Lab Help

Word Embeddings: Training the CBOW model

In previous lecture notebooks you saw how to prepare data before feeding it to a continuous bag-of-words model, the model itself, its architecture and activation functions. This notebook will walk you through:

- Forward propagation.
- Cross-entropy loss.
- Backpropagation.
- Gradient descent.

Which are concepts necessary to understand how the training of the model works.

Let's dive into it!

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

Forward propagation

Let's dive into the neural network itself, which is shown below with all the dimensions and formulas you'll need.

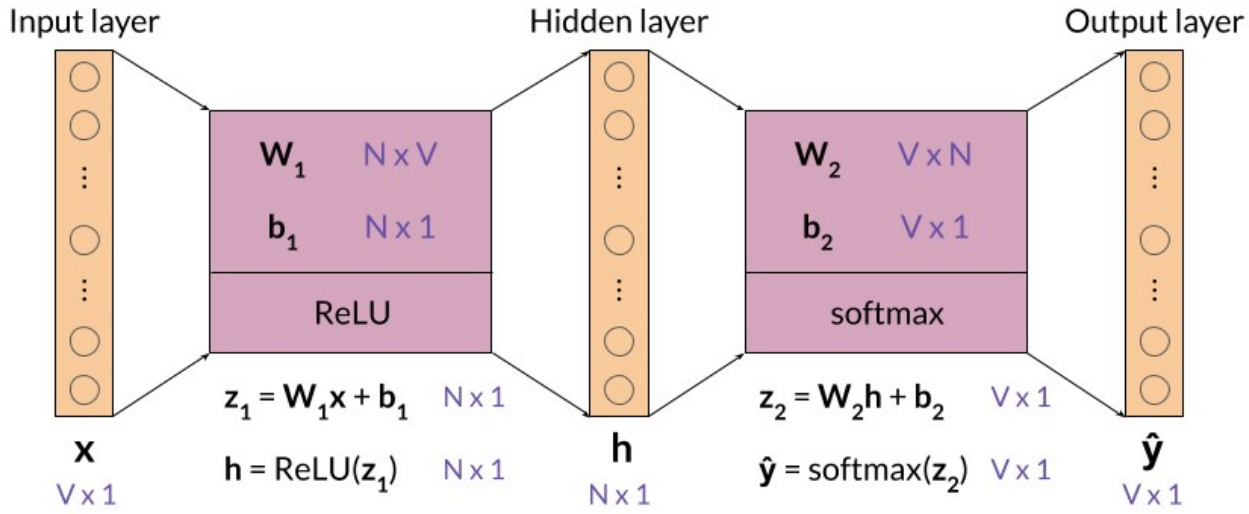

Figure 2

Set N equal to 3. Remember that N is a hyperparameter of the CBOW model that represents the size of the word embedding vectors, as well as the size of the hidden layer.

Also set V equal to 5, which is the size of the vocabulary we have used so far.