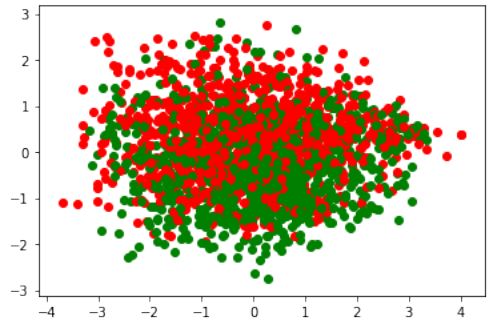
1. The closest sentences are "a solidly entertaining little film" and "a modestly surprising movie" The minimum distance is 1.2074438333511353.
2. [CLS] token is originally trained to be used for sentence sentiment classification. We expect this because sentiment is a very important feature for classification tasks.
3. The plot is shown below. The red plots are negative, and the green plots are positive.



4. According to the evaluate function, the accuracy for Naive Bayes learner is 0.798, for Adaboost learner is 0.786, for k-nearest neighbors is 0.732, for classic multilayer neural network is 0.85, for 2 hidden layers neural network is 0.832, for logistic is 0.854.
5. K-nearest neighbor does not work well because there is a lot overlap between negative and positive sentence vectors which can be observed in the plots by PCA. Also, the plots are close to each other, and the neighbors do not have the same labels stably for the most plots. Thus, it is hard for K-nearest neighbor to work well.
6. Because this data is mostly linear separable. A single perceptron works well when the data is linear separable. However, a more complex neural network could be overfitting in this case.
7. The first two components of PCA are the features with maximum covariance but the feature may not contain the most information and reduce the entropy most. However, we need the features with the most information for the decision tree. Also, the feature selected might be synthetic feature. Thus, one of the first two components of PCA would be likely to show up as a weak learner in Adaboost.
8. The sentence is: the most repugnant adaptation of a classic text since roland joff and demi moore 's the scarlet letter.

```
: print(df[0][11])
```

```
the most repugnant adaptation of a classic text since roland joff and demi moore 's the scarlet letter
```

9. The abbreviation returned:

```
[('the', 'DT'), ('most', 'RBS'), ('repugnant', 'JJ'), ('adaptation', 'NN
'), ('of', 'IN'), ('a', 'DT'), ('classic', 'JJ'), ('text', 'NN'), ('since
', 'IN'), ('roland', 'NN'), ('joff', 'NN'), ('and', 'CC'), ('demi', 'NN'),
 ('moore', 'NN'), ("'s", 'POS'), ('the', 'DT'), ('scarlet', 'NN'), ('lette
r', 'NN')]
```

Meaning for “most” in this sentence: forming the superlative of adjectives and adverbs, especially those of more than one syllable.
Meaning for “since” in this sentence: in the intervening period between (the time mentioned) and the time under consideration, typically the present.
10. The author adds “bias” to get some nonlinear interaction between features.
After we fix capitalization, the successfully tagger found NNP:

```
[('The', 'DT'), ('most', 'RBS'), ('repugnant', 'JJ'), ('adaptation', 'NN
'), ('of', 'IN'), ('a', 'DT'), ('classic', 'JJ'), ('text', 'NN'), ('since
', 'IN'), ('Roland', 'NNP'), ('Joff', 'NNP'), ('and', 'CC'), ('Demi', 'NNP
```

```
'), ('Moore', 'NNP'), ("'s", 'POS'), ('the', 'DT'), ('scarlet', 'NN'), ('l
etter', 'NN')]
```

11. The result is shown below:

```
: #nltk.download('maxent_ne_chunker')
  #nltk.download('words')
  entities = nltk.chunk.ne_chunk(tagged)
  print(entities)
```

```
(S
  the/DT
  most/RBS
  repugnant/JJ
  adaptation/NN
  of/IN
  a/DT
  classic/JJ
  text/NN
  since/IN
  roland/NN
  joff/NN
  and/CC
  demi/NN
  moore/NN
  's/POS
  the/DT
  scarlet/NN
  letter/NN)
```

Thus, all proper nouns are thought as “NN”, noun.

12. Code.