

### Question 3.1.1

How many examples are in the file data.tsv?

There's 10000 examples

### Question 3.1.2

Select two random examples each from the positive set (subjective) and two from the negative set. For all four examples, explain, in English, why it has the given label.

Data 1628: quite simply , a joy to watch and--especially--to listen to .

This sentence is subjective as it talks about the author's personally feelings that can't be proven true or false by others.

Data 1446: just isn't as weird as it ought to be .

This sentence is subjective as it compares the weirdness of the movie with the author's expectations.

Data 8296: he has just graduated college , he has a beautiful girlfriend , jane , and a publisher has picked up his first novel .

This sentence is objective since it is only concluding the plot of the movie.

Data 8771: she has been programmed to eliminate other cybernetic organisms .

This sentence is objective since it's just stating a fact.

### Question 3.1.3

Find one example from each of the positive and negative sets that you think has the incorrect label, and explain why each is wrong

Data 200: minority report is exactly what the title indicates , a report .

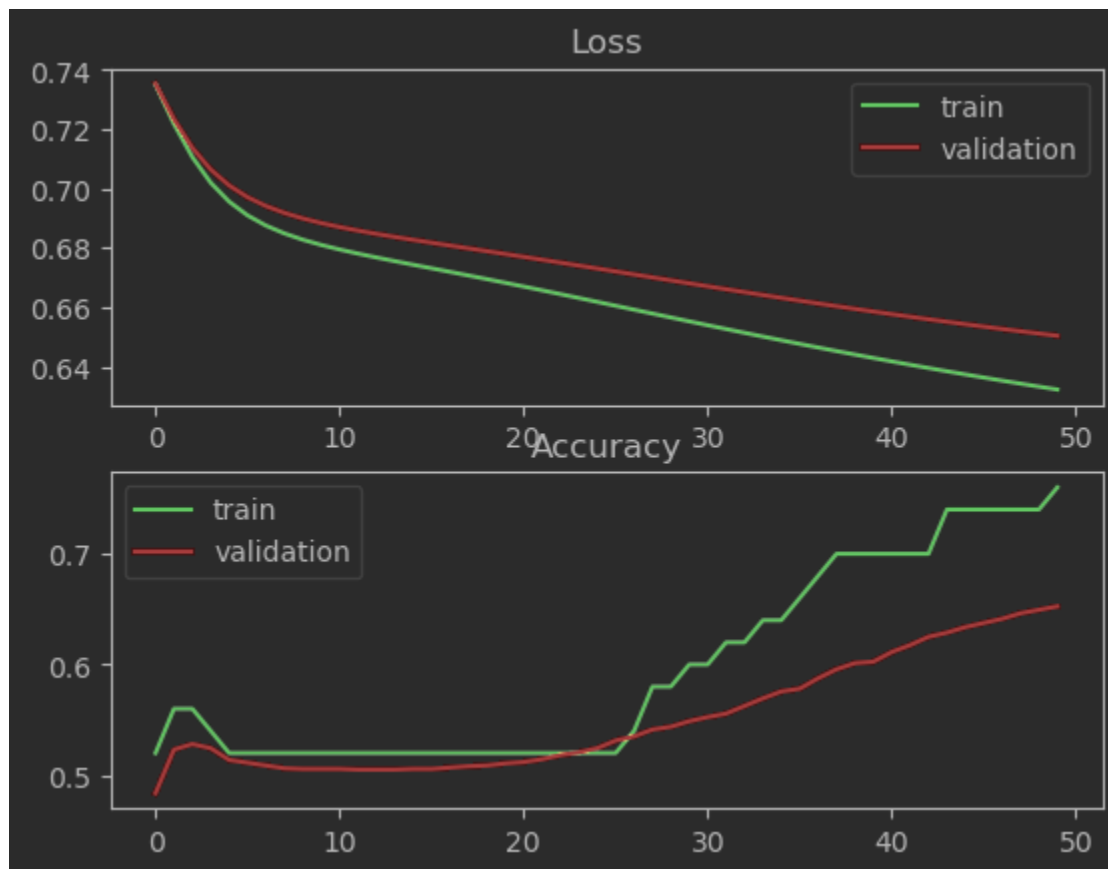
This sentence is simply stating the type of the movie with no personal bias, but is labelled as subjective.

Data 5749: through a funny , sometimes uncomfortable , and combative evening we learn what that brief romance meant to them and how they feel about the subsequent course of their lives .

This sentence is incorrectly labelled as objective though it contains some of author's personal opinion.

#### Question 4.4

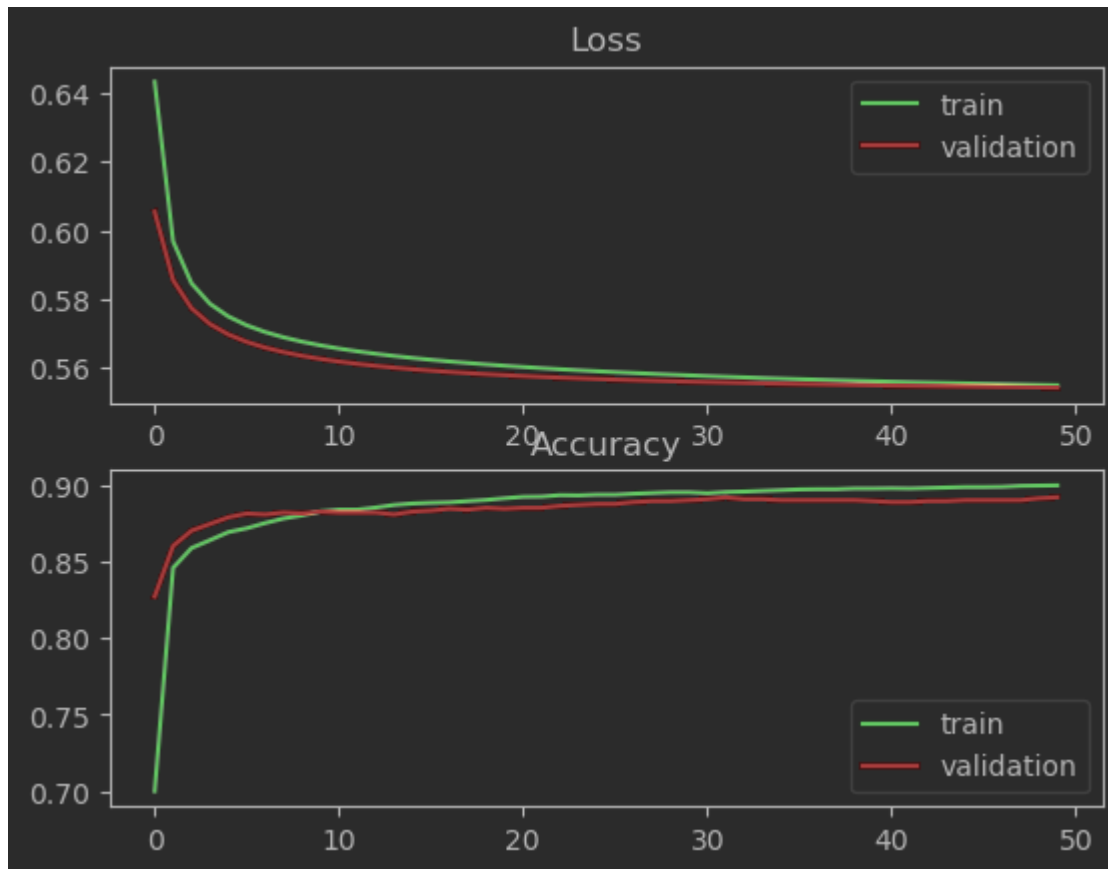
Provide the training loss and accuracy plot for the overfit data in your Report.



#### Question 4.5

Give the training and validation loss and accuracy curves vs. epoch in your report, and report the final test accuracy. Evaluate the test data and provide the accuracy result in your report.

Answer this questions: In the baseline model, what information contained in the original sentence is being ignored?



Final test accuracy is 89.15%.

The words that aren't recognized by GloVe and punctuations are ignored.

#### Question 4.6

Use that function to determine the 20 closest words to those trained parameters of the neuron. You should see some words that make it clear what the classifier is doing. Do some of the words that you generated make sense? Explain.

The top 20 words from my model are:

flattering 0.522

visuals 0.512

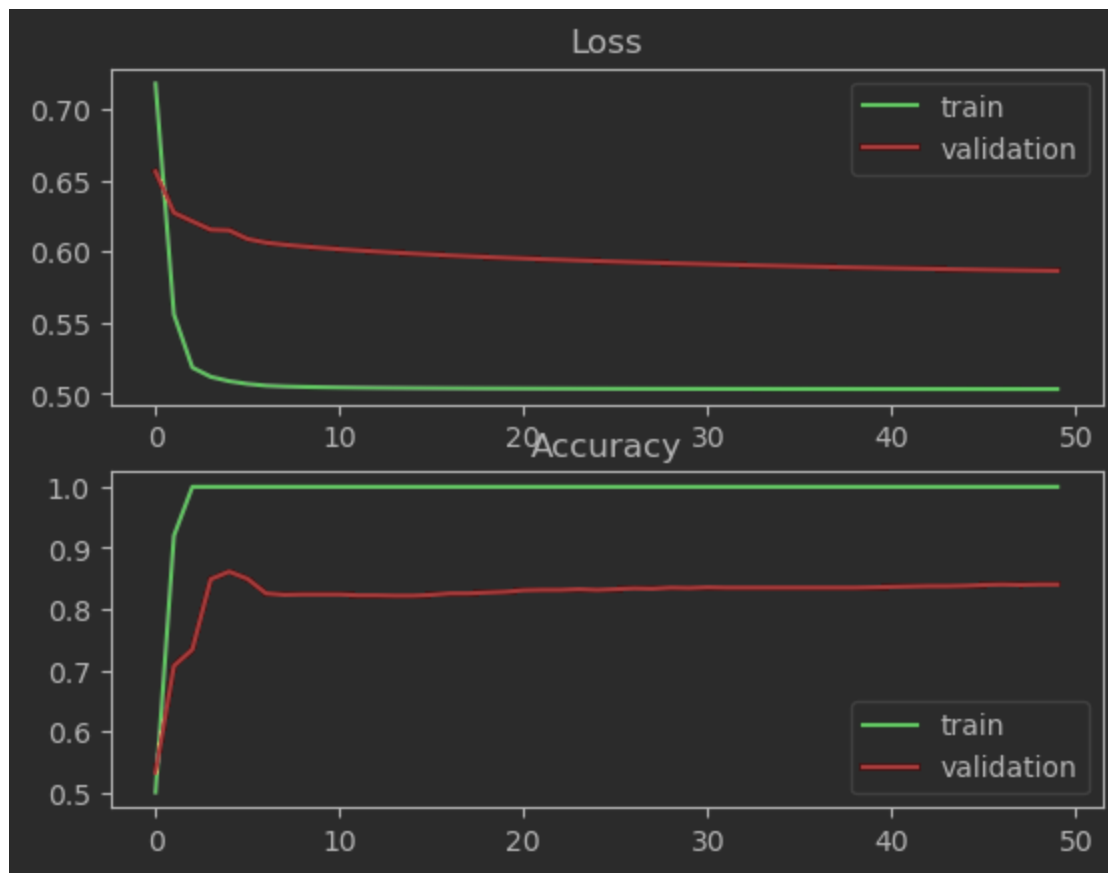
verbiage 0.511

portentous	0.492
insipid	0.489
succinct	0.487
campy	0.485
mesmerizing	0.482
disquieting	0.481
laudatory	0.480
perfunctory	0.477
nuanced	0.472
understatement	0.468
simplistic	0.468
cheesy	0.467
crispness	0.467
watchable	0.466
pizazz	0.464
riveting	0.464
reverential	0.464

Quite a few of these words are related to subjective feelings after watching movies, such as "campy", "insipid" or "disquieting". There are also words that describes the quality of a movie, like "verbiage", "watchable" or "succinct".

### Question 5.1

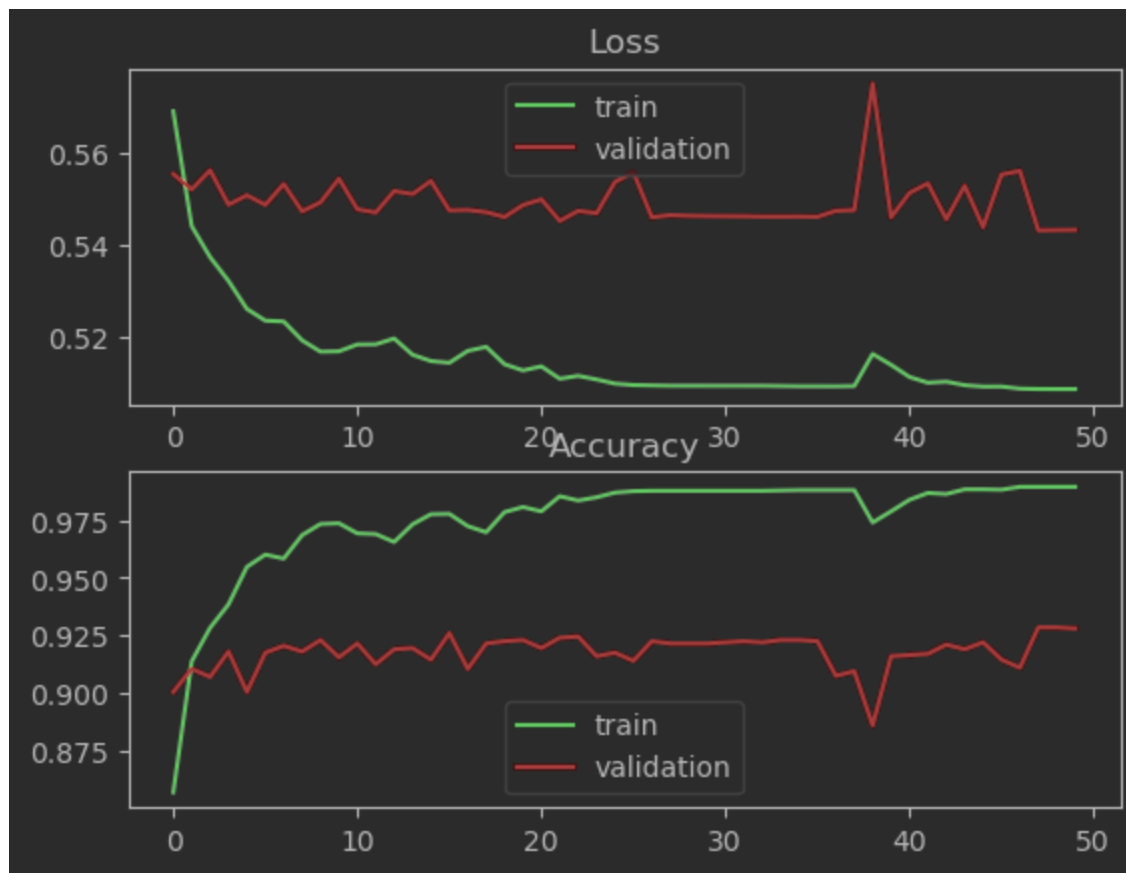
Report the training accuracy that you were able to achieve with the overfit dataset.



I was able to get to 100% training accuracy.

#### Question 5.2.1

Report the accuracy and the full hyperparameter settings. Give the training and validation curves for that best model, and describe your overall hyperparameter tuning approach.

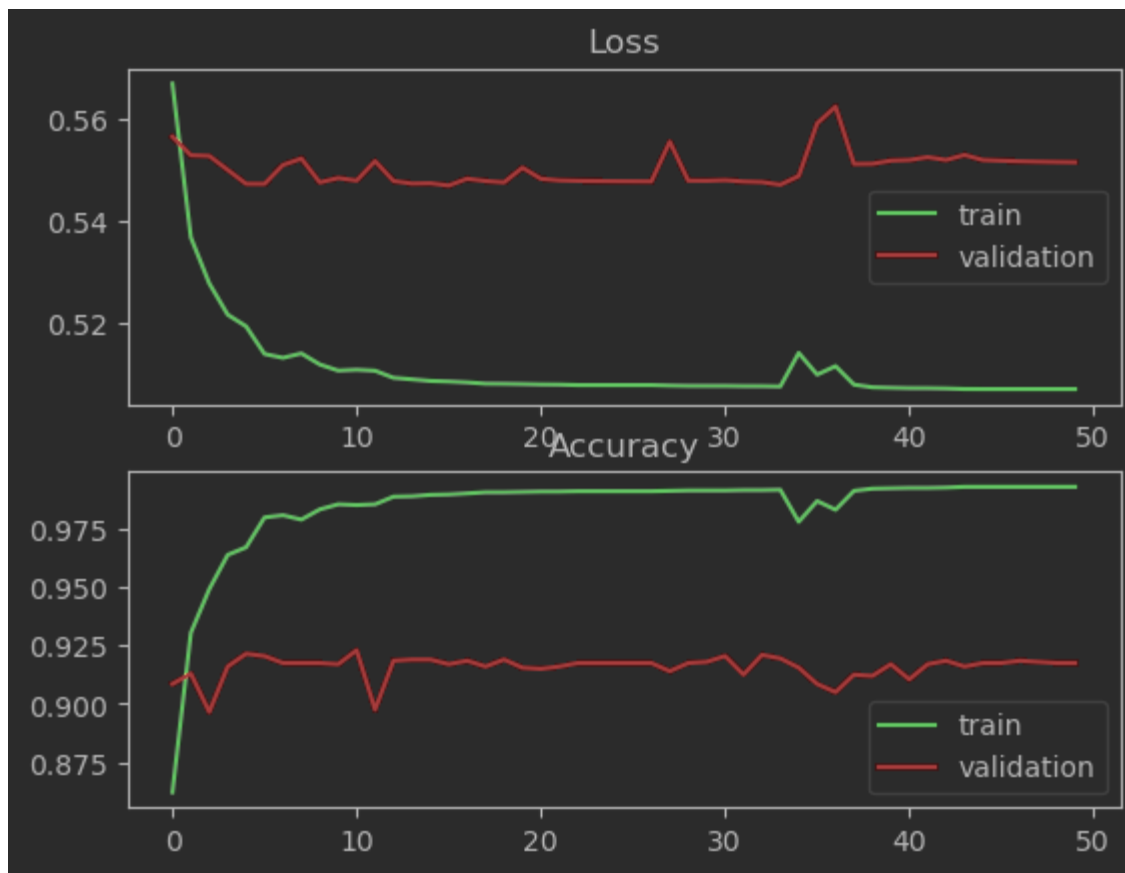


I was able to reach 92.80% test accuracy with  $k_1=3$ ,  $n_1=230$ ,  $k_2=4$ ,  $n_2=230$ . The parameters for the training is 50 epochs with a learning rate of  $1e-3$  and a batch size of 32.

I first arbitrarily set the kernel width to 2 and 7, and then found the best kernel number and batch size via trial and error. I then decided that kernel number of 220 for both and a batch size of 32 was the best, and proceeded to find out the best kernel width under these conditions. I came to the conclusion that  $(k_1, k_2) = (2, 3), (2, 4), (2, 5), (3, 4)$  were the best and then tried these against kernel numbers between 200 and 240, and picked  $k_1=3$ ,  $n_1=230$ ,  $k_2=4$ ,  $n_2=230$ .

### Question 5.2.2

Report the accuracy of the result, and comment on the result.



The test accuracy after unfreezing the embedding layer is 91.75%, which is about the same as before, but the model reached max training accuracy asymptote about twice as fast. The reduced test accuracy could be attributed to noise as both models are given more than enough epochs to train.

### Question 5.3

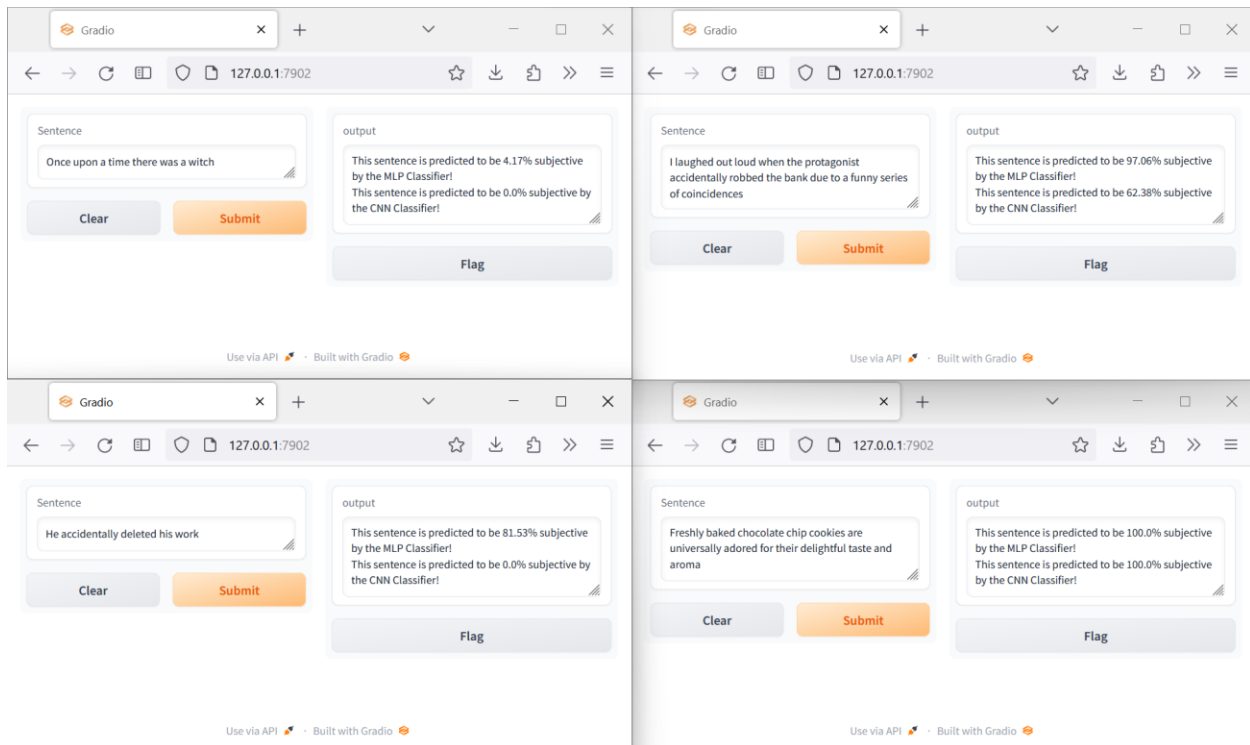
Use that function to determine the five closest words to each of the words in the the kernels trained in your best classifier. Do those words make sense? Do the set of words in each given kernel give a broader insight into what the model is looking for? Explain.

The words makes some sense but not much. Most of the words in the groups of closest words belong to the same part of speech, but that was expected since that's a feature of word embedding not unique to the trained model. However, the sequence of these groups of words that makes up the kernels does make sense. For example there's virtually no sequence that has 2 groups of

consecutive adverbs, and adverbs are more likely to be followed by verbs instead of nouns, etc..

### Question 6.1

Comment on how the two models performed and whether they are behaving as you expected. Do they agree with each other? Which model seems to be performing the best?



The models surprisingly didn't perform so well. They agree with each other if the sentence is definitely objective/subjective but not so much if the sentence is long and complicated. The MLP model was heavily influenced by the number of subjective/objective words in the sentence and its result can vary drastically by changing just 1 word. The CNN model seems to be performing better.