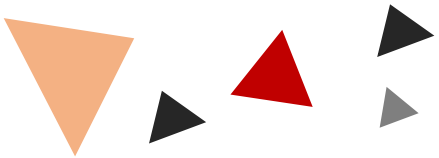


Deep Classification, Embedding & Text Generation

YIMIN LI
liym15@uchicago.edu
May 22, 2020

CONTENTS

Data

- ✓ **Amazon review data (2018) with 233.1 million reviews**
- ✓ **In this homework, we conducted on a small scope:**
- **Sub-Category: Video Games (400,000 inputs)**
- **Select First 5,000 queries in the video games**
- **Goal: Understand reviewers' behavior and extract high-level information from it**

	overall	verified	reviewTime	reviewerID	asin	reviewerName	reviewText	summary	unixReviewTime
0	5	True	10 17, 2015	A1HP7NVNPFMA4N	700026657	Ambrosia075	This game is a bit hard to get the hang of, bu...	but when you do it's great.	1445040000
1	4	False	07 27, 2015	A1JGAP0185YJI6	700026657	travis	I played it a while but it was alright. The st...	But in spite of that it was fun, I liked it	1437955200
2	3	True	02 23, 2015	A1YJWEXHQBWK2B	700026657	Vincent G. Mezera	ok game.	Three Stars	1424649600
3	2	True	02 20, 2015	A2204E1TH211HT	700026657	Grandma KR	found the game a bit too complicated, not what...	Two Stars	1424390400
4	5	True	12 25, 2014	A2RF5B5H74JLPE	700026657	jon	great game, I love it and have played it since...	love this game	1419465600

Data

	overall	verified	reviewTime	reviewerID	asin	reviewerName	reviewText	summary	unixReviewTime
0	5	True	10 17, 2015	A1HP7NVNPFMA4N	700026657	Ambrosia075	This game is a bit hard to get the hang of, bu...	but when you do it's great.	1445040000
1	4	False	07 27, 2015	A1JGAP0185YJI6	700026657	travis	I played it a while but it was alright. The st...	But in spite of that it was fun, I liked it	1437955200
2	3	True	02 23, 2015	A1YJWEXHQBWK2B	700026657	Vincent G. Mezera	ok game.	Three Stars	1424649600
3	2	True	02 20, 2015	A2204E1TH211HT	700026657	Grandma KR	found the game a bit too complicated, not what...	Two Stars	1424390400
4	5	True	12 25, 2014	A2RF5B5H74JLPE	700026657	jon	great game, I love it and have played it since...	love this game	1419465600

	label	reviewerID	reviewText
0	1	A1HP7NVNPFMA4N	This game is a bit hard to get the hang of, bu...
1	0	A1JGAP0185YJI6	I played it a while but it was alright. The st...
2	0	A1YJWEXHQBWK2B	ok game.
3	0	A2204E1TH211HT	found the game a bit too complicated, not what...
4	1	A2RF5B5H74JLPE	great game, I love it and have played it since...

- **Manually re-labelled the data**

to binary variables

- **Distribution:**

5-star(labelled as 1): 3388

4-star(labelled as 0): 1612

Deep Neural Nets (Keras)

```
Epoch 1/10
4500/4500 [==============================] - 31s 7ms/step - loss: 0.6249 - accuracy: 0.6758
Epoch 2/10
4500/4500 [==============================] - 29s 6ms/step - loss: 0.6035 - accuracy: 0.6987
Epoch 3/10
4500/4500 [==============================] - 28s 6ms/step - loss: 0.5381 - accuracy: 0.7558
Epoch 4/10
4500/4500 [==============================] - 28s 6ms/step - loss: 0.6088 - accuracy: 0.6798
Epoch 5/10
4500/4500 [==============================] - 28s 6ms/step - loss: 0.5047 - accuracy: 0.7704
Epoch 6/10
4500/4500 [==============================] - 28s 6ms/step - loss: 0.5450 - accuracy: 0.7584
Epoch 7/10
4500/4500 [==============================] - 28s 6ms/step - loss: 0.4611 - accuracy: 0.8029
Epoch 8/10
4500/4500 [==============================] - 28s 6ms/step - loss: 0.4631 - accuracy: 0.8096
Epoch 9/10
4500/4500 [==============================] - 28s 6ms/step - loss: 0.4016 - accuracy: 0.8324
Epoch 10/10
4500/4500 [==============================] - 27s 6ms/step - loss: 0.3628 - accuracy: 0.8529
```

- **Keras (LSTM):**

high-level package that we don't bother every detail or hyperparameters associated with neural Network

- **Results:**

Accuracy: 67 – 85%

Run-time: ~290s

Fairly fine accuracy but with high variance (?)

Deep Neural Nets (Keras)

```
Epoch 1/10
4500/4500 [==============================] - 93s 21ms/step - loss: 0.6272 - accuracy: 0.6780
Epoch 2/10
4500/4500 [==============================] - 88s 20ms/step - loss: 0.5946 - accuracy: 0.7084
Epoch 3/10
4500/4500 [==============================] - 89s 20ms/step - loss: 0.5224 - accuracy: 0.7711
Epoch 4/10
4500/4500 [==============================] - 90s 20ms/step - loss: 0.5708 - accuracy: 0.7449
Epoch 5/10
4500/4500 [==============================] - 90s 20ms/step - loss: 0.4352 - accuracy: 0.8193
Epoch 6/10
4500/4500 [==============================] - 90s 20ms/step - loss: 0.3547 - accuracy: 0.8511
Epoch 7/10
4500/4500 [==============================] - 89s 20ms/step - loss: 0.3106 - accuracy: 0.8722
Epoch 8/10
4500/4500 [==============================] - 88s 20ms/step - loss: 0.2598 - accuracy: 0.9027
Epoch 9/10
4500/4500 [==============================] - 89s 20ms/step - loss: 0.2267 - accuracy: 0.9204
Epoch 10/10
4500/4500 [==============================] - 89s 20ms/step - loss: 0.2098 - accuracy: 0.9260
```

Added one more layer

- **Results:**

Accuracy: 67 – 92% (better)

Run-time: ~900s (3 times than before)

Better than one layer with the largest accuracy of 92.6% (super excellent!)

But much more time-consuming (costly)

BERT (bidirectional transformer embedding)

```
======== Epoch 1 / 4 ========
Training...
  Batch    40  of    141.    Elapsed: 0:00:16.
  Batch    80  of    141.    Elapsed: 0:00:31.
  Batch   120  of    141.    Elapsed: 0:00:46.

  Average training loss: 0.53
  Training epcoh took: 0:00:54

Running Validation...
  Accuracy: 0.78
  Validation took: 0:00:02

======== Epoch 2 / 4 ========
Training...
  Batch    40  of    141.    Elapsed: 0:00:15.
  Batch    80  of    141.    Elapsed: 0:00:30.
  Batch   120  of    141.    Elapsed: 0:00:45.

  Average training loss: 0.40
  Training epcoh took: 0:00:53

Running Validation...
  Accuracy: 0.81
  Validation took: 0:00:02
```

```
======== Epoch 3 / 4 ========
Training...
  Batch    40  of    141.    Elapsed: 0:00:15.
  Batch    80  of    141.    Elapsed: 0:00:30.
  Batch   120  of    141.    Elapsed: 0:00:45.

  Average training loss: 0.30
  Training epcoh took: 0:00:53

Running Validation...
  Accuracy: 0.80
  Validation took: 0:00:02

======== Epoch 4 / 4 ========
Training...
  Batch    40  of    141.    Elapsed: 0:00:15.
  Batch    80  of    141.    Elapsed: 0:00:30.
  Batch   120  of    141.    Elapsed: 0:00:45.

  Average training loss: 0.22
  Training epcoh took: 0:00:53

Running Validation...
  Accuracy: 0.80
  Validation took: 0:00:02
```

BERT Model

- **Results:**

Training Loss: 0.22 – 0.52

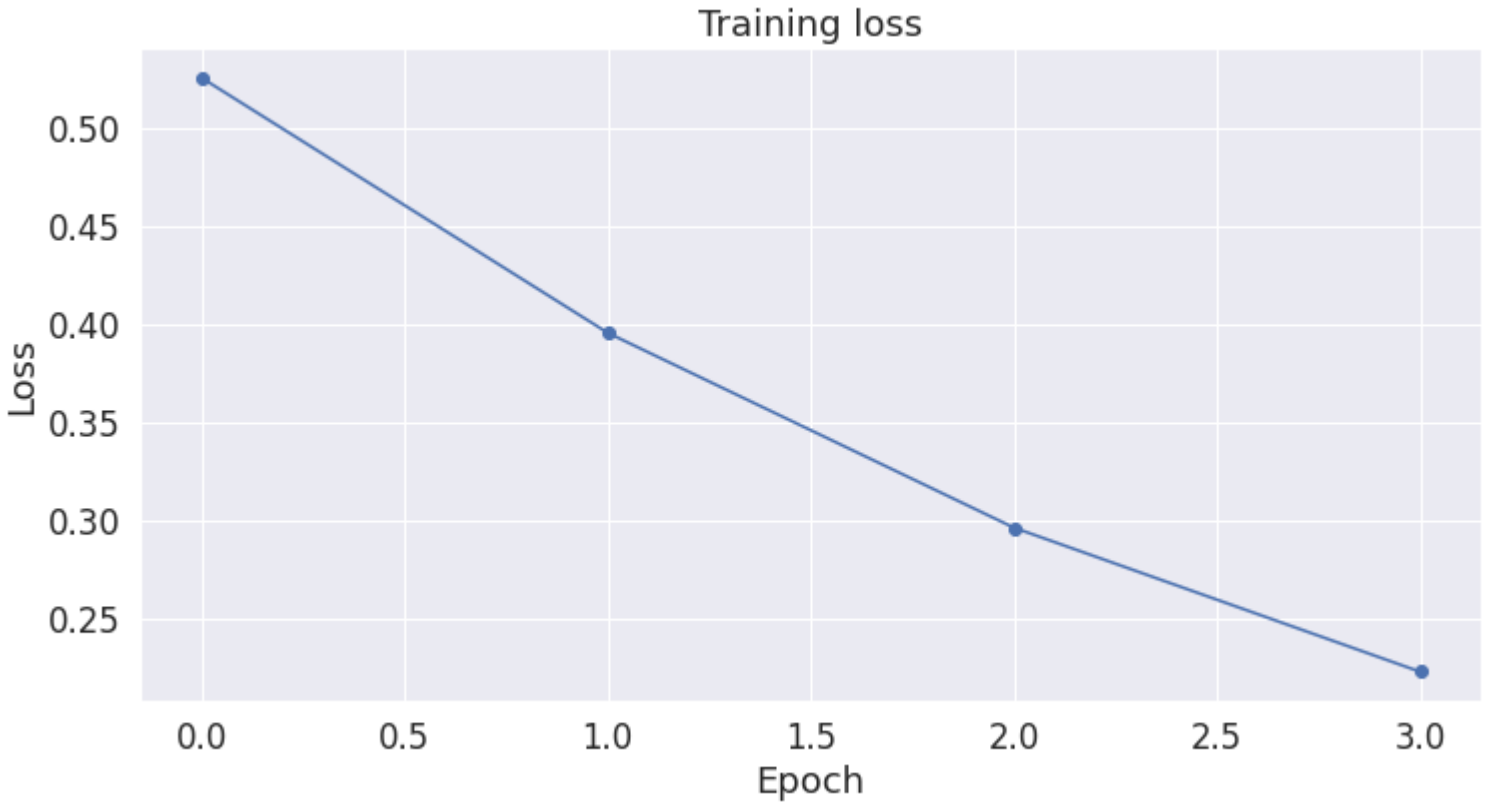
Accuracy: 78 – 81%

Run-time: ~210s on Colab

(~12h on local machine)

Fairly accurate with low variance but with strong reliance on GPU

BERT (bidirectional transformer embedding)



Some Findings:

- **When Epoch = 3.0, training loss can become half of the Epoch = 0.0**
- **As Epoch numbers increase, the training loss decreases.**
- **Generally speaking, it performs better than LSTM.**

Model Evaluation

- ✓ Number of test sentences: 5,000
- ✓ Positive samples (1): 3388 of 5000 (67.76%)
- ✓ MCC (Matthews Correlation Coefficient): 0.845

Excellent Prediction!

Some Findings:

- **Keras and BERT performs much better than the traditional models but also time-consuming (rely on GPU.)**

Model	Error rate	Model	Error rate
Keras(two-layer)	0.18(avg.)	Decision Tree	0.39
BERT	0.19(avg.)	Random Forest	0.41
Logistic Regression	0.36	KNN	0.24
Naïve Bayes	0.32	Neural Network	0.42

Model Evaluation

- ✓ Number of test sentences: 5,000
- ✓ Positive samples (1): 3388 of 5000 (67.76%)
- ✓ MCC (Matthews Correlation Coefficient): 0.845

Excellent Prediction!

Some Findings:

- **Keras and BERT performs much better than the traditional models but also time-consuming (rely on GPU.)**

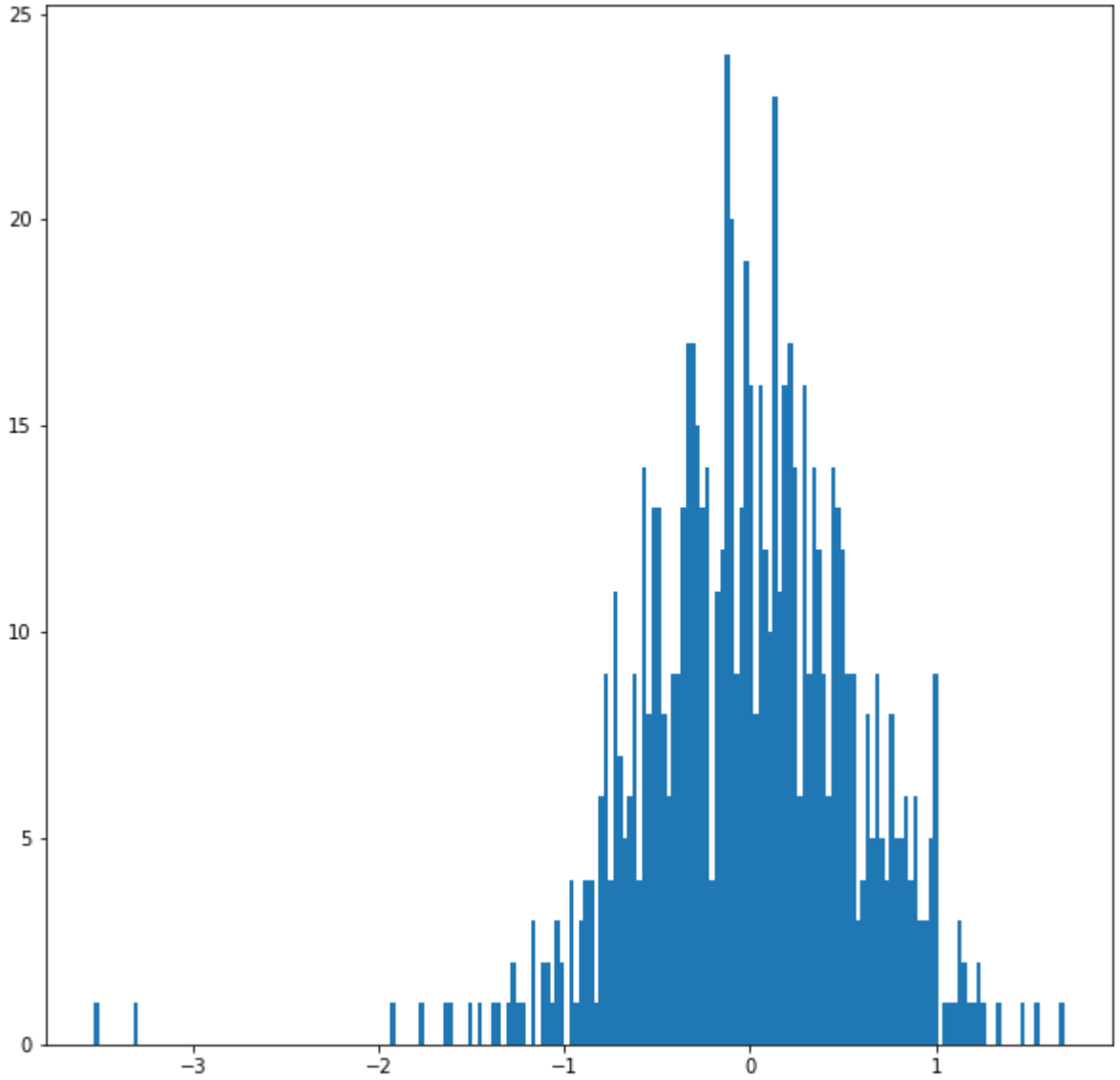
Model	Error rate	Model	Error rate
Keras(two-layer)	0.18(avg.)	Decision Tree	0.39
BERT	0.19(avg.)	Random Forest	0.41
Logistic Regression	0.36	KNN	0.24
Naïve Bayes	0.32	Neural Network	0.42

Word Embeddings

The distribution of range of values

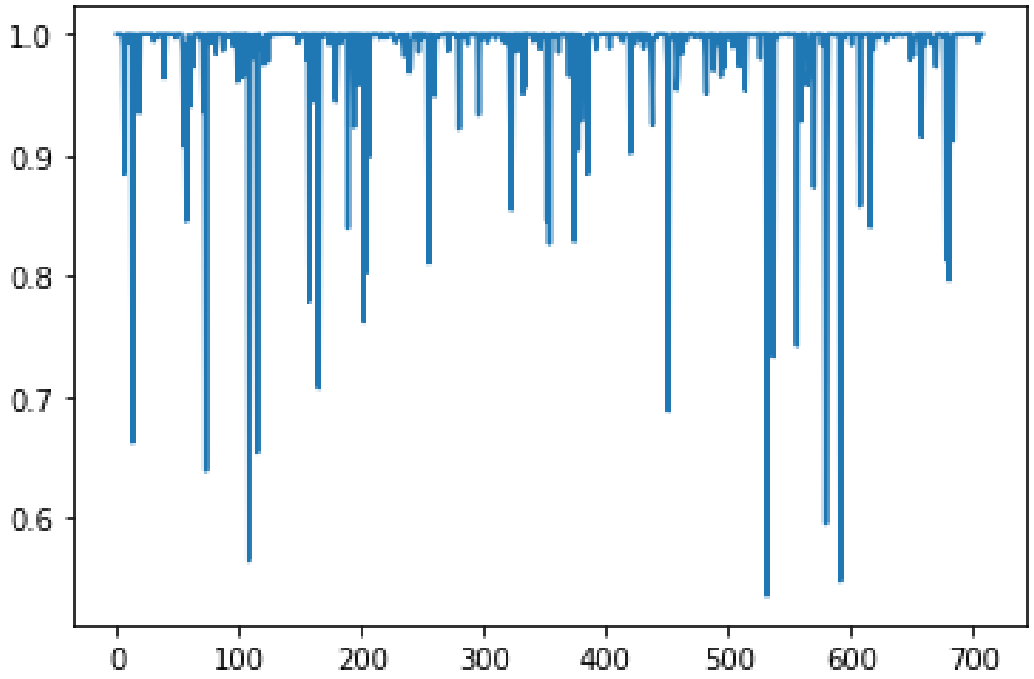
for word embeddings for BERT:

- **Majority around [-1, 1] (a little bit more in the negative side)**
- **With several outliers near -4**

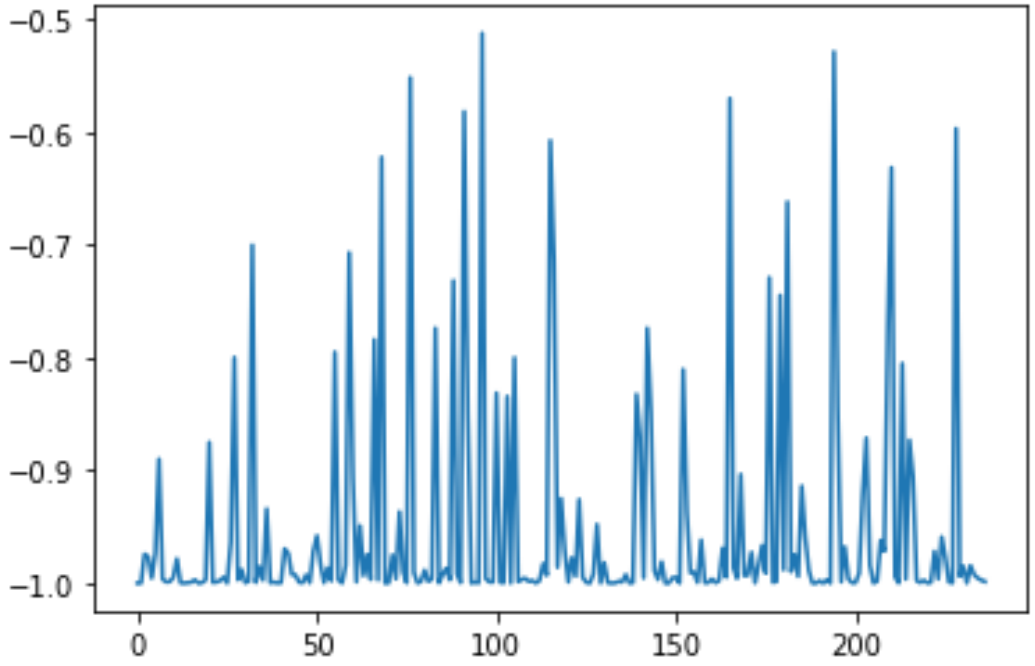


Transformers Pipeline (sentiments)

Positive



Negative



Some Findings:

- **More positive than negative (around 3:1)**
- **Strong positive and weak negative (bigger variance)**

Text Generation (all data)

- ✓ **All dataset: 5000 inputs**
- ✓ **Train/Test split: 0.2 (Train: 4000, Test: 1000)**
- ✓ **Begin words for text generation: This game**

```
=== GENERATED SEQUENCE 1 ===
This game was remarkably innovative and fun.  It has a complete line of games to choose from, such as
```

Findings:

- **High-level pattern: Tend to be very positive (innovative and fun)**
- **It matches with the fairly positive sentimental score shown in the prior presentation**

Text Generation (all data)

✓ **Begin words for text generation: I like this video game because**

```
I like this video game because it's fun and it's fun to play. I also like the fact that you can play as a character and hav
e a lot of fun. I also like the fact that you can play as a character and have a
```

✓ **Begin words for text generation: I don’t like this video game because**

```
I don't like this video game because it's so boring.  I don't like the graphics, but I like the gameplay.  I like the stor
y, but I don't like the characters.  I like the music, but I don
```

- **Interesting to see that it can generate make-sense positive and negative sentence**

Text Generation (five-star data)

- ✓ **Five-star dataset: 500 inputs (all five-star)**
- ✓ **Train/Test split: 0.2 (Train: 400, Test: 100)**
- ✓ **Begin words for text generation: This game**

```
=== GENERATED SEQUENCE 1 ===
This game is fun enough for both new and old players alike, but it's not as intense as the original.
```

Findings:

- **High-level pattern: Positive**
- **Interestingly, it generates a “but” here in all five-star comments.**

Text Generation (not-five-star data)

- ✓ **Five-star dataset: 500 inputs (all not five-star)**
- ✓ **Train/Test split: 0.2 (Train: 400, Test: 100)**
- ✓ **Begin words for text generation: This game**

```
=== GENERATED SEQUENCE 1===
This game is boring. I don't like the graph, but I like the game settings.
```

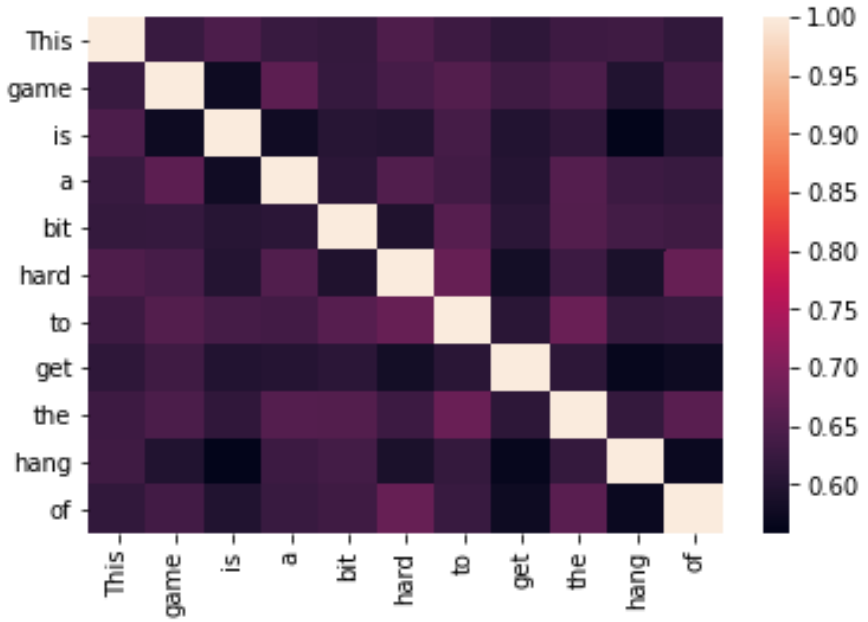
Findings:

- **High-level pattern: Negative (The text-generation really makes sense!)**
- **Interestingly, it also generates a “but” here.**

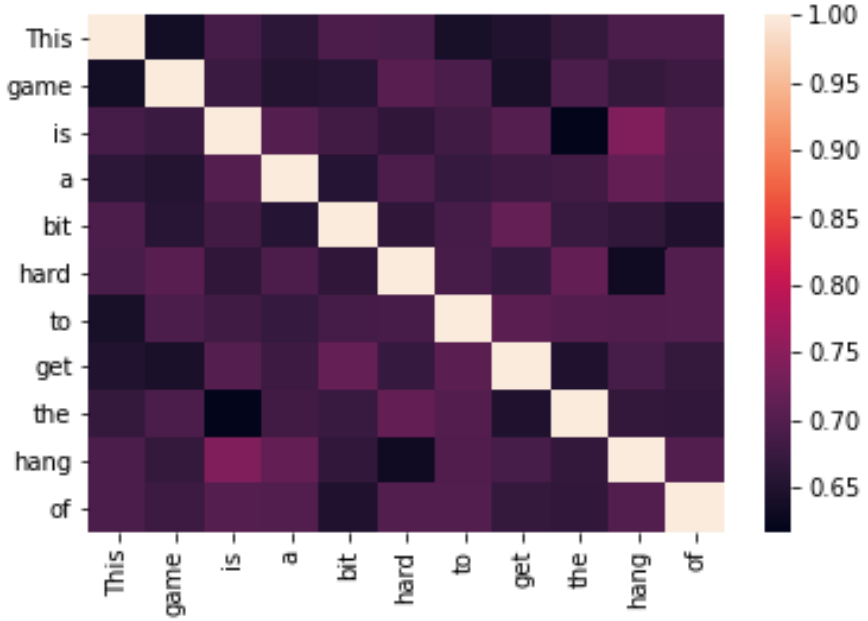
Heatmap

✓ **Text: "This game is a bit hard to get the hang of."**

All reviews



All five-star reviews



Findings:

- **Hard to distinguish between these two datasets for this text**
- **Maybe because words in this text in so common to use.**

Conclusion

✓ **Deep Classification:**

- **Keras and Bert performs better than the traditional ML models (half of the error rates)**
- **But they are more costly (time-consuming and relies on GPU)**

✓ **Word Embeddings:**

- **Most words embeddings around [-1, 1]**
- **More positive sentiments than negative and positive tends to be much stronger**

✓ **Text Generation:**

- **Can excellently generate sentences make sense (fit to the grammar, topic and sentiments)**
- **Interestingly, it “loves” to use “but”**
- **Hard to distinguish sentences heatmap between different models and datasets**

Conclusion

✓ **Deep Classification:**

- **Keras and Bert performs better than the traditional ML models (half of the error rates)**
- **But they are more costly (time-consuming and relies on GPU)**

✓ **Word Embeddings:**

- **Most words embeddings around [-1, 1]**
- **More positive sentiments than negative and positive tends to be much stronger**

✓ **Text Generation:**

- **Can excellently generate sentences make sense (fit to the grammar, topic and sentiments)**
- **Interestingly, it “loves” to use “but”**
- **Hard to distinguish sentences heatmap between different models and datasets**

Conclusion

✓ **Deep Classification:**

- **Keras and Bert performs better than the traditional ML models (half of the error rates)**
- **But they are more costly (time-consuming and relies on GPU)**

✓ **Word Embeddings:**

- **Most words embeddings around [-1, 1]**
- **More positive sentiments than negative and positive tends to be much stronger**

✓ **Text Generation:**

- **Can excellently generate sentences make sense (fit to the grammar, topic and sentiments)**
- **Interestingly, it “loves” to use “but”**
- **Hard to distinguish sentences heatmap between different models and datasets**