Assignment3 Report

Student Name: Guohao Ma

Student Number: 20676560

Q1: Use genism library to get the most similar words to a given word. Find 20 most similar words to "good" and "bad". Are the words most similar to "good" positive, and words most similar to "bad" negative? Why this is or isn't the case? Explain your intuition briefly.

A1: The words most similar to "good" are basically the positive words such as "great" and "nice". So are the words similar to "bad". That's because these most similar words are basically the synonyms of "good" or "bad". In this case, they have the most similar context, which means they will be close in the high dimension embedding space. However, we also observe "bad" among the words most similar to "good" and vice versa. It seems contrary to the common sense that "good" and "bad" are probably on the opposite direction in the embedding space. But the truth is that the environment or the context of "good" and "bad" are quite similar. Just imagine "good" and "not-bad", then it's not surprising to find antonyms like "good" and "bad" will be considered as 'similar' by word2vec, which is not the case actually.