**Assignment3 Report**

-Atish Telang Patil, 20793593

Few words most similar to good are:

* excellent
* decent
* fantastic
* great
* terrible
* nice
* horrible

Few words most similar to bad are:

* negative
* point
* surprised
* honestly
* write
* saying

There are more words similar to “good” than “bad”. Words like ‘excellent’, ‘decent’, ‘great’ are similar to good whereas words like ‘point’, ‘surprised’, ‘write’ are not similar to the word “bad”.

The similarity is calculated using the cosine similarity, each word is represented as a vector.

The cosine similarity is then calculated between two words, the words like ‘terrible’, ‘horrible’ have similar vector representation as the word “good” therefore the cosine similarity for these words are high.

Similar is the case for the words similar to “bad”. The words ‘surprised’, ‘write’, ‘saying’ have the same vector representation that is the reason why they have high cosine similarity.

While calculating the similarity it takes into account nearing neighbours. If only one word is given to find the words similar to it, that will create a bias.

While getting similar words to “good”, if we add another condition to find the words opposite to “bad”, we get better result:

**model.most\_similar(positive=["good"],negative=["bad"],topn=5)**

* perfect
* great
* perfectly
* nicely
* beautiful

It also depends on the vocabulary, if the vocabulary does not have synonyms of “bad” then it will not be possible for the function to return the exact similar words to “bad”, it will just find the words that are near to the “bad” (that have similar vector representation).