

# Assignment 3

## Report

Hariharan Kalimuthu, 2020115015

### ▼ Explain negative sampling. How do we approximate the word2vec training computation using this technique?

Negative sampling is a technique used in word embedding training, like word2vec, to learn word meanings from large amounts of text data. It helps predict the likelihood of a word appearing in a context with another word by only focusing on a few negative examples (words that aren't in the context) instead of all the words in the vocabulary. This makes it faster and more efficient. By sampling a few negative examples, the model learns to predict whether a word is positive or negative based on its context using a sigmoid function. The weights of the word vectors are updated during training using stochastic gradient descent, allowing the model to learn the relationships between words in the text. Overall, negative sampling makes it easier to learn meaningful word representations that capture their meaning.

### ▼ Explain the concept of semantic similarity and how it is measured using word embeddings. Describe at least two techniques for measuring semantic similarity using word embeddings.

Semantic similarity refers to the degree of similarity between the meanings of two or more words or phrases. Word embeddings are numerical representations of words that capture their semantic and syntactic relationships.

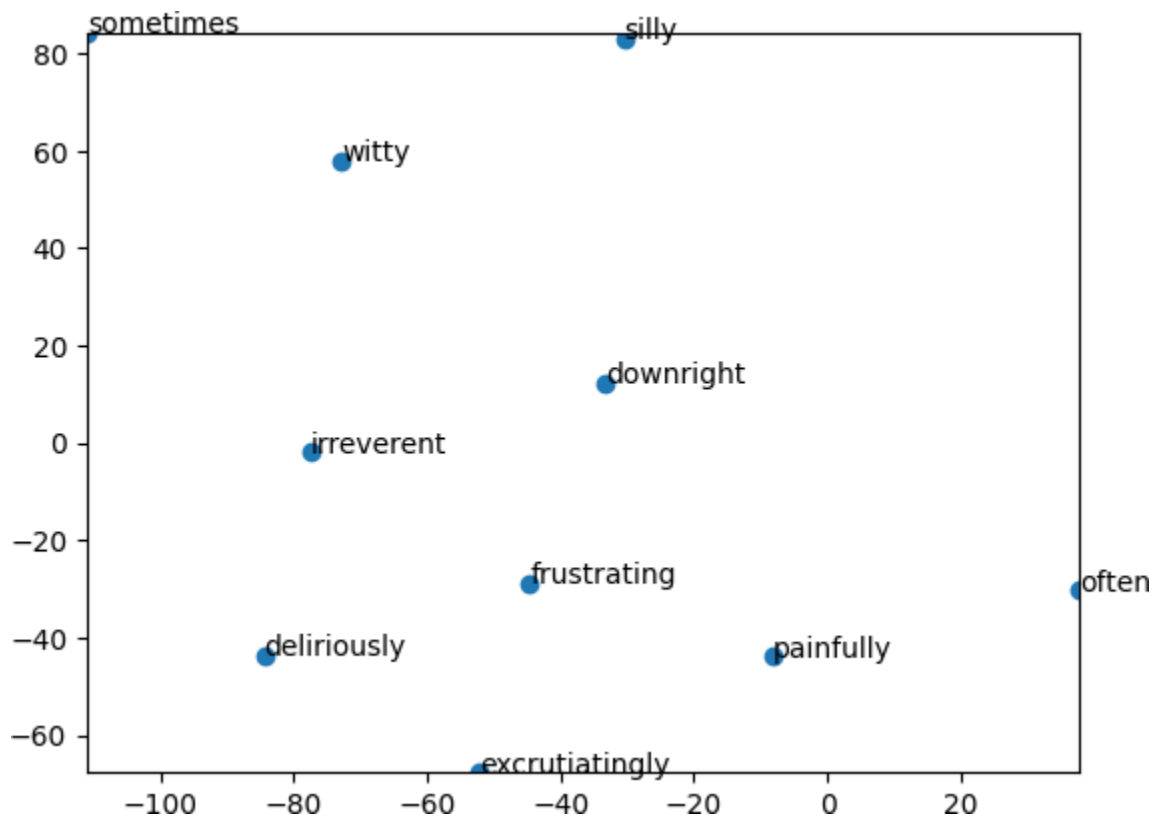
Two techniques:

1. **Cosine Similarity:** Cosine similarity measures the similarity between two words based on the cosine angle between their corresponding vectors.
2. **Word Mover's Distance:** WMD calculates the minimum distance that words in one text document need to move to match the words in another document.

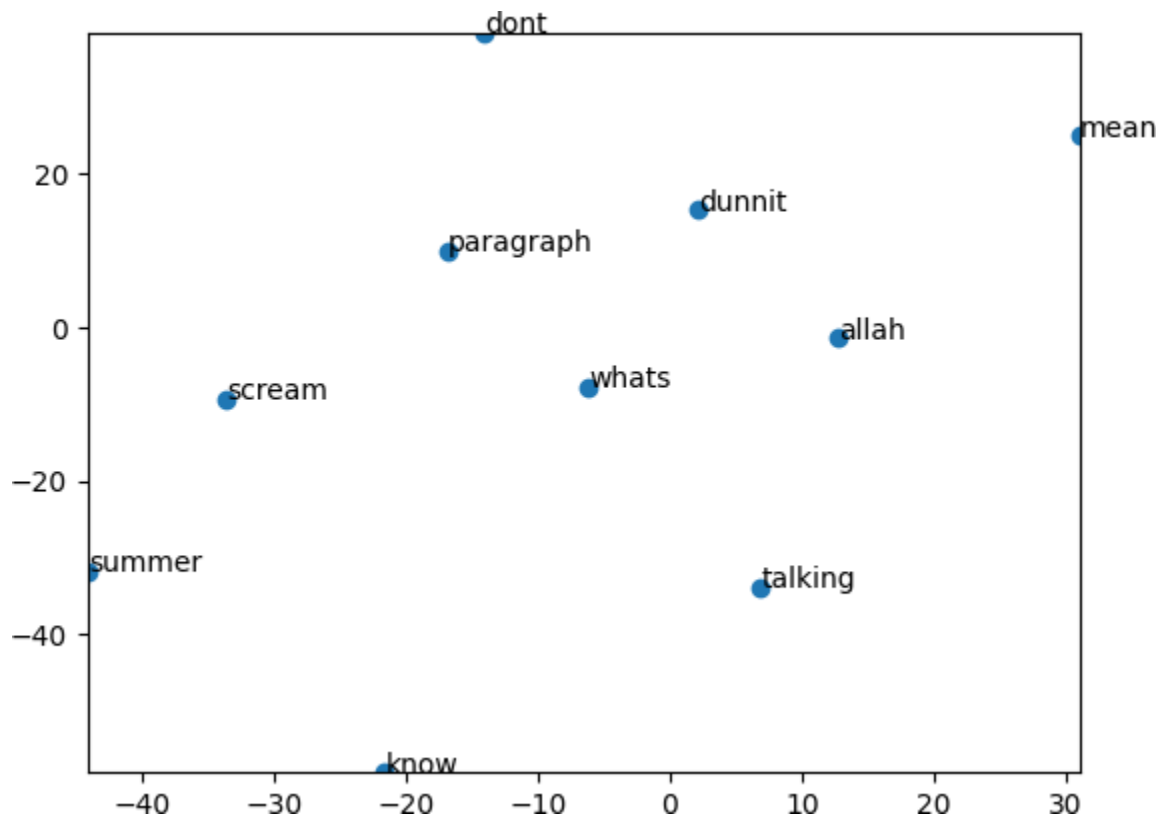
## SVD

### ▼ Question 1 & 2

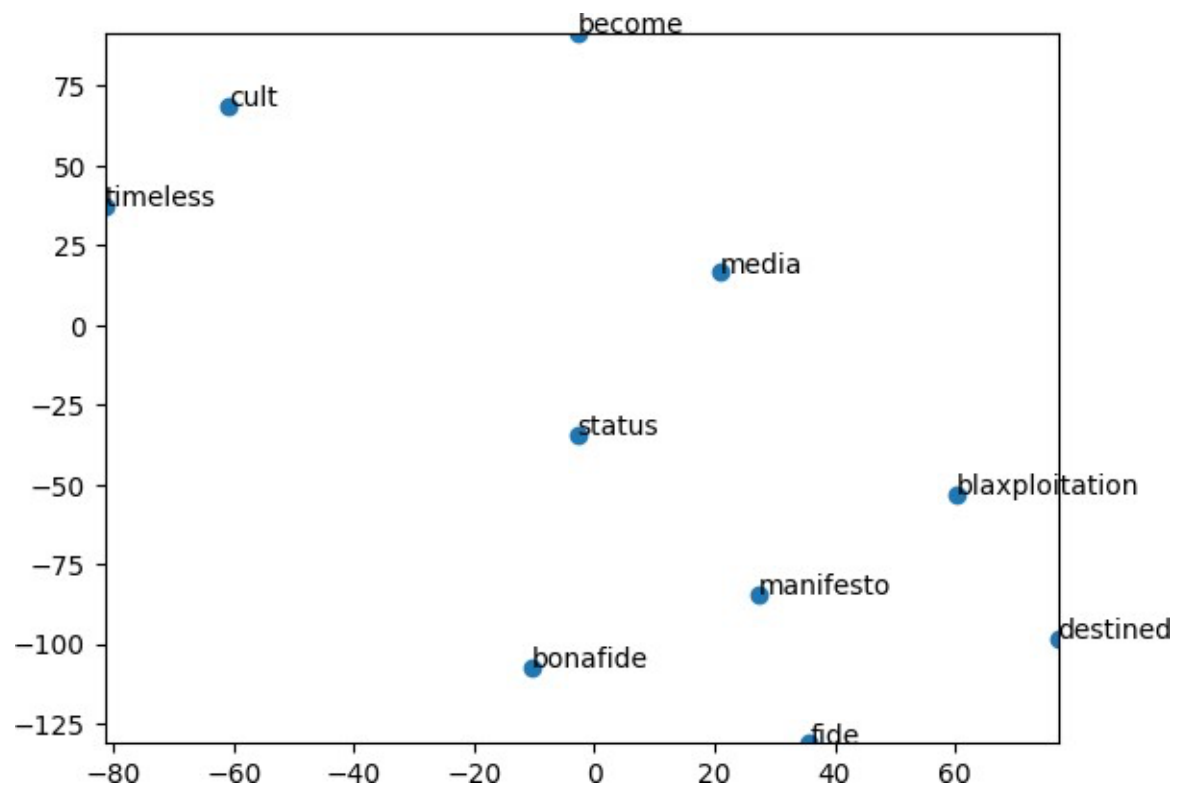
Sometimes



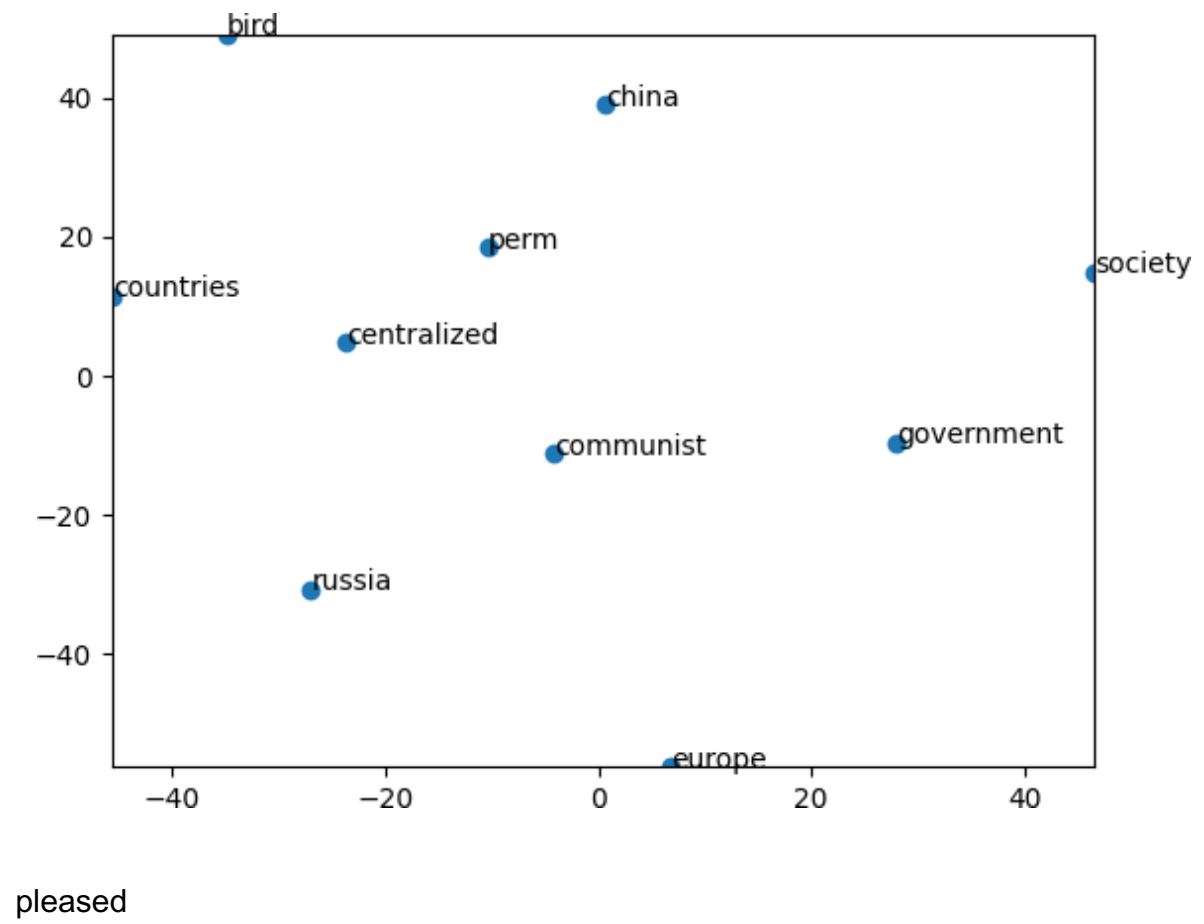
know

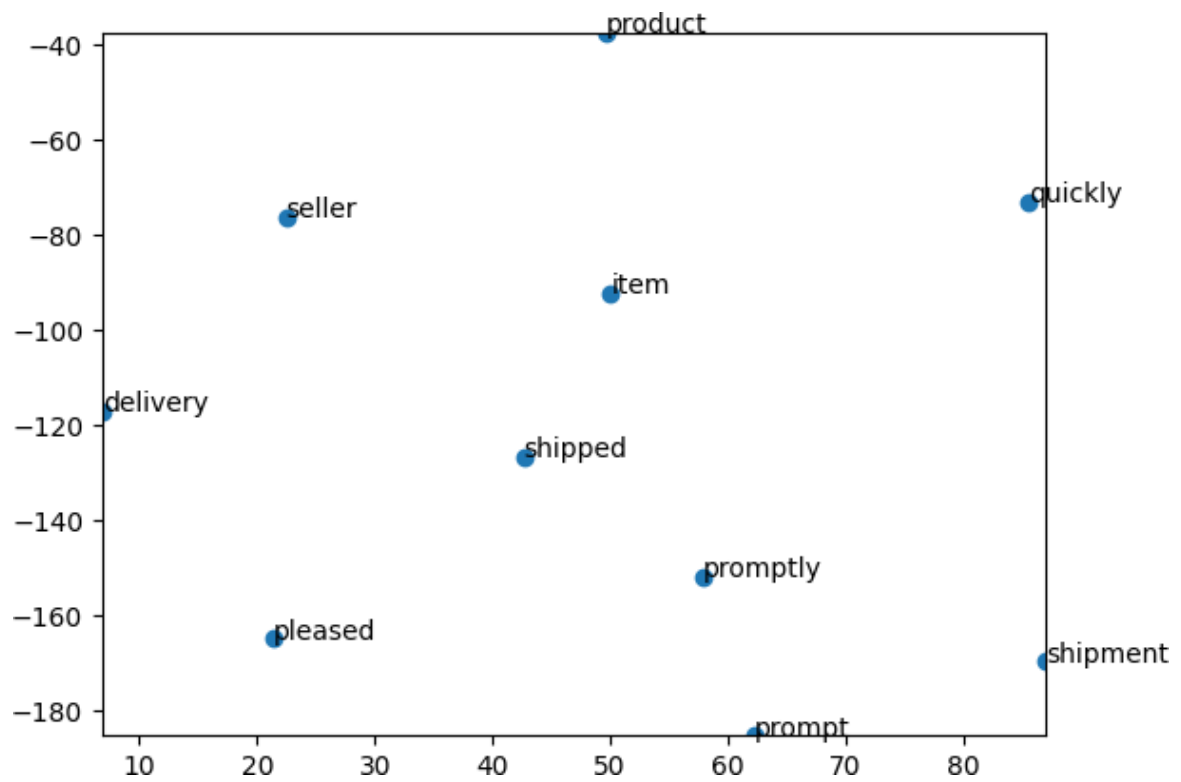


media

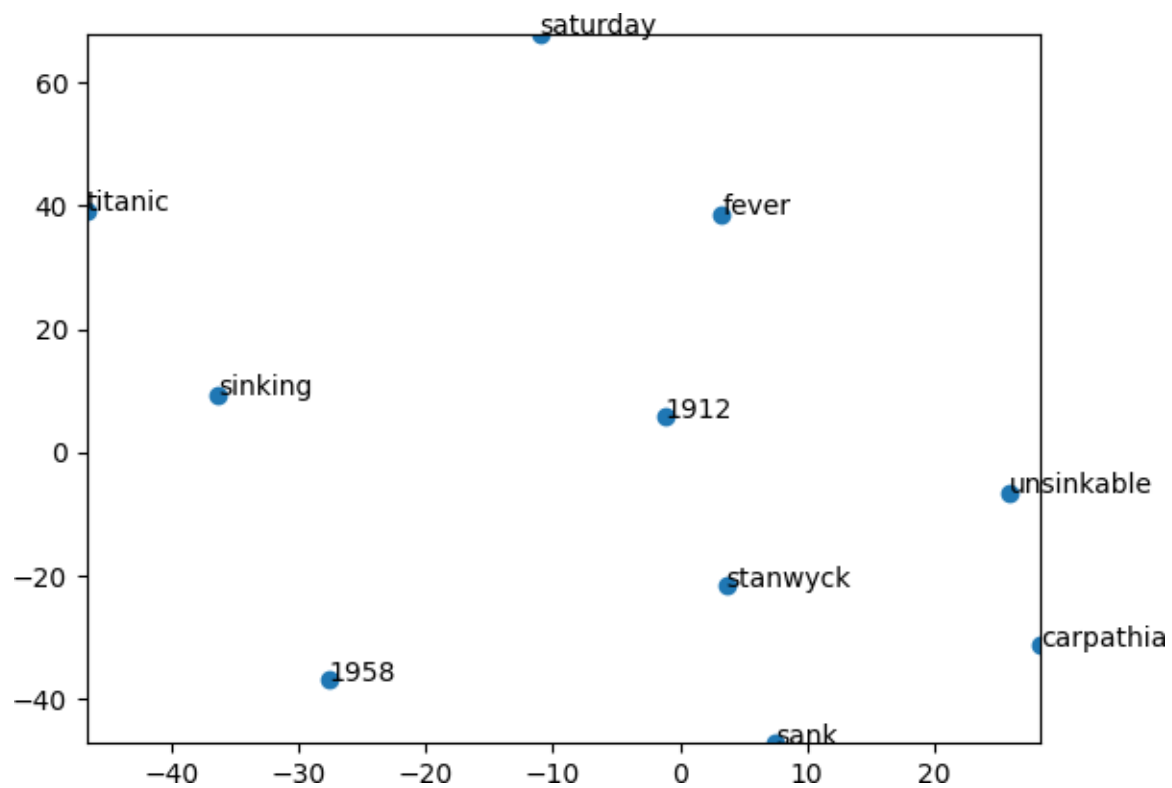


china

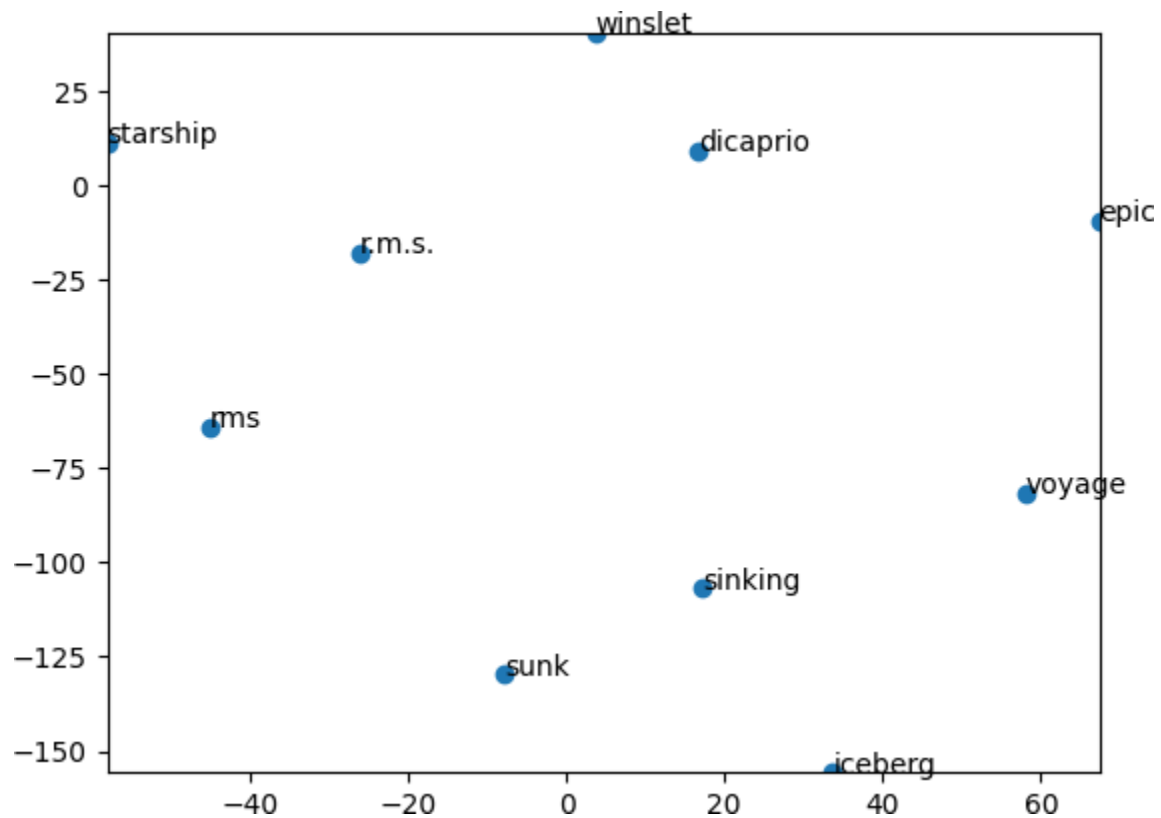




titanic



vs GLoVE 100d



[ files link ]

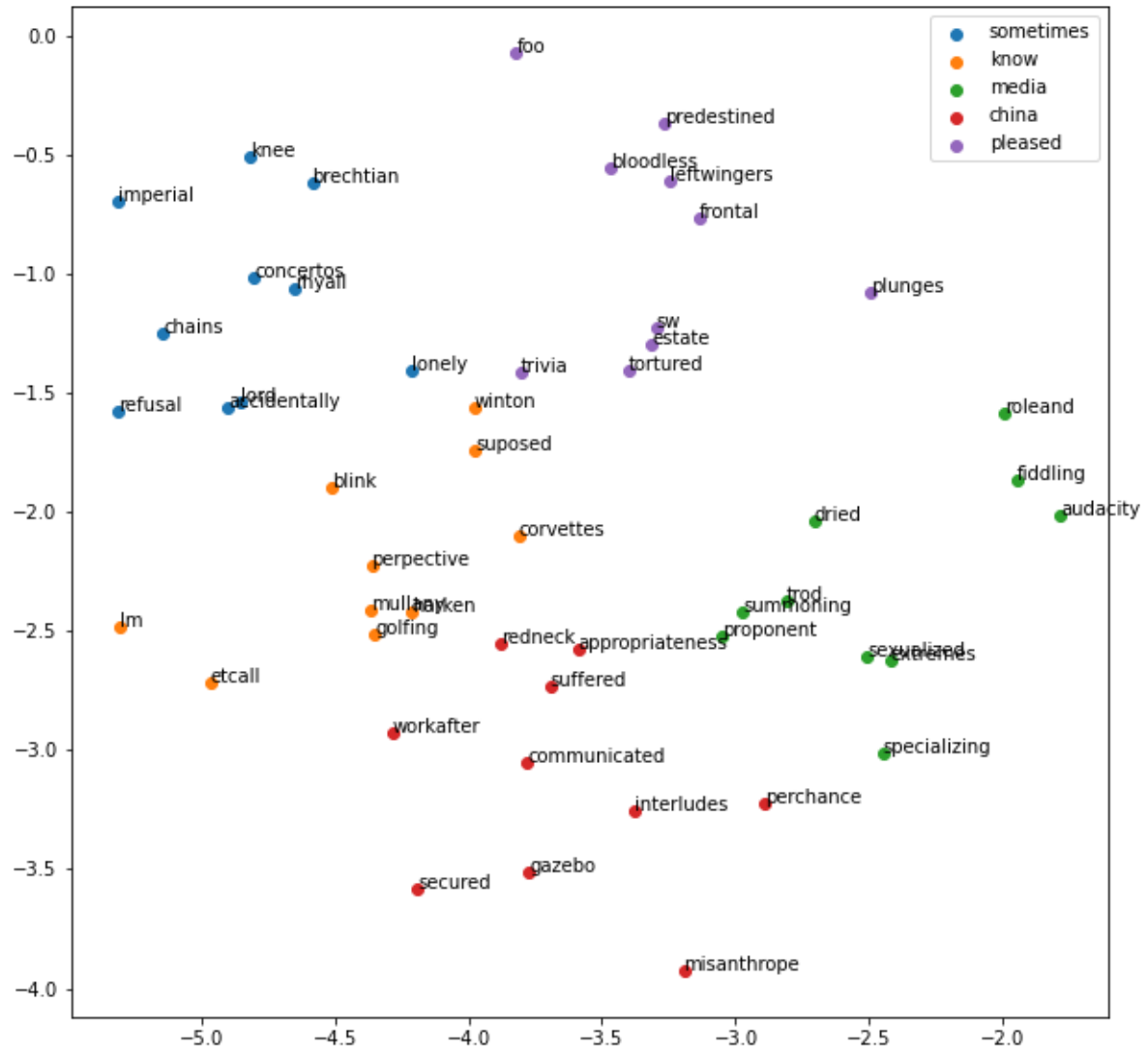
[https://drive.google.com/drive/folders/10jtghCxRPDL6vq5wr\\_pYwMBox\\_FKnAZm?usp=sharing](https://drive.google.com/drive/folders/10jtghCxRPDL6vq5wr_pYwMBox_FKnAZm?usp=sharing)

## CBOW

### ▼ Question 1 & 2

Against 5 different words:





Our Model:

