5. Implement the Continuous Bag of Words (CBOW) Model. Stages can be:

a. Data preparation

b. Generate training data

c. Train model

d. Output

Theory:

Word2vec is considered one of the biggest breakthroughs in the development of natural language processing. The reason behind this is because it is easy to understand and use. Word2vec is basically a word embedding technique that is used to convert the words in the dataset to vectors so that the machine understands. Each unique word in your data is assigned to a vector and these vectors vary in dimensions depending on the length of the word.

The word2vec model has two different architectures to create the word embeddings. They are:

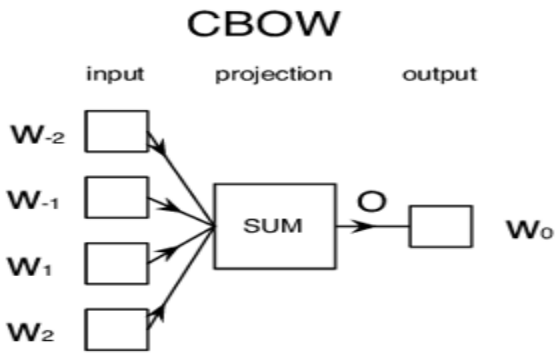
1. Continuous bag of words(CBOW)
2. Skip-gram model

In this article, we will learn about what CBOW is, the model architecture and the implementation of a CBOW model on a custom dataset.

What is the CBOW Model?

The CBOW model tries to understand the context of the words and takes this as input. It then tries to predict words that are contextually accurate. Let us consider an example for understanding this. Consider the sentence: 'It is a pleasant day' and the word 'pleasant' goes as input to the neural network. We are trying to predict the word 'day' here. We will use the one-hot encoding for the input words and measure the error rates with the one-hot encoded target word. Doing this will help us predict the output based on the word with least error.

The Model Architecture



The CBOW model architecture is as shown above. The model tries to predict the target word by trying to understand the context of the surrounding words. Consider the same sentence as above, 'It is a pleasant day'. The model converts this sentence into word pairs in the form (contextword, targetword). The user will have to set the window size. If the window for the context word is 2 then the word pairs would look like this: ([it, a], is), ([is, pleasant], a),([a, day], pleasant). With these word pairs, the model tries to predict the target word considered the context words.

If we have 4 context words used for predicting one target word the input layer will be in the form of four 1XW input vectors. These input vectors will be passed to the hidden layer where it is multiplied by a WXN matrix. Finally, the 1XN output from the hidden layer enters the sum layer where an element-wise summation is performed on the vectors before a final activation is performed and the output is obtained.

The output shows the words that are most similar to the word 'virus' along with the sequence or degree of similarity. The words like symptoms and incubation are contextually very accurate with the word virus which proves that CBOW model successfully understands the context of the data.

Conclusion

In the above article, we saw what a CBOW model is and how it works. We also implemented the model on a custom dataset and got good output. The purpose here was to give you a high-level idea of what word embeddings are and how CBOW is useful. These can be used for text recognition, speech to text conversion etc.