MSCI-641 Assignment 3

Explanation of Results:

According to my results, most of the words similar to "good" are positive words like "decent" , "great" and "wonderful". However, there are also some negative words like "poor" and "bad" in the list of most similar words for good. This is true in the case of the word "bad" as well. While most of the similar words of "bad" are negative like "terrible" , "horrible" , "awful" however there are some positive words like "good" and "alright" included in the list of similar words of "bad". This seems counter intuitive because it seems that positive words should have all positive words in its similar list and so should be the case with negative words. However, positive and negative words are similar to each other because their context is similar to each other. Word2Vec does not capture similarity based on antonyms and synonyms but gives a higher similarity to words appearing in similar context. For example in "He is a good mannered person" & "He is a bad mannered person" the context of both the sentences is same therefore "good" and "bad" would be considered similar to each other. Hence Word2Vec captures Paradigmatic relations – relation that holds between elements of the same category rather than Syntagmatic relations – relation existing between elements that are combined together. As a result some negative words are found in the list of similar words of a positive word and vice versa. In addition, positive words are similar to a positive word and so is the case with negative word in general because again their context is the same. This is quite intuitive because the training corpus contains examples in which most positive and negative words are used in similar context hence in the embedding vector their similarity is high.