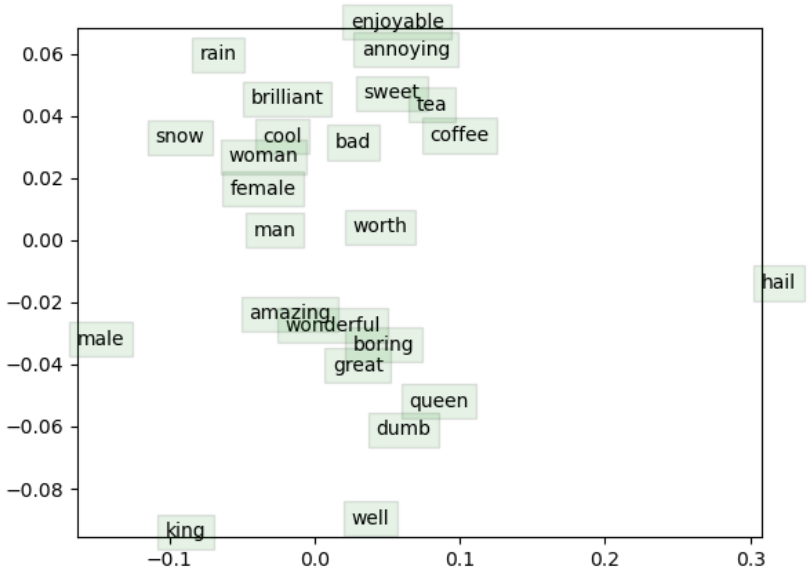
Allison Aprile
Assignment 2: Word Vectors (Report)
March 15, 2021

Cost at Convergence and Training Time

```
iter 39940: 9.611494
iter 39950: 9.629689
iter 39960: 9.717683
iter 39970: 9.776979
iter 39980: 9.813174
iter 39990: 9.854022
iter 40000: 9.812206
sanity check: cost at convergence should be around or below 10
training took 2892 seconds
```

Word Vector Plot



Although they are not very concentrated, the plot above does seem to have two clusters, as well as several outliers. One cluster seems to be around the word ‘bad’, whose closest neighbors are a little surprising to me – for example, ‘sweet’, ‘brilliant’, and ‘cool’ are not typically associated with negativity. The other cluster seems to be around ‘boring’, which also has surprising neighbors of ‘amazing’, ‘wonderful’, and ‘great’. These neighbors seem to be related to each other, but otherwise are not usually associated with ‘boring’. A few other trends that stand out are the closest of synonyms or related topics (for example, ‘woman’ and ‘female’, or ‘tea’ and ‘coffee’). Overall, this graph seems to indicate a lot of juxtaposition in the training data.

(KNN results on the next page)

Allison Aprile
Assignment 2: Word Vectors (Report)
March 15, 2021

KNN Examples

```
KNN Results
---------------
EXAMPLE:  1
        Word:  great
        Nearest Neighbors:  ['academy', 'matured', 'bolster']

EXAMPLE:  2
        Word:  cool
        Nearest Neighbors:  ['hype', 'schmaltzy', 'atypically']

EXAMPLE:  3
        Word:  brilliant
        Nearest Neighbors:  ['championship', 'stephen', 'iconoclastic']

EXAMPLE:  4
        Word:  wonderful
        Nearest Neighbors:  ['tangents', 'manage', 'certain']

EXAMPLE:  5
        Word:  well
        Nearest Neighbors:  ['mistaken', 'weathered', 'enforcement']

EXAMPLE:  6
        Word:  amazing
        Nearest Neighbors:  ['message', 'ingratiating', 'chiaroscuro']

EXAMPLE:  7
        Word:  worth
        Nearest Neighbors:  ['portrayal', 'coast', 'ethereal']

EXAMPLE:  8
        Word:  sweet
        Nearest Neighbors:  ['torture', 'fleetingly', 'helps']

EXAMPLE:  9
        Word:  enjoyable
        Nearest Neighbors:  ['options', 'matter', 'bloodless']

EXAMPLE:  10
        Word:  boring
        Nearest Neighbors:  ['road', 'irreverence', 'retelling']

(base) aaprile@MacBook-Pro a2 %
```

For the KNN tests, I decided to pick the words from the defined ‘visualizeWords’ list in run.py; this way, I could see whether any of the neighbors appeared on the plot. I believe if I had selected a larger k (I chose k = 3), this definitely would have been the result. Nonetheless, the above report shows some reasonable results. For example, ‘great’s’ neighbors are words typically associated with high standards or strength. Also, ‘boring’s’ neighbors are words related to repetition. There are also some odd results, especially for ‘sweet’ and ‘enjoyable’. These results are attributed to the context words for each word embedding.
In future tests, I would maybe experiment with different similarity measures and a higher number of k for analysis.