Report

Arvindh A, 201911010

Remarks:

- Similar word pairs have high cosine similarity and low euclidean distance
- Contextual word embeddings for homonyms are observed to be more similar with related words to one another

Sentences used

```
test_corpus.append("I love the city atmosphere")
test_words.append("city")
test_corpus.append("This country has the largest lake")
test_words.append("country")

test_corpus.append("This phone is brand new")
test_words.append("phone")
test_corpus.append("His laptop has the best GPU")
test_words.append("laptop")

test_corpus.append("He is a good person")
test_words.append("good")
test_corpus.append("I love the bad boys movie")
test_words.append("bad")

test_corpus.append("India is one of the most diverse countries")
test_words.append("India")
test_corpus.append("Switzerland is a cold country")
test_words.append("Switzerland")

test_corpus.append("I use Manjaro, a linux distribution")
test_words.append("linux")
test_corpus.append("Microsoft's windows 10 OS is pretty good")
test_words.append("windows")
```

5 similarity pairs

```
sim(e, w_to_i, "city", "country")
```

```
(0.67066824, 0.4780803)
```

```
sim(e, w_to_i, "phone", "laptop")
```

```
(0.64256305, 0.45688364)
```

```
sim(e, w_to_i, "good", "bad")
```

```
(0.69522154, 0.43069264)
```

```
sim(e, w_to_i, "India", "Switzerland")
```

```
(0.7370452, 0.40488723)
```

```
sim(e, w_to_i, "linux", "windows")
```

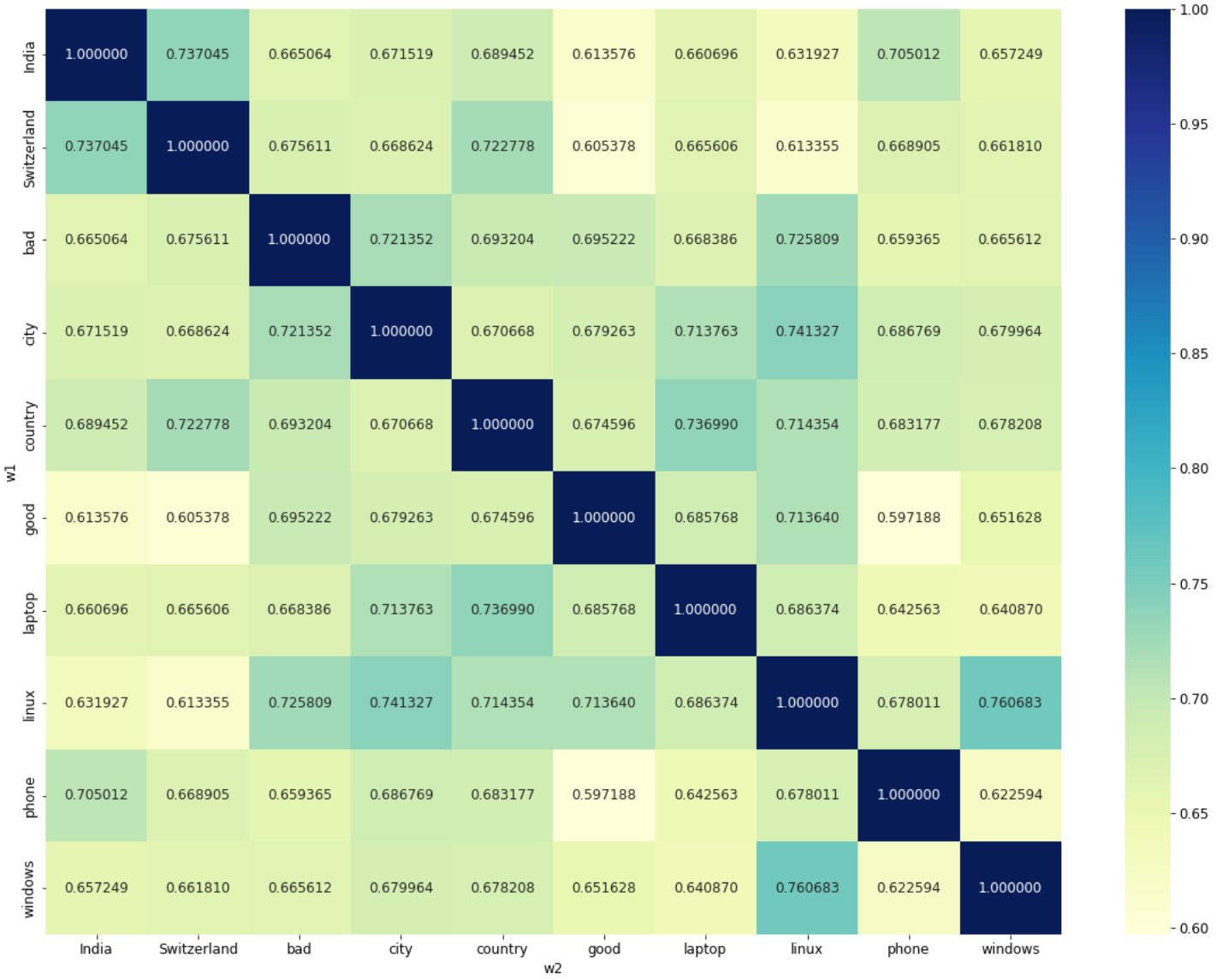
```
(0.76068336, 0.39930138)
```

DataFrame for heatmap plot (10 * 10 entries)

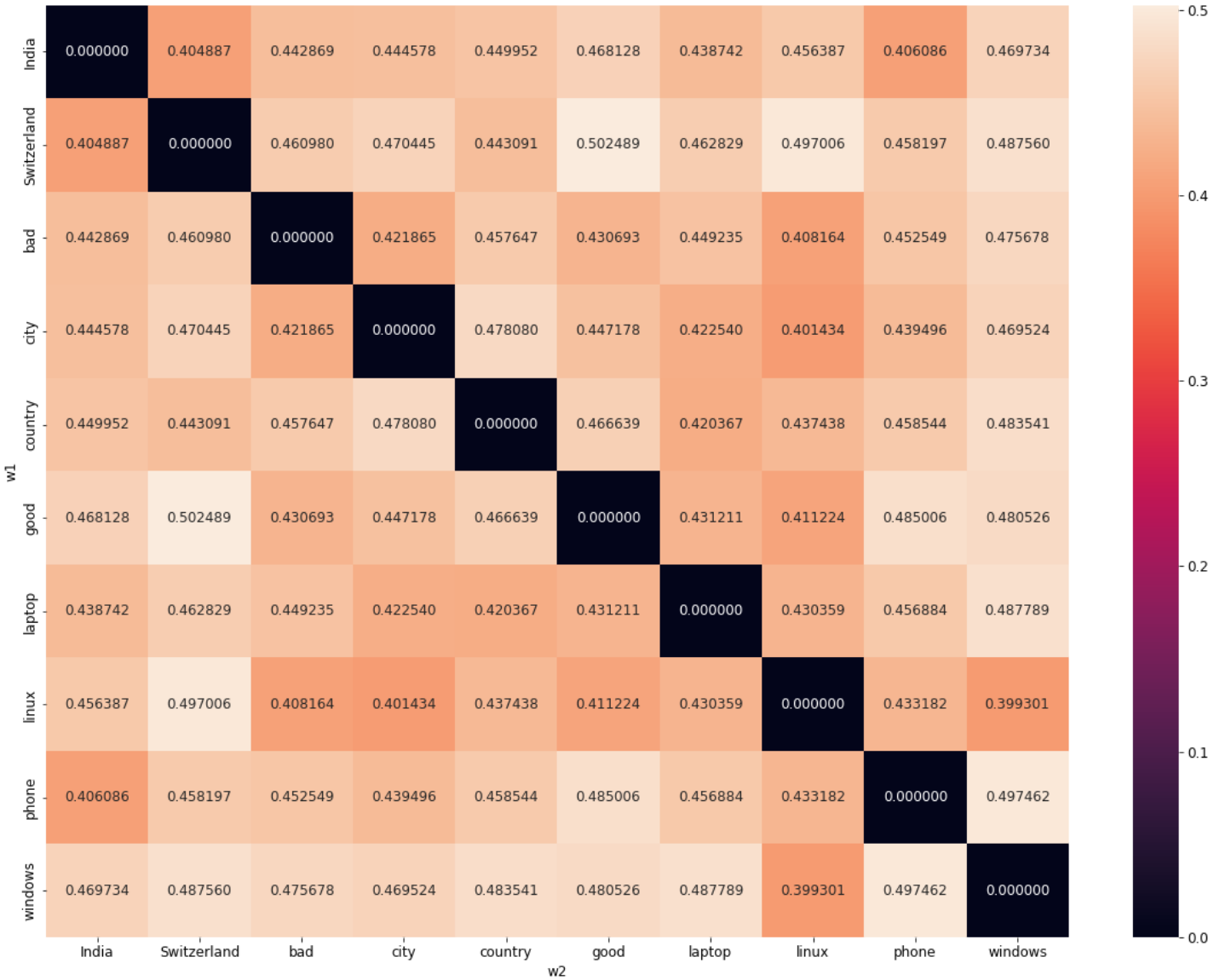
	w1	w2	cos_sim	euc
0	city	city	1.000000	0.000000
1	city	country	0.670668	0.478080
2	city	phone	0.686769	0.439496
3	city	laptop	0.713763	0.422540
4	city	good	0.679263	0.447178
...	...	...	...	...
95	windows	bad	0.665612	0.475678
96	windows	India	0.657249	0.469734
97	windows	Switzerland	0.661810	0.487560
98	windows	linux	0.760683	0.399301
99	windows	windows	1.000000	0.000000

100 rows × 4 columns

Cosine Similarity heatmap



Euclidean Similarity heatmap



Contextual similarity

```
w1 = gen_embeddings("Please bear with me throughout this talk", "bear")
w2 = gen_embeddings("Grizzly bear is dangerous", "bear")
disc = gen_embeddings("This zoo has animals from all over the world", "animals")
```

```
cos_sim(w1, disc)
```

```
0.7366396
```

```
cos_sim(w2, disc)
```

```
0.72614187
```