COMP 551 Assignment-3

Yufei Liu

TOTAL POINTS

95.3 / 99

QUESTION 1

1 Late Submission **1 / 1**

✓ **- 0 pts** **Correct**

- 1 pts Late

QUESTION 2

2 Verbosity **1 / 1**

✓ **- 0 pts** **Correct**

QUESTION 3

3 Q1 **10 / 10**

✓ **- 0 pts** **Correct**

QUESTION 4

4 Q2.a **4 / 4**

✓ **- 0 pts** **Correct**

- 1 pts Majority class score low

- 4 pts not reported or misreported

- 1 pts unreported uniform classifier

- 1 pts misunderstanding

- 1 pts scores too high

QUESTION 5

5 Q2.b **10 / 10**

✓ **- 0 pts** **Correct**

- 10 pts nothing reported

QUESTION 6

6 Q2.c **5 / 5**

✓ **- 0 pts** **Correct**

- 0.5 pts insufficient explanation for best values used

- 0.5 pts insufficent hyperparameters for SVM

- 1 pts No (or insufficient) Naive Bayes hyperparameters tuned

- 0.5 pts insufficient Decision Tree hyperparams

- 1.25 pts no Decision tree hyperparameters

- 1 pts insufficient hyperparamters for Decision tree

- 0.5 pts explain hyperparameters

- 1 pts insufficient explanation

- 2 pts insufficient explanation

- 1.25 pts no SVM hyperparameters

- 0.5 pts test more cases

- 5 pts incorrect or nothing reported

QUESTION 7

7 Q2.d **6 / 6**

✓ **- 0 pts** **Correct**

- 1 pts scores too high

- 1.5 pts scores too low

- 2.5 pts missing results

- 1 pts missing results

- 2 pts misunderstanding

- 5 pts incorrect results

- 1 pts incorrect results

- 0.5 pts poor presentation

- 0 pts Click here to replace this description.

- 1 pts misunderstanding

QUESTION 8

8 Q2.e **4 / 4**

✓ **- 0 pts** **Correct**

- 1 pts further explanation needed

- 2 pts insufficient explanation

- 3 pts insufficient or no explanation provided

- 0.5 pts be more precise

- 4 pts no explanation provided

QUESTION 9

9 Q3.a **2 / 2**

✓ **- 0 pts** **Correct**

- 2 pts Click here to replace this description.

- **0 pts** Click here to replace this description.

QUESTION 10

10 Q3.b 2 / 2

✓ **- 0 pts Correct**

- **2 pts** Click here to replace this description.

QUESTION 11

11 Q3.c 6 / 6

✓ **- 0 pts Correct**

- **1 pts** We're looking for the average F1
- **1 pts** Linear SVM F1 training F1 should be close to Valid F1
- **1 pts** Gaussian NB Validation F1 too high
- **1 pts** DecisionTree F1 Validation F1 too high
- **1 pts** SVM Validation F1 too high
- **2 pts** No validation F1 values
- **6 pts** 6
- **4 pts** F1 Scores too low

QUESTION 12

12 Q3.d 1.5 / 3

- **0 pts** Correct

✓ **- 1.5 pts Discussion about reason for classifier performance missing**

- **3 pts** Question missing

QUESTION 13

13 Q3.e 3 / 3

✓ **- 0 pts Correct**

- **1 pts** No explanation of reason
- **1 pts** Frequency information provides usefulness to representation
- **1.5 pts** This question asks about FBOW and BBOW
- **3 pts** Missing question
- **3 pts** answer unclear

QUESTION 14

14 Q3.f 2 / 2

✓ **- 0 pts Correct**

- **2 pts** Question missing
- **1 pts** Question is missing, but I understand your meaning.
- **1 pts** This question refers to BBOW and FBOW

QUESTION 15

15 Q4.a 4 / 4

✓ **- 0 pts Correct**

- **4 pts** Performance not reported or submitted.
- **3 pts** Incorrect performance. There are only 2, evenly distributed classes.

QUESTION 16

16 Q4.b 4 / 4

✓ **- 0 pts Correct**

- **4 pts** Not submitted.

QUESTION 17

17 Q.4.c 2 / 2

✓ **- 0 pts Correct**

- **2 pts** Needed to report tested values and the best choice based on this dataset.
- **1 pts** Needed to reported tested values.
- **1 pts** Did not test hyper-parameters for all three methods.
- **2 pts** Not submitted.

QUESTION 18

18 Q.4.d 6 / 6

✓ **- 0 pts Correct**

- **2 pts** Poor/unrealistic results for Naive Bayes.
- **2 pts** Poor/unrealistic results for Decision Trees.
- **2 pts** Poor/unrealistic results for SVM.
- **5 pts** Poor/unrealistic results on all models.
- **4 pts** Needed to report results for all models.

QUESTION 19

19 Q4.e 2 / 3

- **0 pts** Correct
- **1 pts** Should justify why the best classifier performed the best.

✓ **- 1 pts Partially incorrect or incomplete justification/explanation.**

- **2 pts** Missing explanation.

- **3 pts** Incorrect or not submitted.
- **1 pts** Missing discussion on hyper-parameters.

QUESTION 20

20 Q5.a 2 / 2

✓ - **0 pts** **Correct**
- **2 pts** Not done.

QUESTION 21

21 Q5.b 1.8 / 2

- **0 pts** Correct

✓ - **0.2 pts** **Training data used for validation / One hyper parameter missing**
- **0.4 pts** Click here to replace this description.
- **0.6 pts** Click here to replace this description.
- **0.8 pts** Click here to replace this description.
- **2 pts** cannot find answer to the question

Missing Parameters for Decision trees:
- Criterion

QUESTION 22

22 Q5.c 6 / 6

✓ - **0 pts** **Correct**
- **2 pts** error values of one of the classifiers
- **4 pts** error in score of 2 models
- **6 pts** No answer found or all scores are incorrect
- **0.66 pts** Error in a single value

QUESTION 23

23 Q5.d 2 / 2

✓ - **0 pts** **Correct**
- **0.5 pts** Partially correct
- **2 pts** Incorrect or no answer found.
- **0.25 pts** Mostly correct with minor errors.

QUESTION 24

24 Q5.e 2 / 3

- **0 pts** Correct
- **0.25 pts** Largely correct with minor errors like lack of clarity/incorrect assumptions
- **2 pts** Incomplete answer
- **0.5 pts** Correct with errors like lack of clarity and or incorrect assumptions with incorrect inference.
- **1.5 pts** Partially correct answer
- **0.1 pts** Minor errors

✓ - **1 pts** **Partially correct answer.**
- **3 pts** No answer or incorrect answer.

The reduction in performance is due to more importance given to frequently occurring stop words like 'a', 'the', etc which do not convey any useful information.

QUESTION 25

25 Q5.f 2 / 2

✓ - **0 pts** **Correct**
- **0.25 pts** Largely correct but with minor errors (lack of clarity, incorrect assumptions)
- **0.5 pts** Correct with errors in certain parts (lack of clarity, incorrect inference)
- **1 pts** Partially correct answer with incorrect assumptions or inference
- **2 pts** Incorrect or no answer found.

QUESTION 26

26 Q5.g 4 / 4

✓ - **0 pts** **Correct**
- **4 pts** No Answer Found
- **1 pts** Partially correct answer
- **0.5 pts** Partially correct answer
- **2 pts** Partially correct answer

Assignment 3 Report

Yufei Liu
260561054

Question 1

Check the dataset files in folder “output_datasets”, it should contain 8 files.

1 Late Submission 1 / 1

✓ **- 0 pts** **Correct**

- 1 pts Late

Assignment 3 Report

Yufei Liu
260561054

Question 1

Check the dataset files in folder “output_datasets”, it should contain 8 files.

2 Verbosity 1 / 1

✓ - 0 pts Correct

Assignment 3 Report

Yufei Liu
260561054

Question 1

Check the dataset files in folder “output_datasets”, it should contain 8 files.

3 Q1 10 / 10

✓ - 0 pts Correct

Question 2

(a)

Uniform Random F1 Measure = 0.20350000000000001

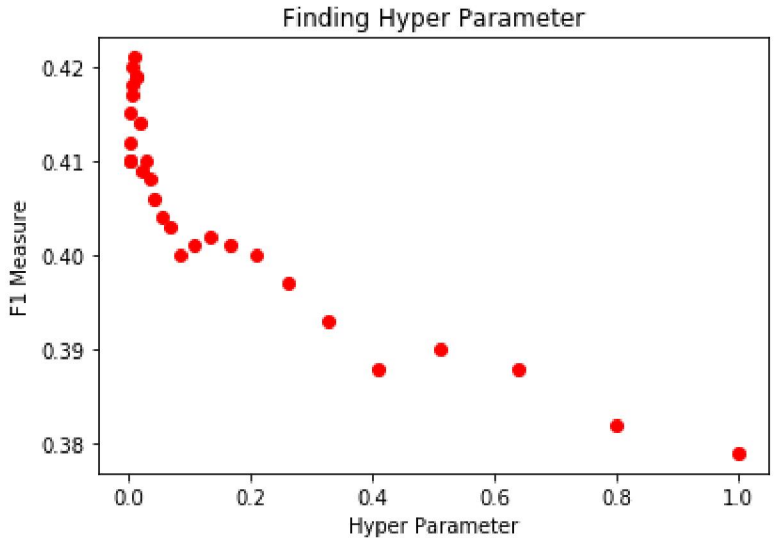
Majority Class F1 Measure = 0.351

(b) & (c) & (d)

Please check corresponding text files. For example, “Assignment3_260561054_2_c.txt”.

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.009223372036854787, it achieves a F1 measure of 0.421 on validation set. The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.009223372036854787
 F1 measure  = 0.421
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.7382857142857144

(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.421

(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.4455

4 Q2.a 4 / 4

✓ **- 0 pts Correct**

- 1 pts Majority class score low

- 4 pts not reported or misreported

- 1 pts unreported uniform classifier

- 1 pts misunderstanding

- 1 pts scores too high

Question 2

(a)

Uniform Random F1 Measure = 0.20350000000000001

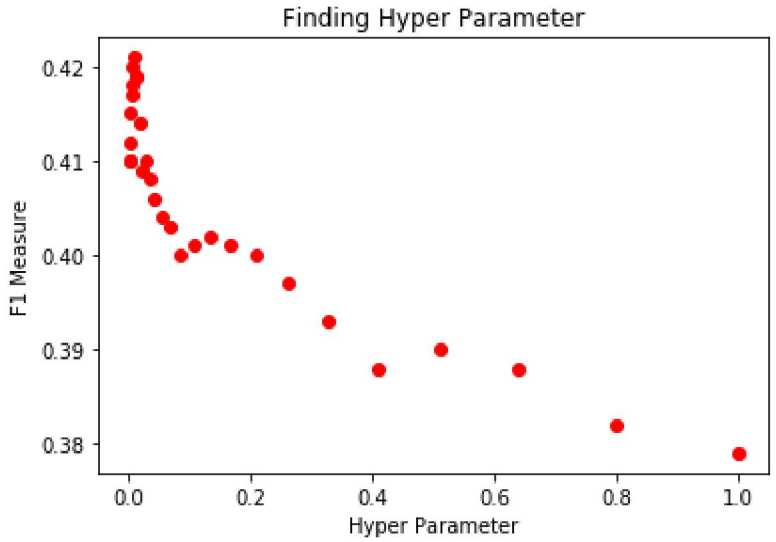
Majority Class F1 Measure = 0.351

(b) & (c) & (d)

Please check corresponding text files. For example, “Assignment3_260561054_2_c.txt”.

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.009223372036854787, it achieves a F1 measure of 0.421 on validation set. The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.009223372036854787
 F1 measure  = 0.421
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.7382857142857144

(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.421

(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.4455

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.418 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 10.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.46514285714285714
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.41999999999999993
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.418

Linear SVM

By varying C and dual, the best F1 measure 0.513 is achieved where C = 0.009223372036854787, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result. However, these results do not provide enough evidence to judge which algorithm is better because their F1 measures all remain in a low value range (below or around 0.5).

A simple guess is that the Yelp dataset does not have enough volume, and it has relatively large class number, 5 classes (5 ratings).

Because of this reason, I'll comment why Linear SVM is better in the IMDB section.

I expect that the F1 measures on IMDB set would be much better as it has a large amount of data volume and less classes.

5 Q2.b **10 / 10**

✓ **- 0 pts** **Correct**

- 10 pts nothing reported

Question 2

(a)

Uniform Random F1 Measure = 0.20350000000000001

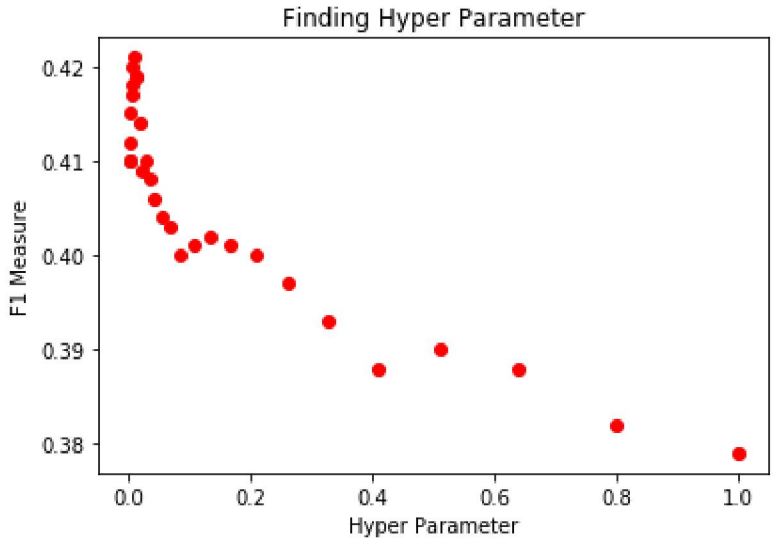
Majority Class F1 Measure = 0.351

(b) & (c) & (d)

Please check corresponding text files. For example, “Assignment3_260561054_2_c.txt”.

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.009223372036854787, it achieves a F1 measure of 0.421 on validation set. The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.009223372036854787
 F1 measure  = 0.421
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.7382857142857144

(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.421

(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.4455

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.418 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 10.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.46514285714285714
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.41999999999999993
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.418

Linear SVM

By varying C and dual, the best F1 measure 0.513 is achieved where C = 0.009223372036854787, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result. However, these results do not provide enough evidence to judge which algorithm is better because their F1 measures all remain in a low value range (below or around 0.5).

A simple guess is that the Yelp dataset does not have enough volume, and it has relatively large class number, 5 classes (5 ratings).

Because of this reason, I'll comment why Linear SVM is better in the IMDB section.

I expect that the F1 measures on IMDB set would be much better as it has a large amount of data volume and less classes.

6 Q2.c 5 / 5

✓ **- 0 pts Correct**

- 0.5 pts insufficient explanation for best values used

- 0.5 pts insufficent hyperparameters for SVM

- 1 pts No (or insufficient) Naive Bayes hyperparameters tuned

- 0.5 pts insufficient Decision Tree hyperparams

- 1.25 pts no Decision tree hyperparameters

- 1 pts insufficient hyperparamters for Decision tree

- 0.5 pts explain hyperparameters

- 1 pts insufficient explanation

- 2 pts insufficient explanation

- 1.25 pts no SVM hyperparameters

- 0.5 pts test more cases

- 5 pts incorrect or nothing reported

Question 2

(a)

Uniform Random F1 Measure = 0.20350000000000001

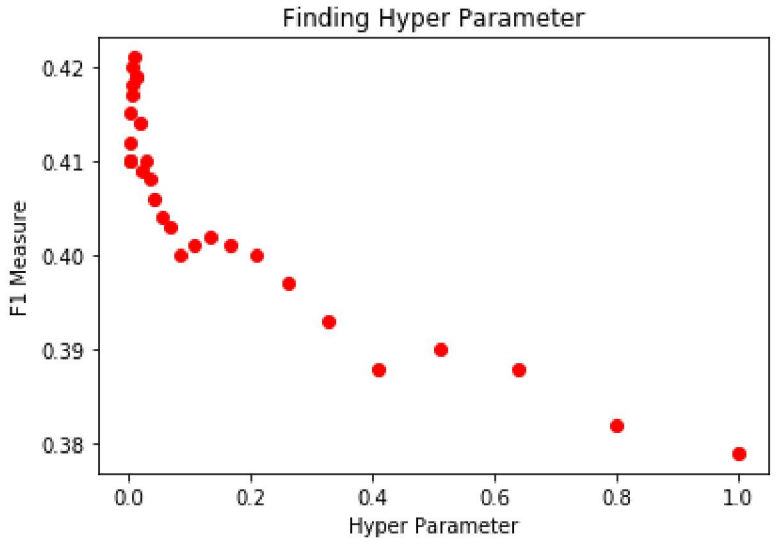
Majority Class F1 Measure = 0.351

(b) & (c) & (d)

Please check corresponding text files. For example, “Assignment3_260561054_2_c.txt”.

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.009223372036854787, it achieves a F1 measure of 0.421 on validation set. The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.009223372036854787
 F1 measure  = 0.421
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.7382857142857144

(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.421

(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.4455

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.418 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 10.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.46514285714285714
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.41999999999999993
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.418

Linear SVM

By varying C and dual, the best F1 measure 0.513 is achieved where C = 0.009223372036854787, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result. However, these results do not provide enough evidence to judge which algorithm is better because their F1 measures all remain in a low value range (below or around 0.5).

A simple guess is that the Yelp dataset does not have enough volume, and it has relatively large class number, 5 classes (5 ratings).

Because of this reason, I'll comment why Linear SVM is better in the IMDB section.

I expect that the F1 measures on IMDB set would be much better as it has a large amount of data volume and less classes.

7 Q2.d 6 / 6

✓ **- 0 pts Correct**

- 1 pts scores too high

- 1.5 pts scores too low

- 2.5 pts missing results

- 1 pts missing results

- 2 pts misunderstanding

- 5 pts incorrect results

- 1 pts incorrect results

- 0.5 pts poor presentation

- 0 pts Click here to replace this description.

- 1 pts misunderstanding

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.418 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 10.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.46514285714285714
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.41999999999999993
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.418

Linear SVM

By varying C and dual, the best F1 measure 0.513 is achieved where C = 0.009223372036854787, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result. However, these results do not provide enough evidence to judge which algorithm is better because their F1 measures all remain in a low value range (below or around 0.5).

A simple guess is that the Yelp dataset does not have enough volume, and it has relatively large class number, 5 classes (5 ratings).

Because of this reason, I'll comment why Linear SVM is better in the IMDB section.

I expect that the F1 measures on IMDB set would be much better as it has a large amount of data volume and less classes.

8 Q2.e 4 / 4

✓ **- 0 pts** **Correct**

- 1 pts further explanation needed

- 2 pts insufficient explanation

- 3 pts insufficient or no explanation provided

- 0.5 pts be more precise

- 4 pts no explanation provided

Question 3

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8014285714285714

(FBOW) Validation F1 Measure of Gaussian NB = 0.3

(FBOW) Testing F1 Measure of Gaussian NB = 0.312

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.408 is achieved when min_sample_split = 0. 0.006782230728489994 and max_depth = 8.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.48414285714285715

(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.408

(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.416

Linear SVM

By varying C and dual, the best F1 measure 0.508 is achieved where C = 12.839184645488634, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284

(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513

(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. Their F1 measures all remain in a low value range.

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

9 Q3.a **2 / 2**

✓ **- 0 pts** **Correct**

- 2 pts Click here to replace this description.

- 0 pts Click here to replace this description.

Question 3

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8014285714285714

(FBOW) Validation F1 Measure of Gaussian NB = 0.3

(FBOW) Testing F1 Measure of Gaussian NB = 0.312

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.408 is achieved when min_sample_split = 0. 0.006782230728489994 and max_depth = 8.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.48414285714285715

(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.408

(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.416

Linear SVM

By varying C and dual, the best F1 measure 0.508 is achieved where C = 12.839184645488634, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284

(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513

(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. Their F1 measures all remain in a low value range.

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

10 Q3.b **2 / 2**

✓ **- 0 pts** **Correct**

- 2 pts Click here to replace this description.

Question 3

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8014285714285714

(FBOW) Validation F1 Measure of Gaussian NB = 0.3

(FBOW) Testing F1 Measure of Gaussian NB = 0.312

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.408 is achieved when min_sample_split = 0. 0.006782230728489994 and max_depth = 8.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.48414285714285715

(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.408

(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.416

Linear SVM

By varying C and dual, the best F1 measure 0.508 is achieved where C = 12.839184645488634, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284

(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513

(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. Their F1 measures all remain in a low value range.

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

11 Q3.c 6 / 6

✓ **- 0 pts Correct**

- **1 pts** We're looking for the average F1
- **1 pts** Linear SVM F1 training F1 should be close to Valid F1
- **1 pts** Gaussian NB Validation F1 too high
- **1 pts** DecisionTree F1 Validation F1 too high
- **1 pts** SVM Validation F1 too high
- **2 pts** No validation F1 values
- **6 pts** 6
- **4 pts** F1 Scores too low

Question 3

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8014285714285714

(FBOW) Validation F1 Measure of Gaussian NB = 0.3

(FBOW) Testing F1 Measure of Gaussian NB = 0.312

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.408 is achieved when min_sample_split = 0. 0.006782230728489994 and max_depth = 8.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.48414285714285715

(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.408

(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.416

Linear SVM

By varying C and dual, the best F1 measure 0.508 is achieved where C = 12.839184645488634, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284

(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513

(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. Their F1 measures all remain in a low value range.

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

12 Q3.d **1.5 / 3**

- **0 pts** Correct

✓ - **1.5 pts** **Discussion about reason for classifier performance missing**

- **3 pts** Question missing

Question 3

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8014285714285714

(FBOW) Validation F1 Measure of Gaussian NB = 0.3

(FBOW) Testing F1 Measure of Gaussian NB = 0.312

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.408 is achieved when min_sample_split = 0. 0.006782230728489994 and max_depth = 8.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.48414285714285715

(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.408

(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.416

Linear SVM

By varying C and dual, the best F1 measure 0.508 is achieved where C = 12.839184645488634, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.8435714285714284

(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.513

(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.5085

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. Their F1 measures all remain in a low value range.

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

13 Q3.e 3 / 3

✓ **- 0 pts** **Correct**

- 1 pts No explanation of reason

- 1 pts Frequency information provides usefulness to representation

- 1.5 pts This question asks about FBOW and BBOW

- 3 pts Missing question

- 3 pts answer unclear

(f) For Yelp dataset, the performances of these two representations are similar. But ideally, I think the Frequency Bag of Representation should be worse because the most frequent words like “the”, “a” and “I” has little meaning, which does not contribute too much to the ratings. And those frequent words weighted a lot, although the classifier should notice this issue through training, considering the Yelp dataset does not have a large volume, the model is not mature enough.

14 Q3.f 2 / 2

✓ **- 0 pts** **Correct**

- 2 pts Question missing

- 1 pts Question is missing, but I understand your meaning.

- 1 pts This question refers to BBOW and FBOW

Question 4

(a)

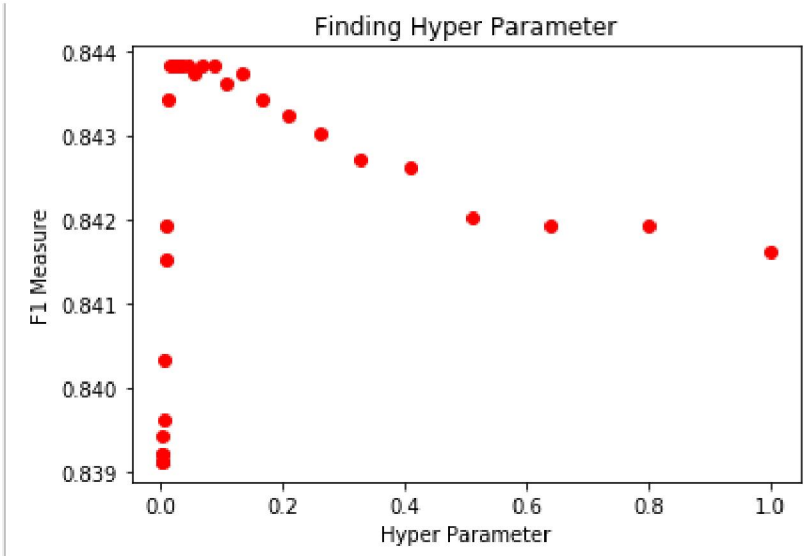
Uniform Random F1 Measure = 0.497523170344

(b) & (c) & (d)

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.08589934592000005, it achieves a F1 measure of 0.8438312337532493 on validation set.

The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.08589934592000005
 F1 measure  = 0.8438312337532493
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.8686929280810505
(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.8438312337532493
(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.8284925588094095

15 Q4.a **4 / 4**

✓ **- 0 pts** **Correct**

- 4 pts Performance not reported or submitted.

- 3 pts Incorrect performance. There are only 2, evenly distributed classes.

Question 4

(a)

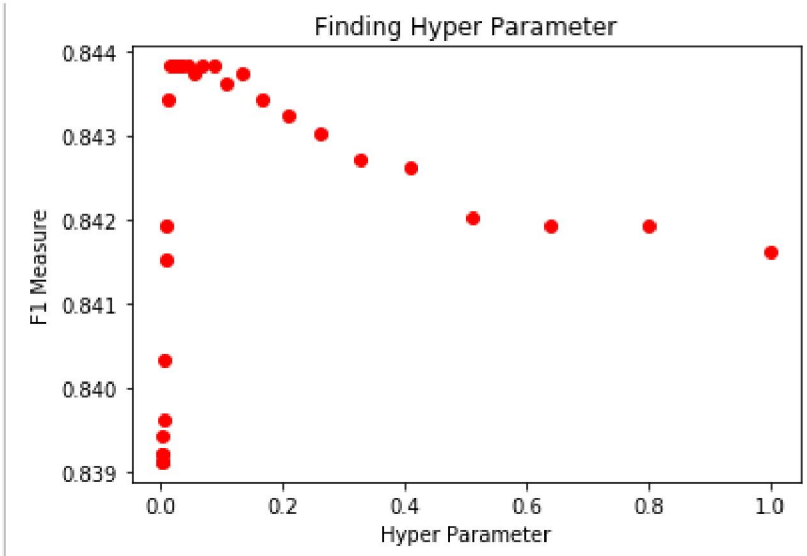
Uniform Random F1 Measure = 0.497523170344

(b) & (c) & (d)

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.08589934592000005, it achieves a F1 measure of 0.8438312337532493 on validation set.

The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.08589934592000005
 F1 measure  = 0.8438312337532493
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.8686929280810505
(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.8438312337532493
(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.8284925588094095

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0. 6959608078384323 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.7064587082583483
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6959608078384323
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7058329332693231

Linear SVM

By varying C and dual, the best F1 measure 0.8787242551489702 is achieved where C = 0.004747561509942996, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9470772512164234
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.8787242551489702
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.874739958393343

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result.

Linear SVM performs better because we can tune the penalty parameter and minimize the loss function. In addition, the IMDB dataset is large and has clear decision boundary which is an ideal dataset for SVM. Moreover, the reason that decision tree performs worse could be that I did not set correct constrains on hyper parameters as I'm not sure which one should be used relatively to others.

16 Q4.b **4 / 4**

✓ **- 0 pts** **Correct**

- 4 pts Not submitted.

Question 4

(a)

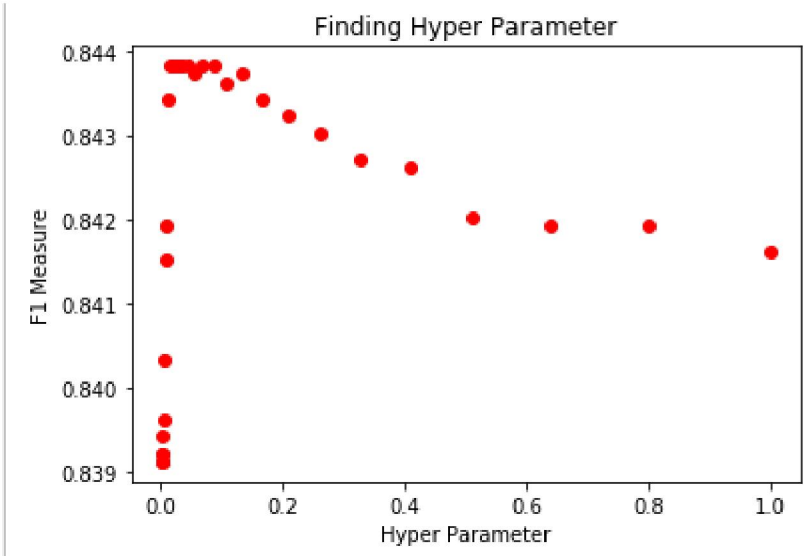
Uniform Random F1 Measure = 0.497523170344

(b) & (c) & (d)

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.08589934592000005, it achieves a F1 measure of 0.8438312337532493 on validation set.

The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.08589934592000005
 F1 measure  = 0.8438312337532493
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.8686929280810505
(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.8438312337532493
(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.8284925588094095

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0. 6959608078384323 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.7064587082583483
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6959608078384323
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7058329332693231

Linear SVM

By varying C and dual, the best F1 measure 0.8787242551489702 is achieved where C = 0.004747561509942996, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9470772512164234
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.8787242551489702
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.874739958393343

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result.

Linear SVM performs better because we can tune the penalty parameter and minimize the loss function. In addition, the IMDB dataset is large and has clear decision boundary which is an ideal dataset for SVM. Moreover, the reason that decision tree performs worse could be that I did not set correct constrains on hyper parameters as I'm not sure which one should be used relatively to others.

17 Q.4.c **2 / 2**

✓ **- 0 pts** **Correct**

- 2 pts Needed to report tested values and the best choice based on this dataset.

- 1 pts Needed to reported tested values.

- 1 pts Did not test hyper-parameters for all three methods.

- 2 pts Not submitted.

Question 4

(a)

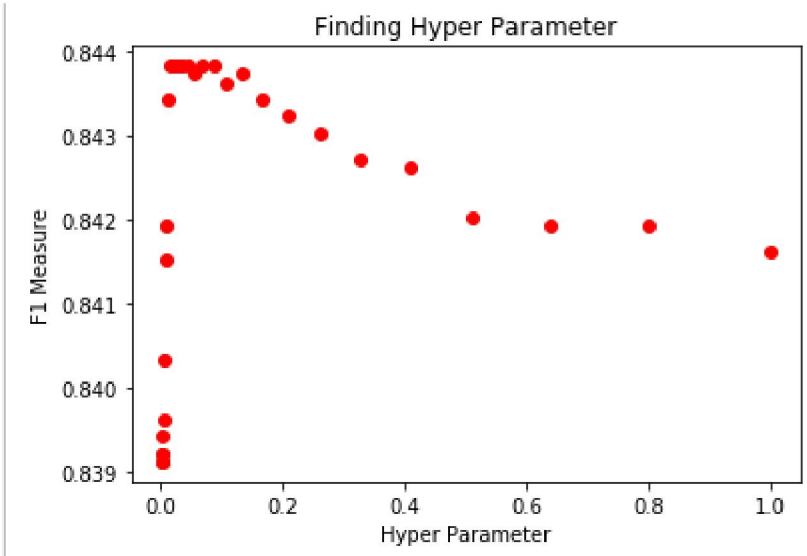
Uniform Random F1 Measure = 0.497523170344

(b) & (c) & (d)

Bernoulli Naive Bayes

By tuning the hyper parameter Laplace smoothing parameter Alpha, which ranges from 1 to 0.003777, we got the best Alpha = 0.08589934592000005, it achieves a F1 measure of 0.8438312337532493 on validation set.

The training process is shown in the figure below:




```
The best smoothing parameter alpha = 0.08589934592000005
 F1 measure  = 0.8438312337532493
```

Applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Bernoulli NB with best hyper-parameter = 0.8686929280810505
(BBOW) Validation F1 Measure of Bernoulli NB with best hyper-parameter = 0.8438312337532493
(BBOW) Testing F1 Measure of Bernoulli NB with best hyper-parameter = 0.8284925588094095

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0. 6959608078384323 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.7064587082583483
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6959608078384323
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7058329332693231

Linear SVM

By varying C and dual, the best F1 measure 0.8787242551489702 is achieved where C = 0.004747561509942996, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9470772512164234
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.8787242551489702
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.874739958393343

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result.

Linear SVM performs better because we can tune the penalty parameter and minimize the loss function. In addition, the IMDB dataset is large and has clear decision boundary which is an ideal dataset for SVM. Moreover, the reason that decision tree performs worse could be that I did not set correct constrains on hyper parameters as I'm not sure which one should be used relatively to others.

18 Q.4.d 6 / 6

✓ **- 0 pts** **Correct**

- 2 pts Poor/unrealistic results for Naive Bayes.

- 2 pts Poor/unrealistic results for Decision Trees.

- 2 pts Poor/unrealistic results for SVM.

- 5 pts Poor/unrealistic results on all models.

- 4 pts Needed to report results for all models.

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0. 6959608078384323 is achieved when min_sample_split = 0.04035360699999998 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.7064587082583483
(BBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6959608078384323
(BBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7058329332693231

Linear SVM

By varying C and dual, the best F1 measure 0.8787242551489702 is achieved where C = 0.004747561509942996, dual = False. Where C ranges from 1 to 0.0037777 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(BBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9470772512164234
(BBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.8787242551489702
(BBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.874739958393343

(e)

By comparing the test results of these three classifiers, we observe that Linear SVM gives the best result on Binary Bag of Word representation of the yelp dataset. And Decision Tree has the worst result.

Linear SVM performs better because we can tune the penalty parameter and minimize the loss function. In addition, the IMDB dataset is large and has clear decision boundary which is an ideal dataset for SVM. Moreover, the reason that decision tree performs worse could be that I did not set correct constrains on hyper parameters as I'm not sure which one should be used relatively to others.

19 Q4.e **2 / 3**

- **0 pts** Correct
- **1 pts** Should justify why the best classifier performed the best.

✓ **- 1 pts** **Partially incorrect or incomplete justification/explanation.**

- **2 pts** Missing explanation.
- **3 pts** Incorrect or not submitted.
- **1 pts** Missing discussion on hyper-parameters.

Question 5

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set, and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8565620209291475
(FBOW) Validation F1 Measure of Gaussian NB = 0.7521495700859829
(FBOW) Testing F1 Measure of Gaussian NB = 0.681148983837414

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.6993601279744052 is achieved when min_sample_split = 0.49 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.706125441578351
(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6993601279744052
(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7032325172027525

Linear SVM

By varying C and dual, the best F1 measure 0.876124775044991 is achieved where C = 31.947999937062274, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9186829300806505
(FBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.876124775044991
(FBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.8739398303728596

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. The reason why SVM is better and Decision Tree is worse is explained in Question 4. (e).

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. (Which is similar to the Yelp dataset). The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

20 Q5.a **2 / 2**

✓ **- 0 pts** **Correct**

- 2 pts Not done.

Question 5

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set, and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8565620209291475
(FBOW) Validation F1 Measure of Gaussian NB = 0.7521495700859829
(FBOW) Testing F1 Measure of Gaussian NB = 0.681148983837414

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.6993601279744052 is achieved when min_sample_split = 0.49 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.706125441578351
(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6993601279744052
(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7032325172027525

Linear SVM

By varying C and dual, the best F1 measure 0.876124775044991 is achieved where C = 31.947999937062274, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9186829300806505
(FBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.876124775044991
(FBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.8739398303728596

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. The reason why SVM is better and Decision Tree is worse is explained in Question 4. (e).

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. (Which is similar to the Yelp dataset). The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

21 Q5.b **1.8 / 2**

- **0 pts** Correct

✓ - **0.2 pts** **Training data used for validation / One hyper parameter missing**

- **0.4 pts** Click here to replace this description.
- **0.6 pts** Click here to replace this description.
- **0.8 pts** Click here to replace this description.
- **2 pts** cannot find answer to the question

Missing Parameters for Decision trees:

- Criterion

Question 5

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set, and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8565620209291475
(FBOW) Validation F1 Measure of Gaussian NB = 0.7521495700859829
(FBOW) Testing F1 Measure of Gaussian NB = 0.681148983837414

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.6993601279744052 is achieved when min_sample_split = 0.49 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.706125441578351
(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6993601279744052
(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7032325172027525

Linear SVM

By varying C and dual, the best F1 measure 0.876124775044991 is achieved where C = 31.947999937062274, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9186829300806505
(FBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.876124775044991
(FBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.8739398303728596

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. The reason why SVM is better and Decision Tree is worse is explained in Question 4. (e).

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. (Which is similar to the Yelp dataset). The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

22 Q5.c 6 / 6

✓ **- 0 pts** **Correct**

- 2 pts error values of one of the classifiers

- 4 pts error in score of 2 models

- 6 pts No answer found or all scores are incorrect

- 0.66 pts Error in a single value

Question 5

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set, and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8565620209291475
(FBOW) Validation F1 Measure of Gaussian NB = 0.7521495700859829
(FBOW) Testing F1 Measure of Gaussian NB = 0.681148983837414

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.6993601279744052 is achieved when min_sample_split = 0.49 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.706125441578351
(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6993601279744052
(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7032325172027525

Linear SVM

By varying C and dual, the best F1 measure 0.876124775044991 is achieved where C = 31.947999937062274, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9186829300806505
(FBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.876124775044991
(FBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.8739398303728596

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. The reason why SVM is better and Decision Tree is worse is explained in Question 4. (e).

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. (Which is similar to the Yelp dataset). The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

23 Q5.d 2 / 2

✓ **- 0 pts Correct**

- 0.5 pts Partially correct

- 2 pts Incorrect or no answer found.

- 0.25 pts Mostly correct with minor errors.

Question 5

(a) & (b) & (c)

Gaussian Native Bayes

There is no hyper parameters need to be tuned here. The F1 measures on training set, validation set, and testing set are shown below,

(FBOW) Training F1 Measure of Gaussian NB = 0.8565620209291475
(FBOW) Validation F1 Measure of Gaussian NB = 0.7521495700859829
(FBOW) Testing F1 Measure of Gaussian NB = 0.681148983837414

Decision Tree

I chose hyper parameters min_sample_split and max_depth. For min_sample_split, it ranges from 1 to 0.004747561 and for max_depth it ranges from 1 to 15. The max_depth hyper parameter was tuned to prevent overfitting because if we set it to default, the training F1 measure would be 1, validation and testing F1 measure would be below 0.5.

The best F1 measure 0.6993601279744052 is achieved when min_sample_split = 0.49 and max_depth = 6.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Decision Tree with best hyper-parameter = 0.706125441578351
(FBOW) Validation F1 Measure of Decision Tree with best hyper-parameter = 0.6993601279744052
(FBOW) Testing F1 Measure of Decision Tree with best hyper-parameter = 0.7032325172027525

Linear SVM

By varying C and dual, the best F1 measure 0.876124775044991 is achieved where C = 31.947999937062274, dual = False. Where C ranges from 1 to 237.3763138 and dual is a Boolean.

By applying the best hyper parameters on training set, validation set and testing set, we got:

(FBOW) Training F1 Measure of Linear SVM with best hyper-parameter = 0.9186829300806505
(FBOW) Validation F1 Measure of Linear SVM with best hyper-parameter = 0.876124775044991
(FBOW) Testing F1 Measure of Linear SVM with best hyper-parameter = 0.8739398303728596

(d)

The performance result is similar to the Binary Bag of Word representation where Linear SVM has the best F1 measures. But Gaussian Naïve Bayer has the worst F1 measures. The reason why SVM is better and Decision Tree is worse is explained in Question 4. (e).

(e)

For these two representations, the performances of Linear SVM and Decision Tree remain stable, however, there is a big drop in performance of Naïve Bayes. (Which is similar to the Yelp dataset). The Gaussian Naïve Bayer has much lower F1 measure than Bernoulli Naïve Bayer. The reasons could be that the probability distribution is not a Gaussian distribution, and there are no hyper parameters to be adjusted for Gaussian Naïve Bayer.

24 Q5.e 2 / 3

- **0 pts** Correct
- **0.25 pts** Largely correct with minor errors like lack of clarity/incorrect assumptions
- **2 pts** Incomplete answer
- **0.5 pts** Correct with errors like lack of clarity and or incorrect assumptions with incorrect inference.
- **1.5 pts** Partially correct answer
- **0.1 pts** Minor errors

✓ **- 1 pts Partially correct answer.**

- **3 pts** No answer or incorrect answer.

The reduction in performance is due to more importance given to frequently occurring stop words like 'a', 'the', etc which do not convey any useful information.

(f)

For IMDB dataset, the performances of these two representations are still similar. The reason could be that in Frequency Bag of Words, the most frequent words like “the”, “a” and “I” has little meaning, which does not contribute too much to the ratings. Therefore, the frequency of each word is not important anymore. The number itself does not say anything except it only shows that this word is present in the current review, which is really similar to a binary representation. Therefore, the Frequency Bag of Word gradually degenerates to Binary Bag of Word, resulting in similar performance at the end.

(g)

Obviously the IMDB dataset has much better performance, with all F1 measures above 0.7 comparing to Yelp dataset where all F1 measures are below 0.5. This behavior is expected because Yelp dataset has smaller data volume and more classes. Instead, IMDB dataset has only two classes and it has sufficient dataset to train the model, thus leading to better performance.

25 Q5.f **2 / 2**

✓ **- 0 pts Correct**

- 0.25 pts Largely correct but with minor errors (lack of clarity, incorrect assumptions)

- 0.5 pts Correct with errors in certain parts (lack of clarity, incorrect inference)

- 1 pts Partially correct answer with incorrect assumptions or inference

- 2 pts Incorrect or no answer found.

(f)

For IMDB dataset, the performances of these two representations are still similar. The reason could be that in Frequency Bag of Words, the most frequent words like “the”, “a” and “I” has little meaning, which does not contribute too much to the ratings. Therefore, the frequency of each word is not important anymore. The number itself does not say anything except it only shows that this word is present in the current review, which is really similar to a binary representation. Therefore, the Frequency Bag of Word gradually degenerates to Binary Bag of Word, resulting in similar performance at the end.

(g)

Obviously the IMDB dataset has much better performance, with all F1 measures above 0.7 comparing to Yelp dataset where all F1 measures are below 0.5. This behavior is expected because Yelp dataset has smaller data volume and more classes. Instead, IMDB dataset has only two classes and it has sufficient dataset to train the model, thus leading to better performance.

26 Q5.g 4 / 4

✓ **- 0 pts Correct**

- 4 pts No Answer Found

- 1 pts Partially correct answer

- 0.5 pts Partially correct answer

- 2 pts Partially correct answer