Question 3.1.1

Question: Review the data to see how it is organized in the file. How many examples are in the file data.tsv?

Answer: There are a total of 10,000 examples in the file.

Question 3.1.2

Question: 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.

Answer:
1. "it's another stale , kill–by–numbers flick , complete with about–as–original–as–a–bar–code plot ."
   Explanation: This sentence contains subjective language, such as "stale" and "about–as–original–as–a–bar–code", that expresses the author's negative opinion.
2. "this story of unrequited love doesn't sustain its momentum and the result is a ploddingly slow drama ."
   Explanation: The sentence has subjective terms like "doesn't sustain its momentum" and "ploddingly slow drama", which indicate the author's viewpoint on the pace and quality of the drama.
3. "as this bizarro weekend unfolds , we see exactly what makes these two tick ."
   Explanation: This sentence is more of a descriptive statement about the events in a movie or story. It doesn't contain any explicit positive or negative opinions.
4. "washington heights" tells the story of carlos"
   Explanation: This sentence seems to be the beginning of a summary or description of the plot of "Washington Heights". It's neutral in tone and doesn't offer a subjective viewpoint.

Question 3.1.3

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

Answer:
1. "i'm not exactly sure what this movie thinks it's about ..."
   Explanation: This sentence seems more like a statement of confusion or uncertainty about the movie's theme or message. While it does contain a personal perspective ("I'm

not exactly sure"), it doesn't explicitly convey a positive or negative opinion about the movie. This makes it borderline between subjective and objective.

2. "shortly thereafter , kilo is involved in a drug deal gone bad ..."
   Explanation: This sentence offers a description of events in a story. However, the phrase "drug deal gone bad" has a slightly subjective undertone, implying a negative outcome. It could be considered subjective and therefore potentially mislabeled.

Question 4.4

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

Answer:

Training Loss for Overfitting

Training Accuracy for Overfitting
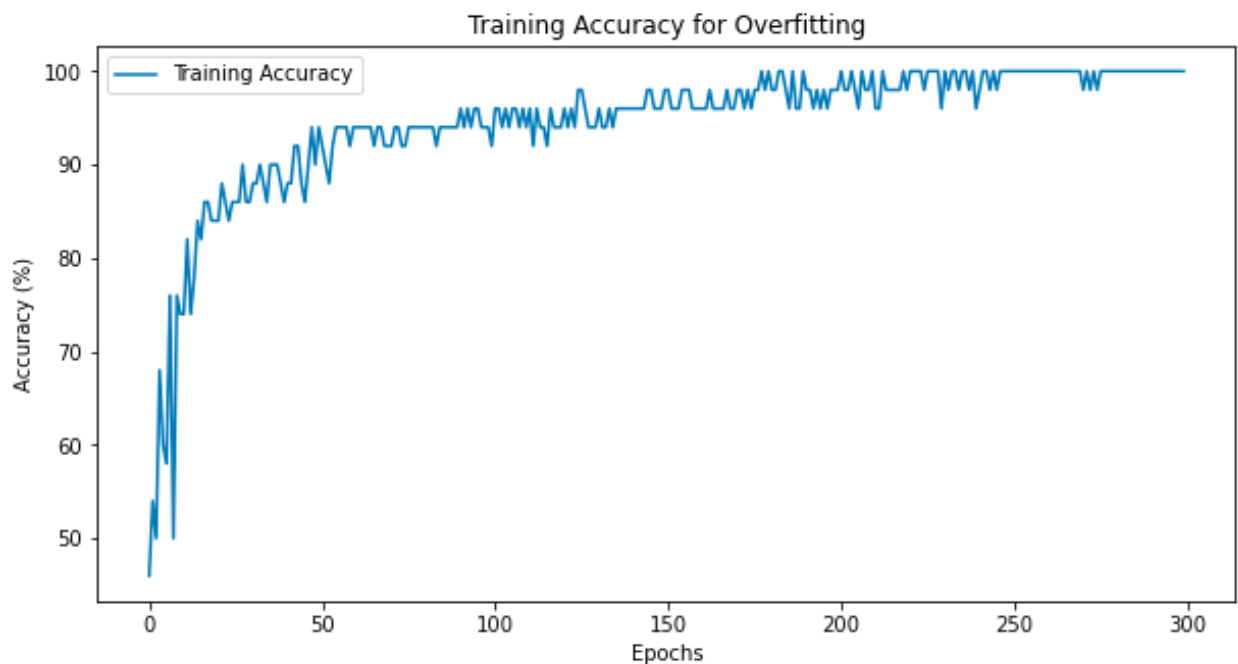
Question 4.5

Question1: 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.

Answer1:

```
Epoch 1/50, Training Loss: 0.6815, Training Accuracy: 61.02%, Validation Loss: 0.6658, Validation Accuracy: 64.83%
Test Loss: 0.6662, Test Accuracy: 63.64%
Epoch 2/50, Training Loss: 0.6525, Training Accuracy: 74.64%, Validation Loss: 0.6387, Validation Accuracy: 78.79%
Test Loss: 0.6395, Test Accuracy: 78.47%
Epoch 3/50, Training Loss: 0.6272, Training Accuracy: 78.50%, Validation Loss: 0.6146, Validation Accuracy: 82.60%
Test Loss: 0.6157, Test Accuracy: 82.82%
Epoch 4/50, Training Loss: 0.6048, Training Accuracy: 81.03%, Validation Loss: 0.5931, Validation Accuracy: 83.79%
Test Loss: 0.5944, Test Accuracy: 84.42%
Epoch 5/50, Training Loss: 0.5846, Training Accuracy: 82.45%, Validation Loss: 0.5737, Validation Accuracy: 84.61%
Test Loss: 0.5752, Test Accuracy: 85.21%
...
Epoch 49/50, Training Loss: 0.3481, Training Accuracy: 87.58%, Validation Loss: 0.3388, Validation Accuracy: 88.36%
Test Loss: 0.3421, Test Accuracy: 87.06%
Epoch 50/50, Training Loss: 0.3468, Training Accuracy: 87.67%, Validation Loss: 0.3373, Validation Accuracy: 88.49%
Test Loss: 0.3407, Test Accuracy: 87.11%
```

Training and Validation Accuracy vs. Epochs
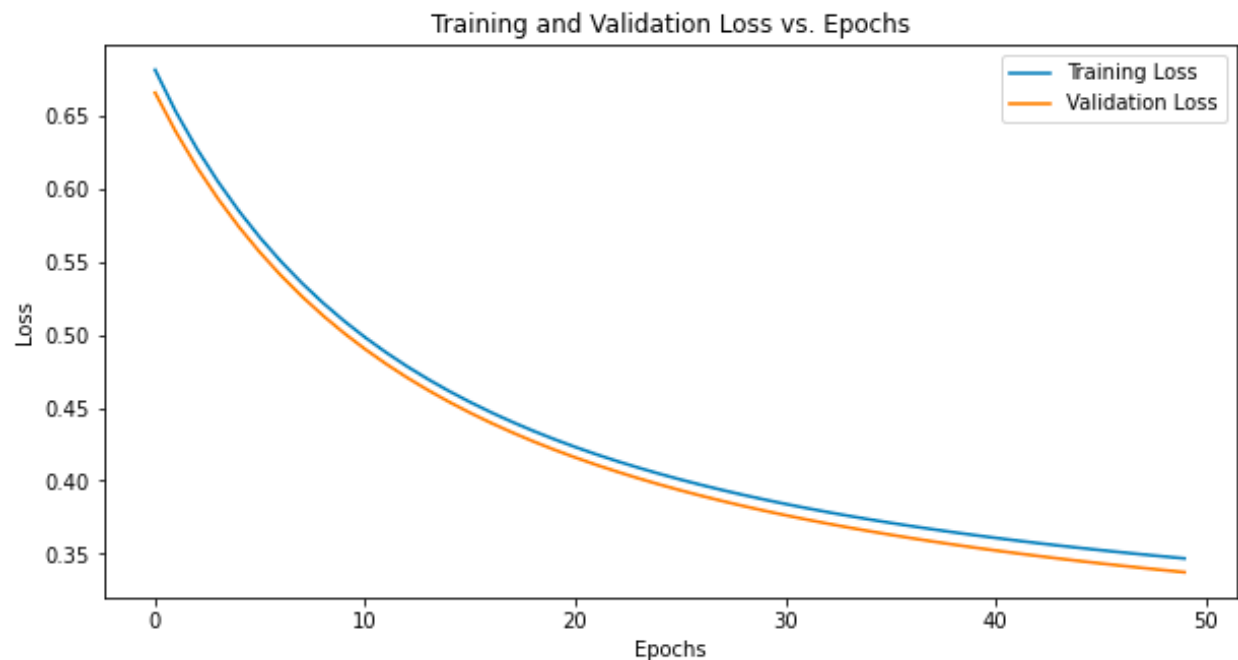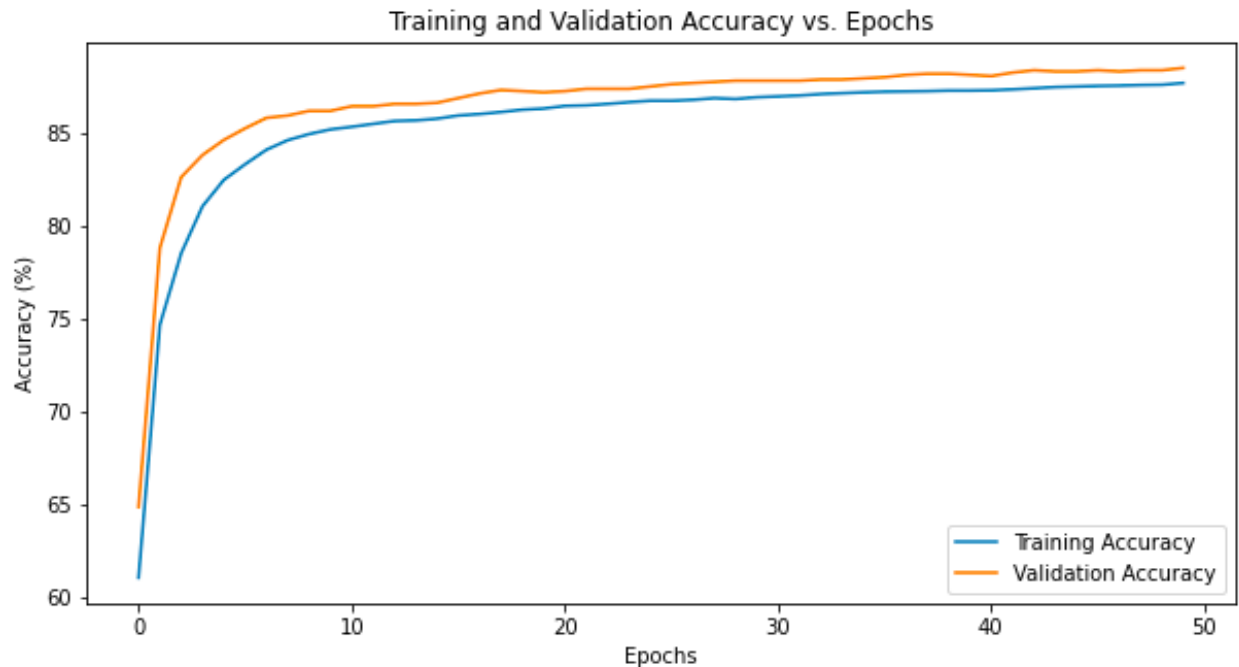
Question2: In the baseline model, what information contained in the original sentence is being ignored?

Answer2:
1. Order information: Information about the order and structure of the words in the sentence is completely ignored.
2. Grammatical and syntactic relationships: By simply averaging word vectors, we cannot capture the grammatical and syntactic relationships between words.
3. Emphasis and embellishment: Word emphasis and embellishment information is also lost.
4. Multiple meanings of words: Certain words may have different meanings in different contexts.
5. Long and short sentence differences: For particularly long or short sentences, calculating the average may cause information loss or dilute key information.
6. Negation and complex structure: Simple averaging methods may not properly handle negation and complex statement structure.

Question 4.6

Question: 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.

Answer:

```
visuals        0.56
simplistic     0.56
pleasing       0.55
watchable      0.54
insipid        0.54
realistic      0.53
flattering     0.53
cheesy   0.52
engrossing     0.52
laughable      0.52
cartoonish     0.52
disquieting    0.52
cartoony       0.51
encapsulates   0.51
verisimilitude 0.51
```

These words do have meaning, and they provide insight into how the model understands and values certain words in a sentence for classification.

As can be seen from the generated vocabulary:

1. Emotionally neutral words: Words like "visuals, ""realistic," and "nuanced" are relatively emotionally neutral and don't necessarily carry explicitly positive or negative emotions.
2. Negative emotion words: such as "insipid", "laughable", "cartoonish" and "silliness". These words often have some negative or pejorative meaning.
3. Positive emotion words: Words like "pleasing", "refreshingly" and "engrossing" carry positive emotions.
4. Punning emotional words: For example, "campy" may be seen as negative in some contexts (referring to the low quality of the work), but may be praised in other contexts (funny, witty).

The combination of these words indicates that the model tries to find various features in the data that are associated with positive, negative, or neutral affect.

Question 5.1

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

Answe   **r:**

```
Epoch 1/10, Loss: 1.4132, Accuracy: 50.00%
Epoch 2/10, Loss: 1.2742, Accuracy: 96.00%
Epoch 3/10, Loss: 1.2268, Accuracy: 70.00%
Epoch 4/10, Loss: 1.1659, Accuracy: 78.00%
Epoch 5/10, Loss: 1.0696, Accuracy: 98.00%
Epoch 6/10, Loss: 1.0121, Accuracy: 100.00%
Epoch 7/10, Loss: 0.9440, Accuracy: 100.00%
Epoch 8/10, Loss: 0.9090, Accuracy: 100.00%
Epoch 9/10, Loss: 0.8335, Accuracy: 100.00%
Epoch 10/10, Loss: 0.7971, Accuracy: 100.00%
Best Training Accuracy: 100.00%
```

Question 5.2.1

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

Answer:
Best accuracy is: 91.86%
Best Hyperparams:
K1: 2, K2: 4, n1: 4, n2: 16, lr: 0.0001

Training and Validation Loss

Training and Validation Accuracy
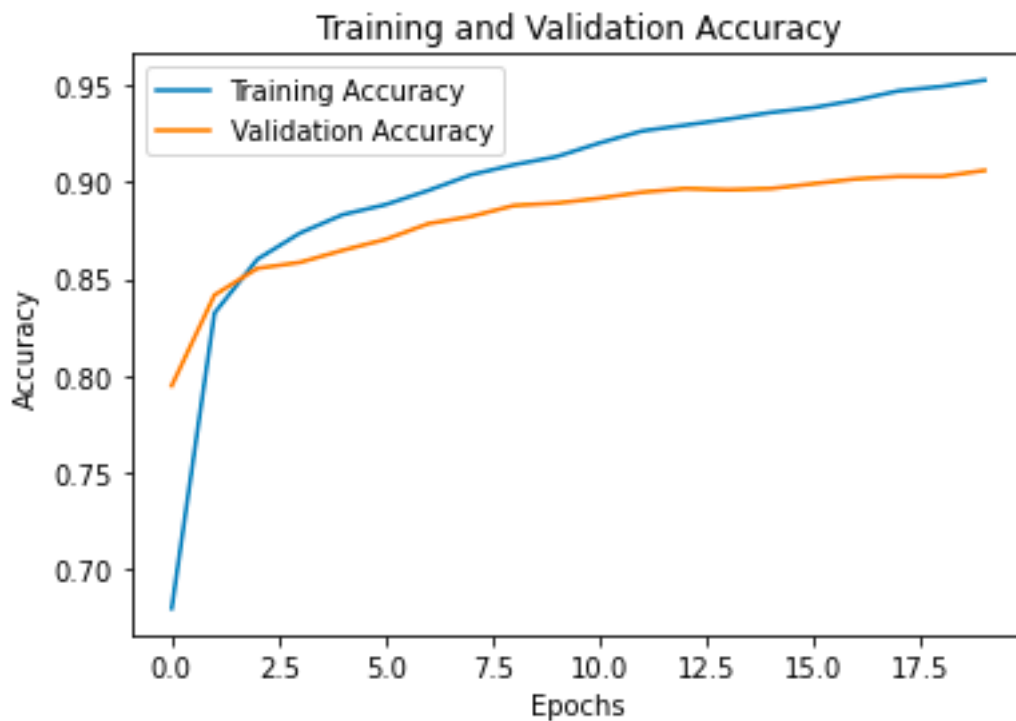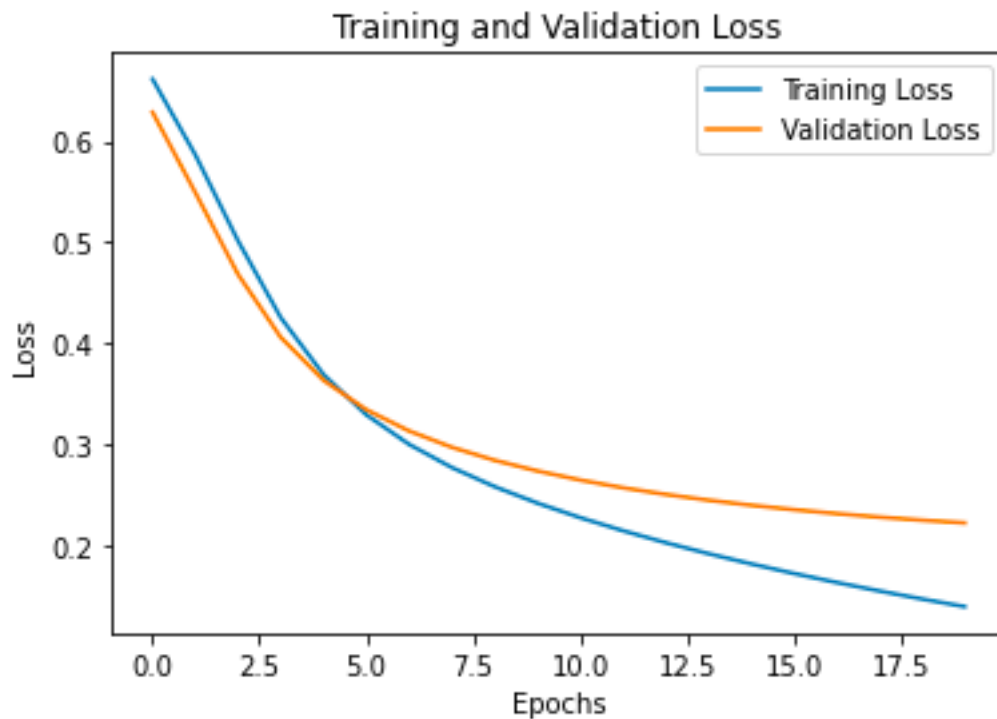
Grid Search is used in tuning the model hyperparameters. When using Grid Search, the parameters tuned are k1, n1, k2, n2 and learning rate lr. And we use 30 epochs, validate the model per 3 epochs to save tuning time. The model achieved best test accuracy of 91.86% with the following params: {'k1': 2, 'n1': 4, 'k2': 4, 'n2': 16, 'lr': 0.0001}

Question 5.2.2

Question: Report the accuracy of the result, and comment on the result. Save this model in a .pt file as you did in Section 4.7, for use below in Section 6

Answer: The accuracy is 90.93%. Which is slightly lesser than model with unfrozen embedding layer.
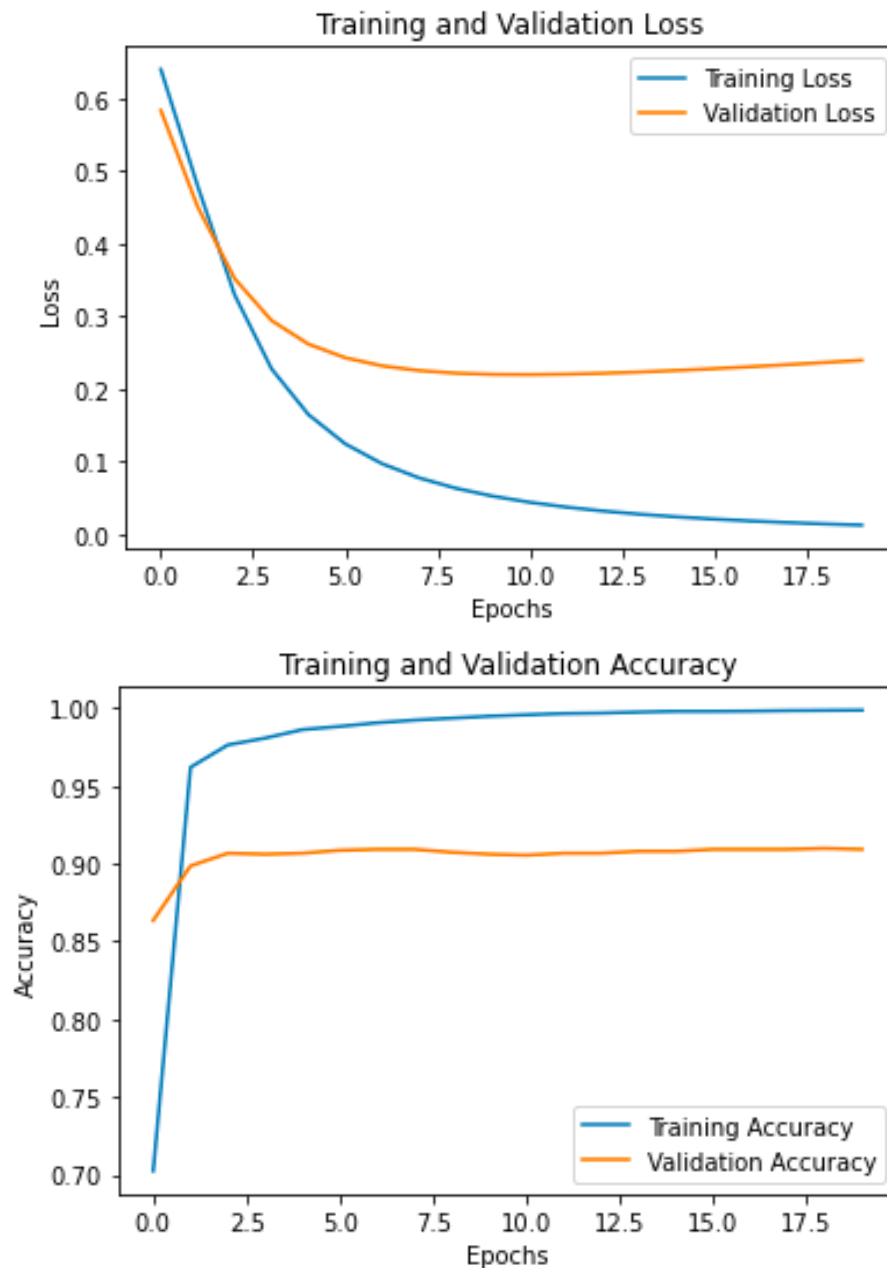
Question 5.3

Question: 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.

Answer: using some example to explain

```
odict_keys(['embedding.weight', 'conv1.weight', 'conv2.weight', 'fc.weight', 'fc.bias'])
funeral          0.56
fiancee          0.50
gunned   0.50
gunmen   0.49
slain    0.49


abducted         0.51
jailed   0.48
kidnapped        0.48
imprisoned       0.47
custody          0.46
```

```
physician-assisted       0.45
dejanews         0.45
netcentric       0.44
vitellia         0.43
novitiate        0.42


shavon   0.48
boukman          0.48
traz     0.48
tuho     0.47
nicholaus        0.46
```

We expect the model to be able to distinguish between subjective and objective sentences. Subjective sentences usually contain words with subjective information such as

emotions, opinions, beliefs, and prejudices, while objective sentences usually describe facts, real events, or universal truths.

Words for the first and second kernels:
(funeral, fiancee, gunned, gunmen, slain)
(abducted, jailed, kidnapped, imprisoned, custody)

Both sets of words are related to some more emotional or subjective topics, such as death, disappearance, arrest, etc. This could mean that these convolutional kernels are more sensitive to capture these subjective themes or emotionalized content.

Some kernels seem to be more sensitive to certain subjective or emotional topics, which may be useful for subjective/objective sentence classification.

Question 6

Question: 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?

Answer:

1. **Strong subjective:** I think the movie we watched today is great.
Both of the model works perfect.

Their predictions are different. I think this is why this sentence is weak subjective, since in the one hand, the sentence is truth, but in the other hand, there is some subjective word like most. And, I agree with CNN.

3. **Strong Objective:** The People's Republic of China built in 1949.

Both of them works perfect.

4. **Weak Objective:** The Ming Dynasty is one of the most strong dynasty in the world

This is tricky, but the result is perfect. The sentence "in the world" is kind of subjective. But the other parts are objective. So, both of them think this is a subjective case. Perfect!

The CNN model performs better than the baseline model, since the sentence is between objective and subjective.