Report

Arvindh A, 201911010

Remarks:

- Written from scratch so not really accurate
- Contextual meanings can be noticed

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)

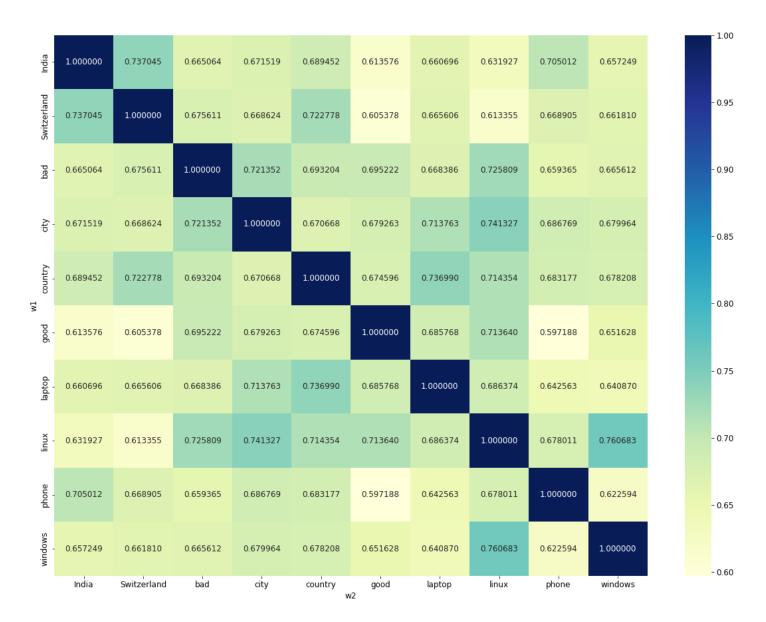
sim(e, w_to_i, "linux", "windows")

(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

lndia -	0.000000	0.404887	0.442869	0.444578	0.449952	0.468128	0.438742	0.456387	0.406086	0.469734	- 0.5
Switzerland	0.404887	0.000000	0.460980	0.470445	0.443091	0.502489	0.462829	0.497006	0.458197	0.487560	
paq	0.442869	0.460980	0.000000	0.421865	0.457647	0.430693	0.449235	0.408164	0.452549	0.475678	- 0.4
aity	0.444578	0.470445	0.421865	0.000000	0.478080	0.447178	0.422540	0.401434	0.439496	0.469524	
1 country	0.449952	0.443091	0.457647	0.478080	0.000000	0.466639	0.420367	0.437438	0.458544	0.483541	- 0.3
wl good	0.468128	0.502489	0.430693	0.447178	0.466639	0.000000	0.431211	0.411224	0.485006	0.480526	
laptop	0.438742	0.462829	0.449235	0.422540	0.420367	0.431211	0.000000	0.430359	0.456884	0.487789	- 0.2
Inux	0.456387	0.497006	0.408164	0.401434	0.437438	0.411224	0.430359	0.000000	0.433182	0.399301	
phone	0.406086	0.458197	0.452549	0.439496	0.458544	0.485006	0.456884	0.433182	0.000000	0.497462	- 0.1
windows	0.469734	0.487560	0.475678	0.469524	0.483541	0.480526	0.487789	0.399301	0.497462	0.000000	
	India	Switzerland	bad	aity	country	good 12	laptop	linux	phone	windows	- 0.0

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
```