Carlos Adrian Bueso

Dr. Dmitriy Dligach - COMP329 Natural Language Processing - Homework 4 Written Report

For this project, we had to train a word2vec model on a corpus of our choice. In my case, I decided to use the Brown Corpus, easily accessible from the NLTK library in python. The Brown corpus consists of 500 texts, each consisting of just over 2,000 words. Another library, gensim, was used for implementing the word2vec algorithm. Because I had previous experience with Google Colaboratory, I decided to use it for training the model.