CBOW Model: Backpropagation and Gradient Descent

- Backpropagation:** calculate partial derivatives of cost with respect to weights and biases.

When computing the back-prop in this model, you need to compute the following:

$$\frac{\partial J_{batch}}{\partial \mathbf{W}_1}, \frac{\partial J_{batch}}{\partial \mathbf{W}_2}, \frac{\partial J_{batch}}{\partial \mathbf{b}_1}, \frac{\partial J_{batch}}{\partial \mathbf{b}_2}$$

- Gradient descent:** update weights and biases

Now to update the weights you can iterate as follows:

$$\begin{aligned} \mathbf{W}_1 &:= \mathbf{W}_1 - \alpha \frac{\partial J_{batch}}{\partial \mathbf{W}_1} \\ \mathbf{W}_2 &:= \mathbf{W}_2 - \alpha \frac{\partial J_{batch}}{\partial \mathbf{W}_2} \\ \mathbf{b}_1 &:= \mathbf{b}_1 - \alpha \frac{\partial J_{batch}}{\partial \mathbf{b}_1} \\ \mathbf{b}_2 &:= \mathbf{b}_2 - \alpha \frac{\partial J_{batch}}{\partial \mathbf{b}_2} \end{aligned}$$

A smaller alpha allows for more gradual updates to the weights and biases, whereas a larger number allows for a faster update of the weights. If α is too large, you might not learn anything, if it is too small, your model will take forever to train.

Mark as completed

