

Report

Intro to NLP Assignment 2

Aryan Jain

2019101056

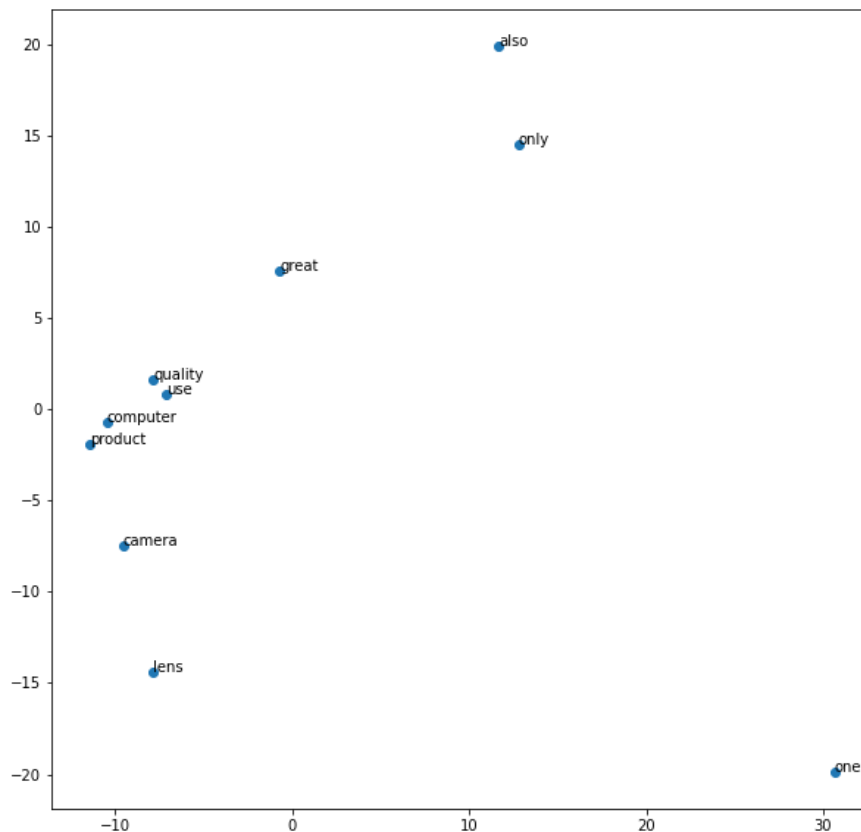
2.1) Word2Vec is a neural network-based model that is popularly used in natural language processing applications. When the size of training data increases excessively, the word2vec models face the issues. To address this issue, there is an approach called **negative sampling** used with word2vec models which allows only a small percentage of network weights to get modified during training.

When passing input for word2vec, we have 'network's label' or 'correct output' for target-context pairs as 1. In negative sampling, we pick a small number of 'negative' words at random to update the weights. A 'negative' term is one for which we want the network to output a 0 in this context. We will also continue to update the weights for our 'positive' term. As a result, just the weights relating to them will be updated, and the loss will only be propagated back for them.

Plots for the 2.2.1

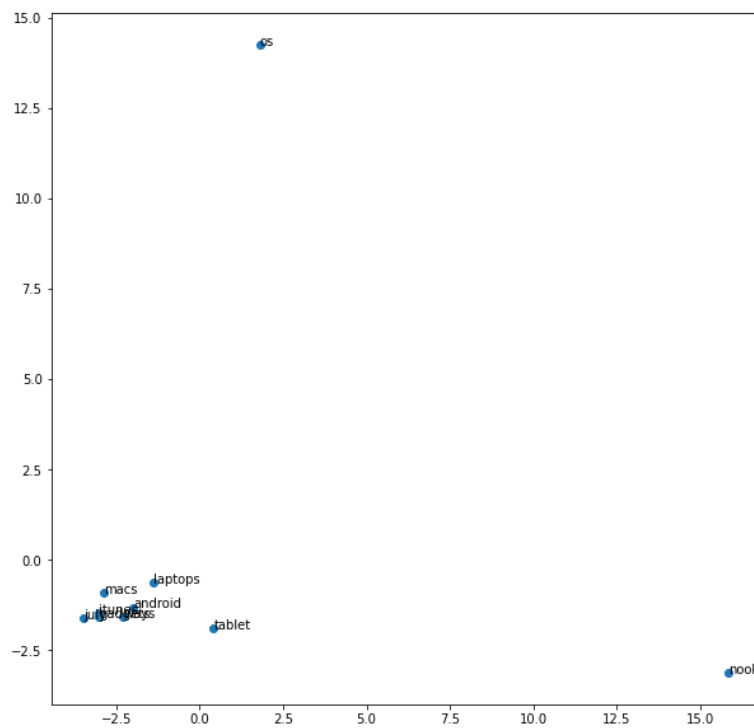
Top 10 closest word for the word camera are:

```
['camera', 'lens', 'one', 'also', 'quality', 'use', 'computer', 'product', 'great', 'only']
```



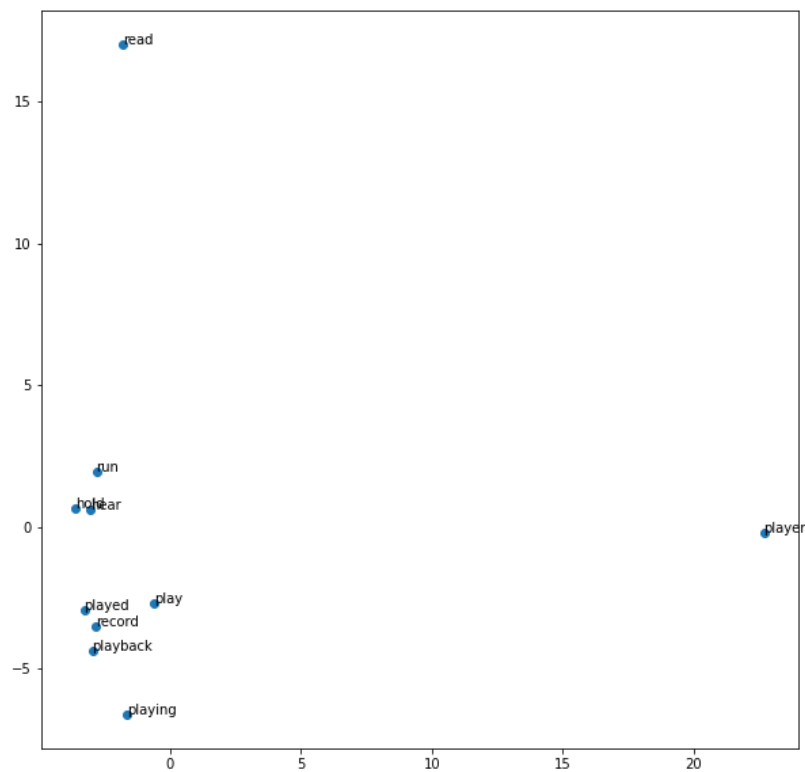
Top 10 closest word for the **noun** android are:

['android', 'tablet', 'laptops', 'macs', 'gadgets', 'itunes', 'nook', 'july', 'ways', 'os']



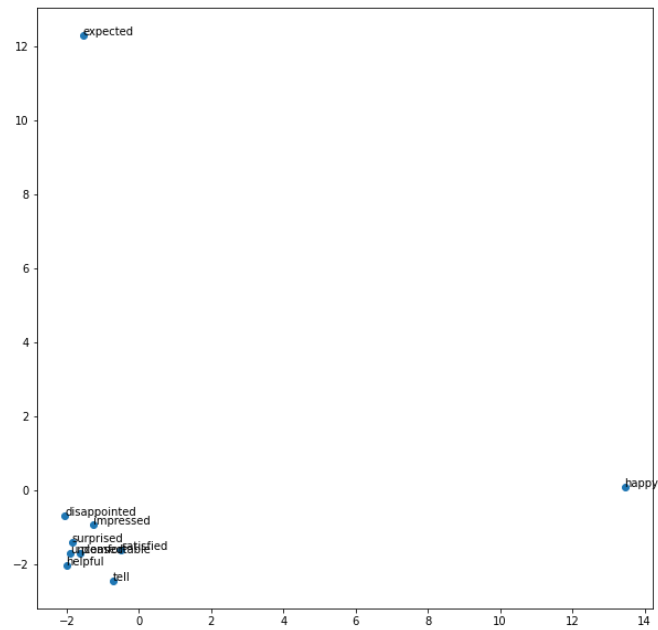
Top 10 closest word for the **verb** play are:

['play', 'playing', 'read', 'record', 'run', 'hold', 'played', 'playback', 'hear', 'player']



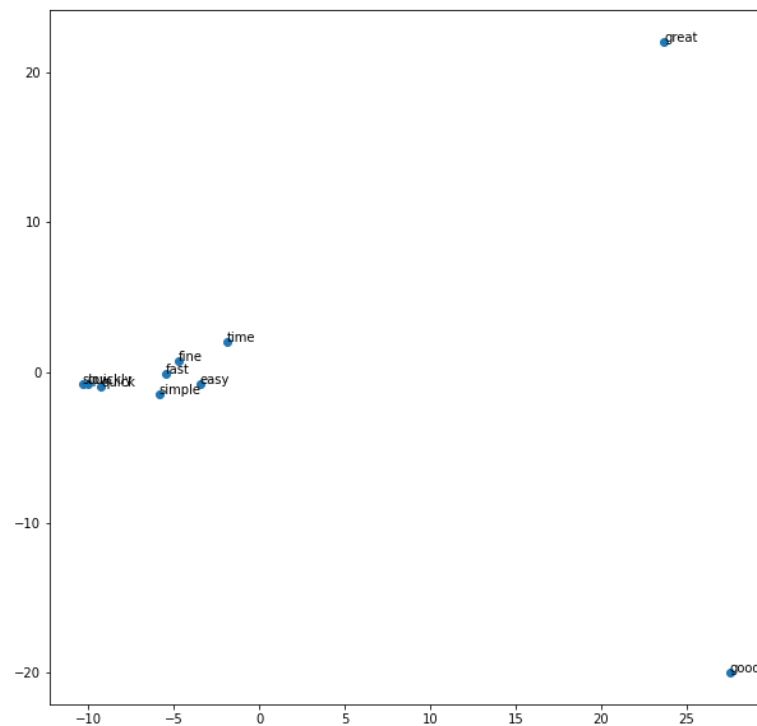
Top 10 closest word for the **adjective** happy are:

['happy', 'satisfied', 'impressed', 'pleased', 'disappointed', 'surprised', 'tell', 'expected', 'uncomfortable', 'helpful']



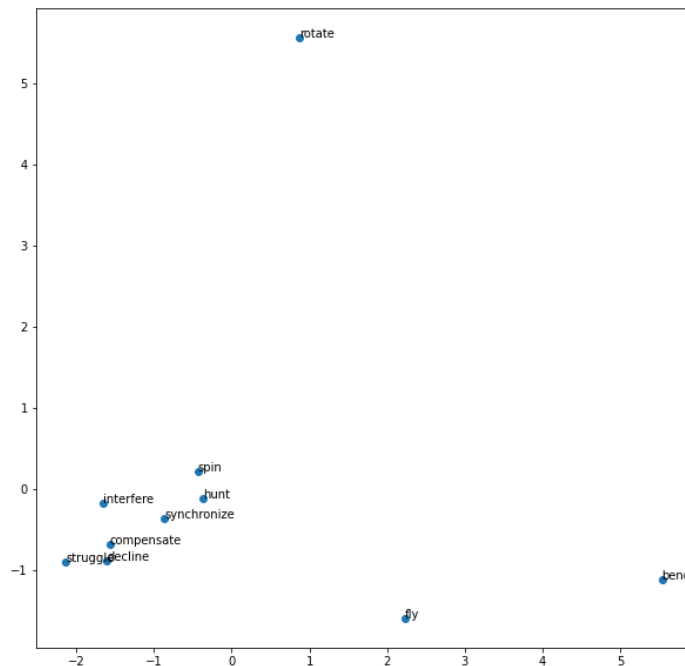
Top 10 closest word for the **adjective** quick are:

['quick', 'fast', 'quickly', 'easy', 'simple', 'time', 'slow', 'fine', 'great', 'good']



Top 10 closest word for the **verb** hunt are:

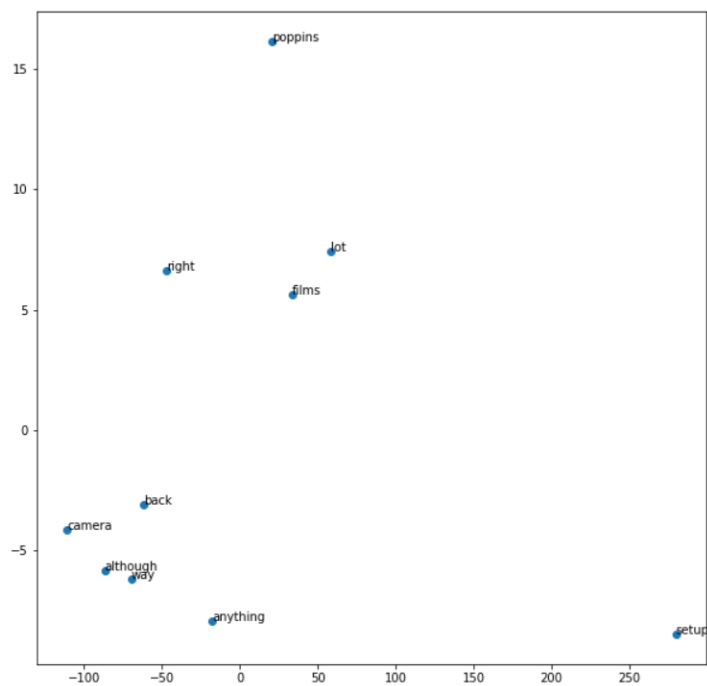
['hunt', 'spin', 'rotate', 'compensate', 'struggle', 'fly', 'decline', 'interfere', 'bend', 'synchronize']



Plots for the 2.2.2

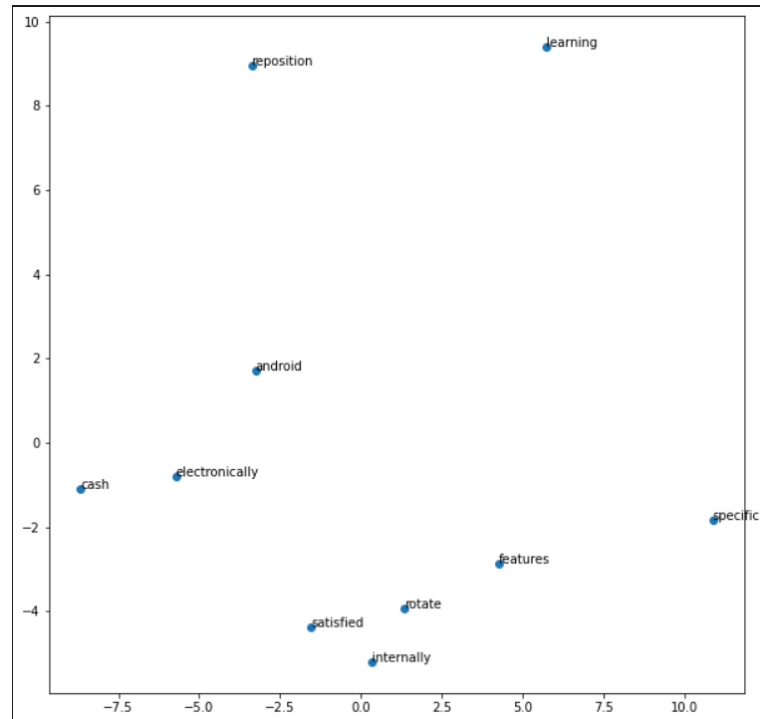
Top 10 closest words for 'camera' are:

'camera', 'although', 'right', 'lot', 'back', 'poppins', 'setup', 'films', 'way', 'anything'



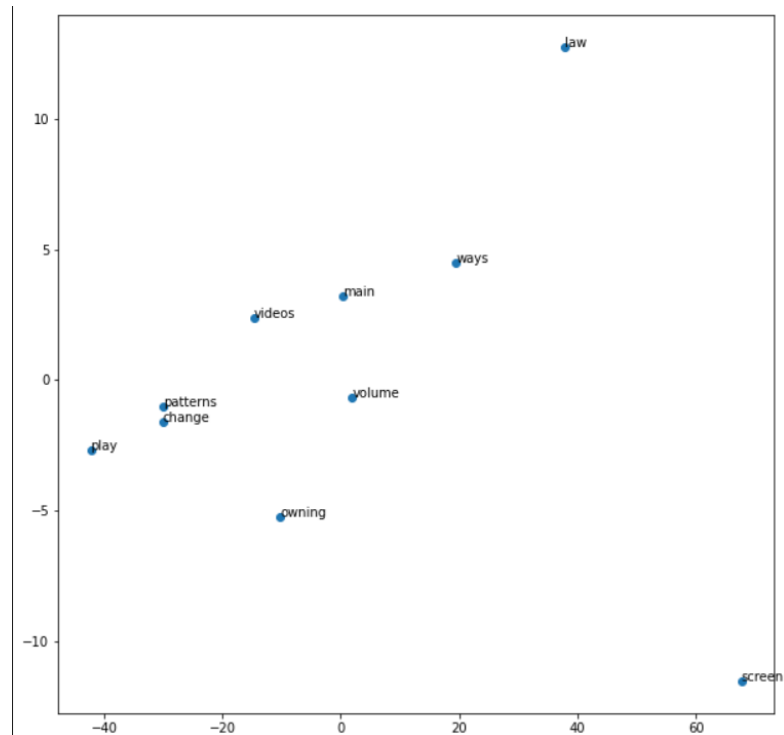
Top 10 closest words for **noun** 'android' are:

'android', 'features', 'satisfied', 'cash', 'internally', 'rotate', 'electronically', 'specific', 'reposition', 'learning'



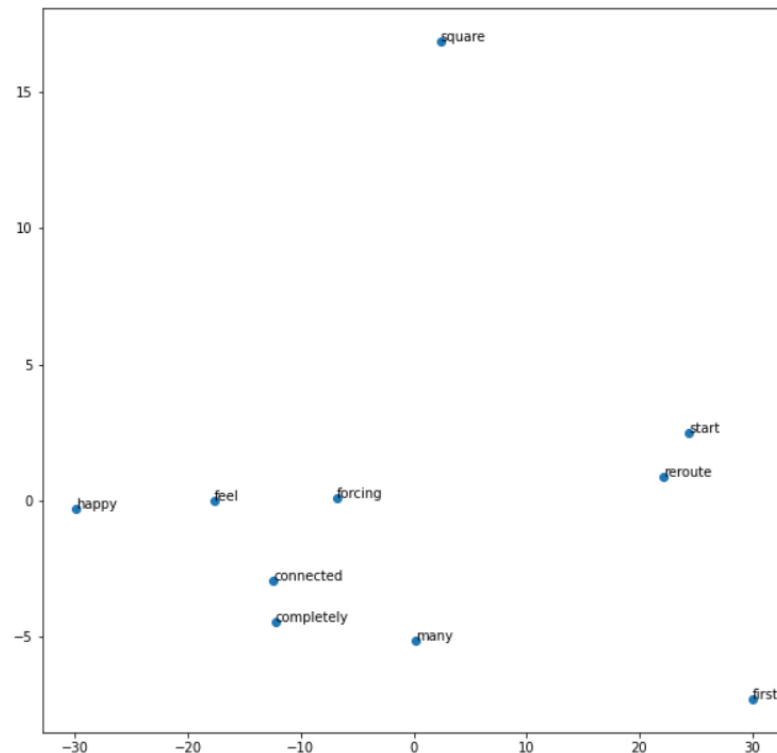
Top 10 closest words for **verb** 'play' are:

'play', 'patterns', 'main', 'owning', 'law', 'screen', 'change', 'volume', 'videos', 'ways'



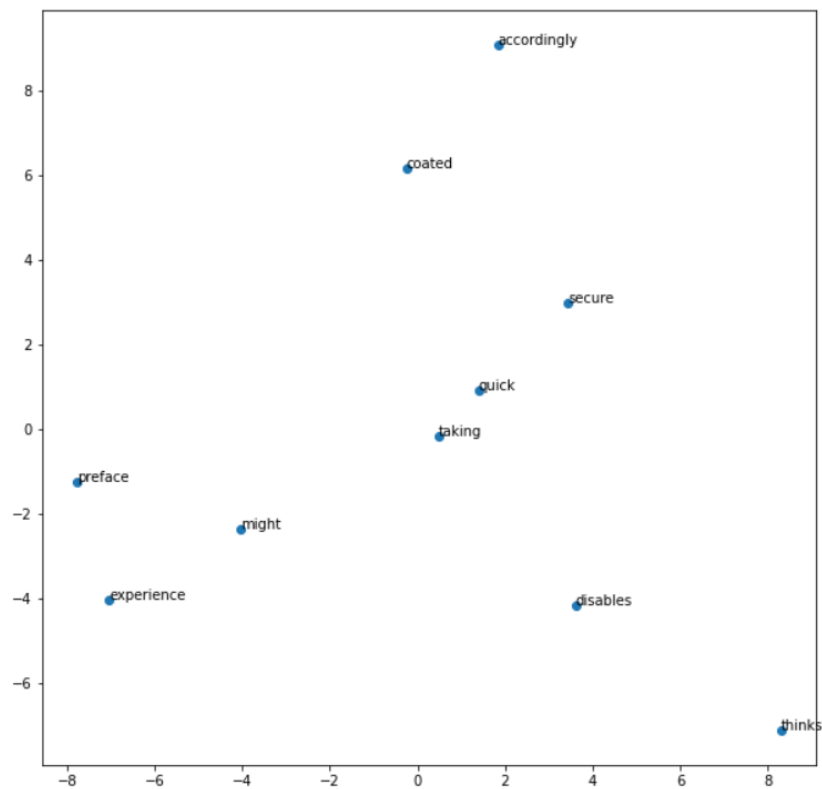
Top 10 closest words for **adjective** 'happy' are:

'happy', 'forcing', 'connected', 'reroute', 'feel', 'start', 'first', 'completely', 'square', 'many'



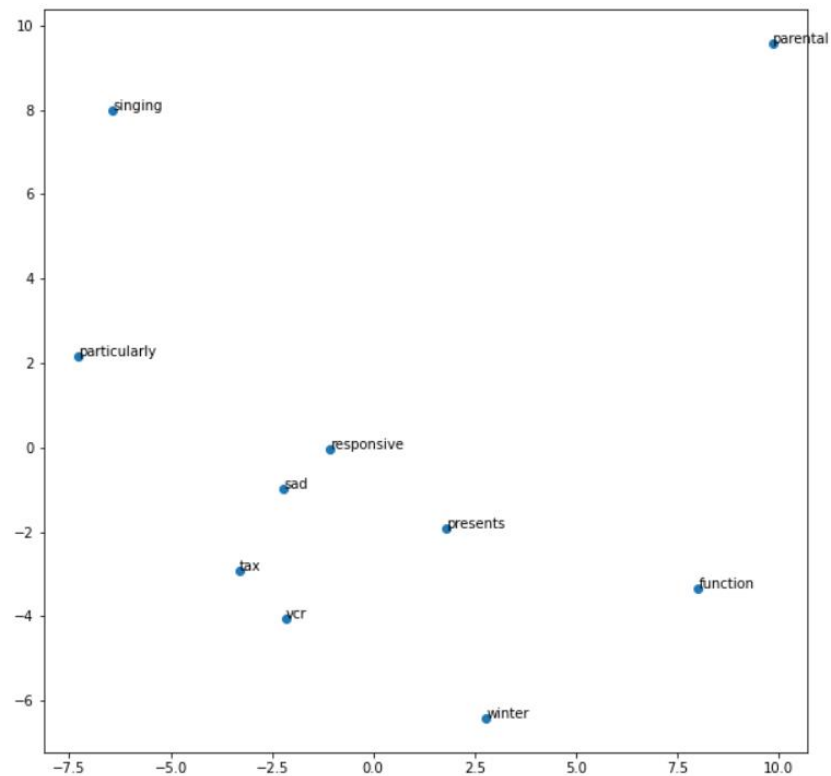
Top 10 closest words for the **adjective** 'quick' are:

'quick', 'coated', 'accordingly', 'secure', 'might', 'preface', 'experience', 'thinks', 'disables', 'taking'



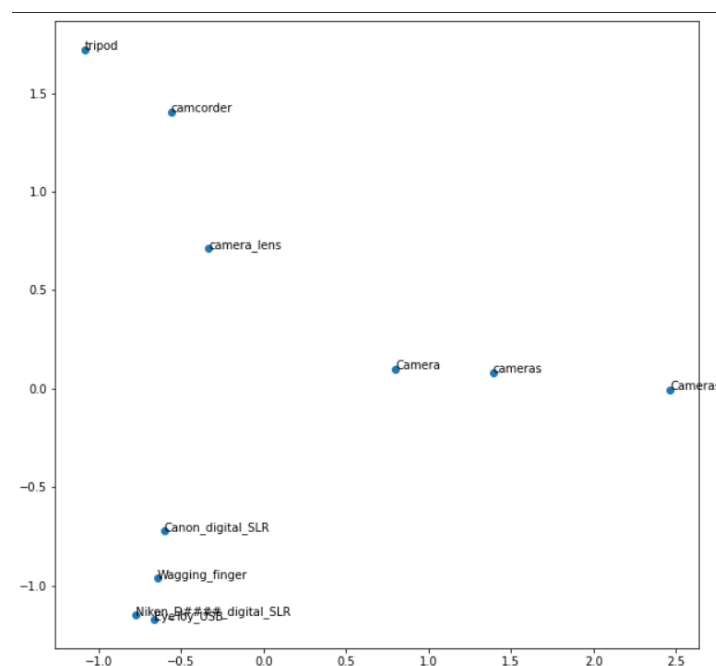
Top 10 closest words for the **verb** 'sad' are:

'sad', 'tax', 'winter', 'particularly', 'function', 'vcr', 'responsive', 'singing', 'parental', 'presents'



For the pre-trained vectors, the 10 closest words for the word 'camera' are:

'cameras', 'Wagging_finger', 'camera_lens', 'camcorder', 'Camera', 'Canon digital SLR', 'Cameras', 'Nikon D#### digital SLR', 'tripod', 'EyeToy USB'



- ➔ The embedding for q2 in q2.pt file
- ➔ The embedding for q1 is in q1.pkl file is available at:
<https://drive.google.com/file/d/16B6FyqeAknSql6DQK68uXd4CUMWifiON/view?usp=sharing>