

## **Assignment 3**

**MSCI 641**

**13<sup>th</sup> June 2019**

### **Question:**

**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**

### **Answer:**

Looking at the results for both good and bad we can see that the similar words received are in fact similar (synonyms). However, there are still some cases where the antonyms are showing up, i.e. looking at words similar to **good** you can see “bad”, “terrible” and same is the case when looking at words similar to **bad**. This is because of a couple of reasons. One of the reasons is that Word2Vec keeps track of diction. Non-semantic properties of words matter to word2vec, the way the things are said/written has a lot of dependencies which can be cultural/language oriented for example: the British use phrases that the Americans do not use, hence the reviews come from people that have a diverse background due to which such a thing becomes evident.

Another reason is that Word2Vec does not capture similarity based on words actually being synonyms and antonyms but rather based on the similarity of the context in which they are used. For example: The movie got good reviews vs The movie got bad reviews, the context is regarding the movie reviews which is same but they can be good/bad, and hence considered similar to each other.