

# Advanced NLP Assignment-1 Report

T H Arjun, CSD, 2019111012

In all the questions for the 5 words to plot TSNE for I chose the words

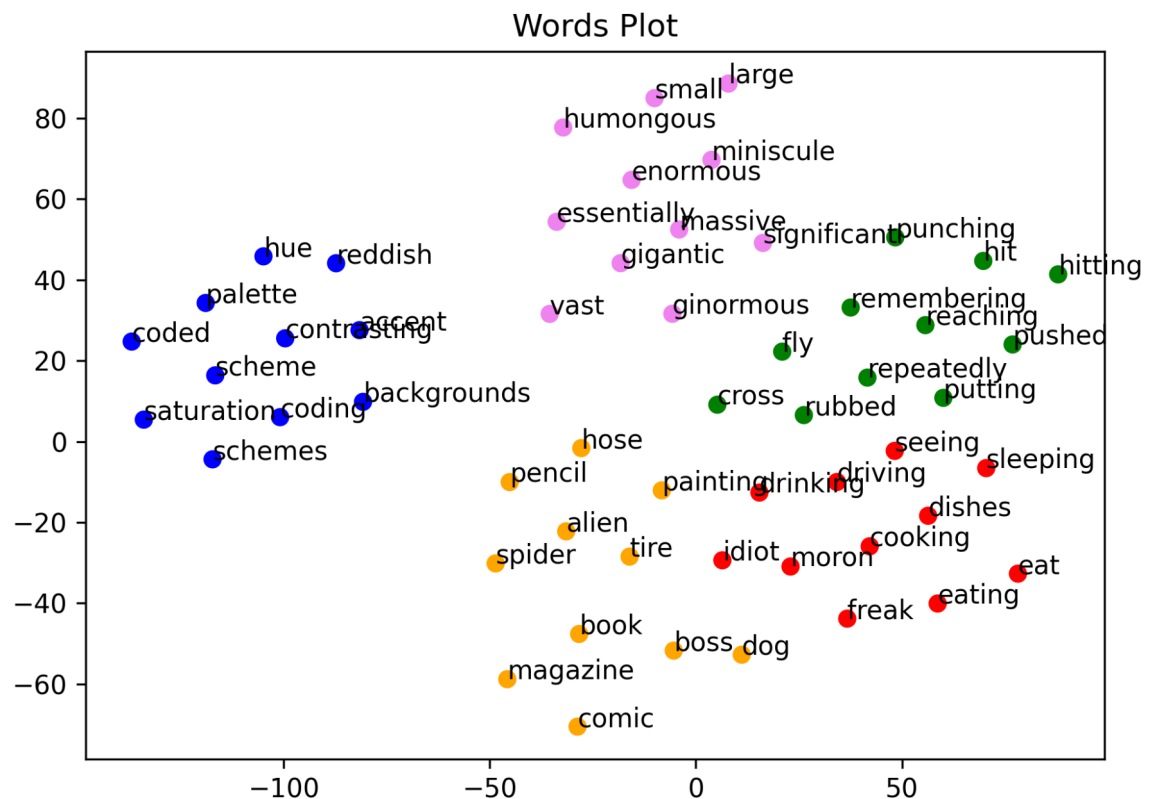
- eating
- punching
- coding
- enormous
- book

They have colours red, green, blue, violet, orange respectively in the TSNE Plots.

- **We notice that they form clusters since their word representations are nearer in the vector space to the words they are similar to. Hence we see colored clusters.**
- **We can also notice that the words similar to the “camera” are also very accurate in both SVD and CBOW that I trained and is very comparable to the pre-trained-language model from Gensim.**

## SVD

### 1. Top-10 word vectors for 5 different words



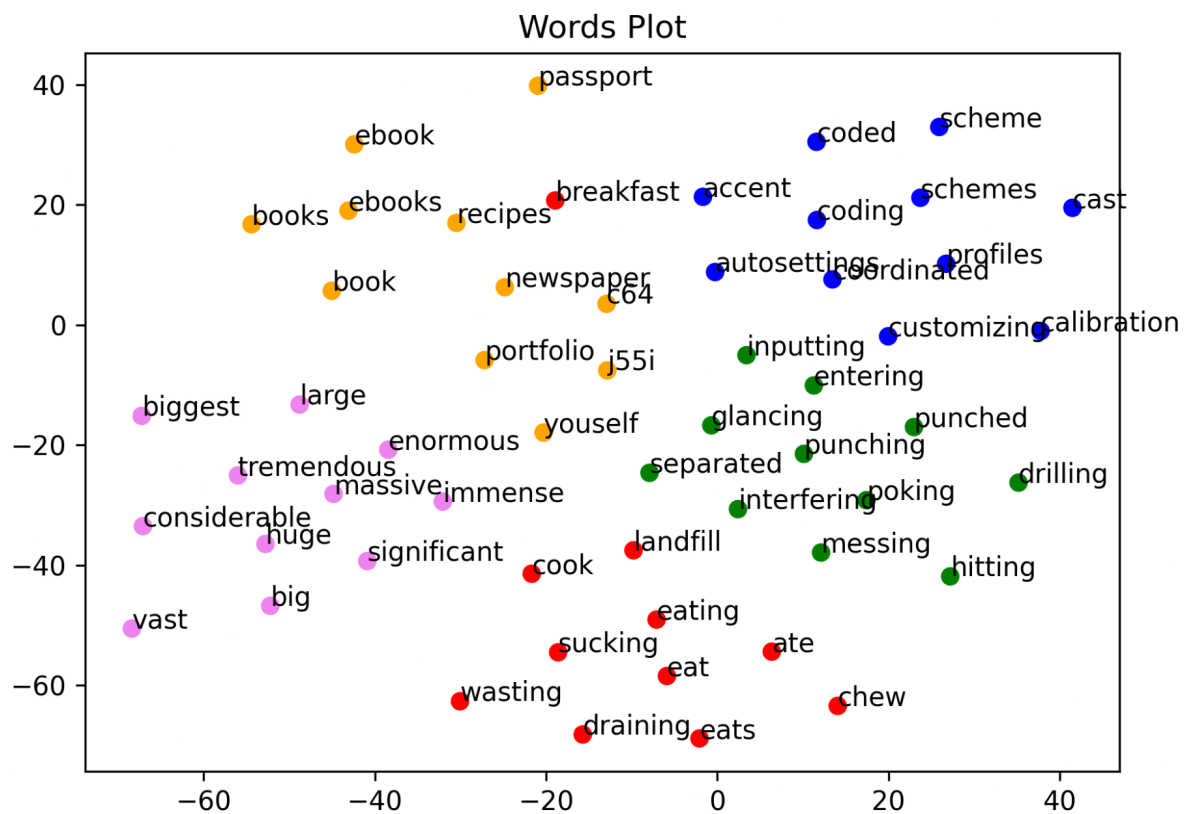
## 2. The top 10 closest words for the word 'camera'

```
[6]: sorted_words[:10]
```

```
[6]: ['camer',  
      'cam',  
      'camcorder',  
      'scope',  
      'camara',  
      'microscope',  
      'telescope',  
      'monopod',  
      'binocular',  
      'board']
```

# CBOW

## 1. Top-10 word vectors for 5 different words



## 2. The top 10 closest words for the word 'camera'

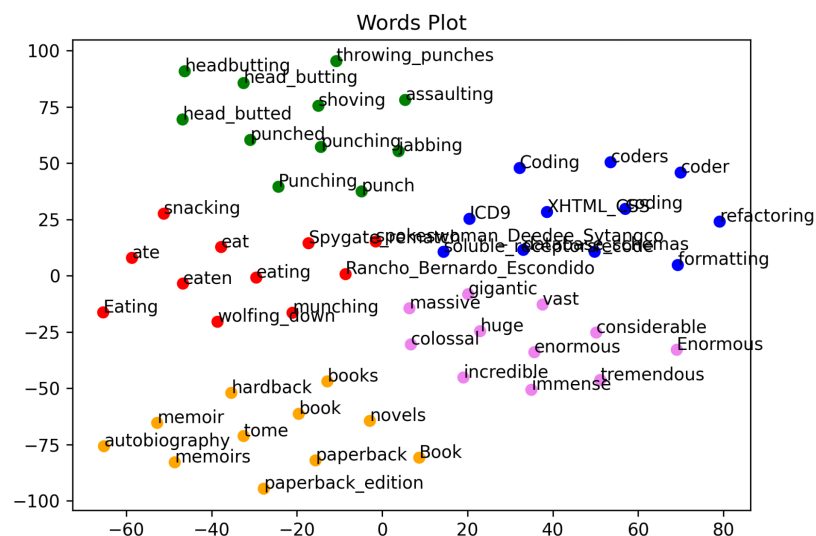
```
[6]: sorted_words[:10]
```

```
[6]: ['cameras',  
      'cam',  
      'dslr',  
      'camcorder',  
      'dlsr',  
      'cams',  
      'slr',  
      'd90',  
      'camerasand',  
      'dslrs']
```

# Gensim

The model I used for this was the google-news gensim word2vec model

## 1. Top-10 word vectors for 5 different words



## 2. The top 10 closest words for the word 'camera'

```
[3]: model.most_similar('camera', topn=10)
```

```
[3]: [('cameras', 0.8131938576698303),  
      ('Wagging_finger', 0.7311819791793823),  
      ('camera_lens', 0.7250816226005554),  
      ('camcorder', 0.7037475109100342),  
      ('Camera', 0.6848660111427307),  
      ('Canon_digital_SLR', 0.6474252343177795),  
      ('Cameras', 0.6350969672203064),  
      ('Nikon_D###_digital_SLR', 0.6259366273880005),  
      ('tripod', 0.6189838647842407),  
      ('EyeToy_USB', 0.6173486709594727)]
```