

MSCI 641

Assignment 3: Word2Vec

Vyas Anirudh Akundy
Student ID: 20765080

1 Intuition:

- The words most similar to "good" are almost all positive and the words most similar to "bad" are almost all negative.
- My intuition is that all the words that are similar to either good or bad appear in the same context and hence have similar word vector representations.
- Taking one of the examples from the lectures, after training a huge corpus, the vectors for "king" and "man" seem to be close to each other though the words are different because they appear in similar contexts
- However one interesting thing to point out is that out of all the similar words for "good", one of them is 'bad' and out of all the similar words for "bad" one of them is 'good'.
- This can be due to the fact that the words "good" and "bad" themselves are similar to each other because of being used in similar sentences. As both of them express sentiment, and one of the dimensions in their vector representations maybe capturing the sentiment of the word, thus "good" and "bad" may have the same value for this particular dimension. This argument also applies to how all "positive"(and few negative) words are similar to "good" and how all "negative"(and few positive) words are similar to "bad"