coursera

Lab Help

Word Embeddings: Hands On

In previous lecture notebooks you saw all the steps needed to train the CBOW model. This notebook will walk you through how to extract the word embedding vectors from a model.

Let's dive into it!

In [1]:
```
import numpy as np
from utils2 import get_dict
```

Before moving on, you will be provided with some variables needed for further procedures, which should be familiar by now. Also a trained CBOW model will be simulated, the corresponding weights and biases are provided:

In [2]:
```
# Define the tokenized version of the corpus
words = ['i', 'am', 'happy', 'because', 'i', 'am', 'learning']

# Define V. Remember this is the size of the vocabulary
V = 5

# Get 'word2Ind' and 'Ind2word' dictionaries for the tokenized corpus
word2Ind, Ind2word = get_dict(words)


# Define first matrix of weights
W1 = np.array([[ 0.41687358,  0.08854191, -0.23495225,  0.28320538,  0.
41800106],
               [ 0.32735501,  0.22795148, -0.23951958,  0.4117634 , -0.
23924344],
               [ 0.26637602, -0.23846886, -0.37770863, -0.11399446,  0.
34008124]])

# Define second matrix of weights
W2 = np.array([[-0.22182064, -0.43008631,  0.13310965],
               [ 0.08476603,  0.08123194,  0.1772054 ],
               [ 0.1871551 , -0.06107263, -0.1790735 ],
               [ 0.07055222, -0.02015138,  0.36107434],
               [ 0.33480474, -0.39423389, -0.43959196]])

# Define first vector of biases
b1 = np.array([[ 0.09688219],
               [ 0.29239497],
               [-0.27364426]])

# Define second vector of biases
b2 = np.array([[ 0.0352008 ],
               [-0.36393384],
               [-0.12775555],
               [-0.34802326],
```