CS505 HW 2 Submission template

Collaboration statement: I collaborated with Minh Nguyen
Link to Code: https://drive.google.com/drive/my-drive

Problem 1

1.a. (3 points) List the parameters here:
We need to learn the mean and standard deviation of each Xi, under each class, plus the two priors for the classes

1.b. (2 points) The total number of parameters needed in terms of n is:
4n + 2

Two pairs of mean and std for each X for two different classes, plus the two priors

2. a. (1 point) Write down NB formulation of P(Y|X) using Summer and Rowdy:

$P_{NB}(Y|X1,X2) = P(X1, X2, Y) / (P(X1, X2, Y) + P(X1, X2, \sim Y))$

(1 point) How do you decide using P(Y|X) if a review is positive or negative?
If this probability is >.5, decide class Y. If not, decide class ~Y.

2. b. (8 points) Fill up this table below with the probabilities (add the rows yourself to fill up all the combinations), both joint and Naive Bayes' probability

X1 (Summer)	X2 (Rowdy)	Y	P(X1, X2, Y)	$P_{NB}(Y\|X1,X2)$	NB Decision
T	T	T	.162	.59	T
T	T	F	.112	.41	T
T	F	T	.378	.89	T
T	F	F	.048	.11	T
F	T	T	.018	.1	F
F	T	F	.168	.9	F
F	F	T	.042	.37	F
F	F	F	.072	.63	F

(2 points) From the table above, the probability of observations where the label is different than predicted (i.e., the expected error rate of the NB classifier) equates to: …

.112 + .048 + .018 + .042 = .22

2. c. (2 points) What is P(Sentiment=1, Summer=1, Rowdy=1): (show how you derive the probability) …

P(Sentiment = 1, Summer = 1, Rowdy = 1) =
P(Summer = 1 | Sentiment = 1) * P(Rowdy = 1 | Sentiment = 1) * P(Sentiment = 1) =
.9 * .3 * .6 = .162

2. d. (1 point) Are any of the NB assumptions violated? Yes

(1 point) Why? Winter and Summer are not independent.

(1 point) What is P(Sentiment=0, Summer=0, Rowdy=1, Winter=1): (show how you derive the probability): …

P(Sentiment = 0, Summer = 0, Rowdy = 1, Winter = 1) =
P(Summer = 0 | Sentiment = 0) * P(Rowdy = 1 | Sentiment = 0) * 1 * P(Sentiment = 0) =
.6 * .7 * 1 * .4 = .168

2. e. (6 points) Fill up this table below with the probabilities (add the rows yourself to fill up all the combinations), both joint and Naive Bayes' probability

X1 (Summer)	X2 (Rowdy)	X3 (Winter)	Y	P(X1,X2, X3,Y)	P_{NB}(Y,X1,X2,X3)	P_{NB}(Y|X1,X2, X3)	NB Decision
T	T	T	T	0	.146	.76	T
T	T	T	F	0	.045	.24	T
T	T	F	T	.162	.016	.19	F
T	T	F	F	.112	.067	.81	F
T	F	T	T	0	.34	.95	T
T	F	T	F	0	.019	.05	T
T	F	F	T	.378	.038	.57	T

T	F	F	F	.048	.029	.43	T
F	T	T	T	.018	.016	.19	F
F	T	T	F	.168	.067	.81	F
F	T	F	T	0	.0018	.02	F
F	T	F	F	0	.1	.98	F
F	F	T	T	.042	.034	.54	T
F	F	T	F	.072	.029	.46	T
F	F	F	T	0	.004	.09	F
F	F	F	F	0	.043	.91	F

(2 points) From the table above, the probability of observations where the label is different than predicted (i.e., the expected error rate of the NB classifier) equates to:

0 + .162 + 0 + .048 + .018 + 0 + .072 = .3

2. f. (3 points) Does the performance of your NB classifier improve with this addition of the new feature "Winter"? Explain your answer: No, the expected error increased. This is likely due to the fact that the new inputs violate the NB assumptions.

3. a. (2 points) What will happen when your NB classifier predicts the probability of a test instance with mention of cleanliness? Explain why this is undesirable: Since the model will learn that the probability of seeing cleanliness is zero, the NB formulation on the test instance will have a numerator and denominator of zero, which ruins the NB's prediction.

3. b. (5 points) Will logistic regression have a similar problem? Explain **concretely** why/why not? …

With logistic regression, the size of each input to both train and test has to be the same. (If the model learns weights for 3 features, it can only make predictions on three features) Therefore, if cleanliness was not encountered in the training set, any occurrences of it will be removed from the testing set before making predictions, so logistic regression would not have the same problem.

Problem 2.

1. (5 points) Accuracy on the test set is: 73%

2. (2 points) Accuracy on the test set is: 74%

   (1 point) Does using binary counts as features improve the accuracy? Yes

3. (4 points) Accuracy on the test set is: 77%

   (1 point) Which classifier performs better on this test set? Logistic Regression

4. (2 points) Accuracy on the test set is: .76%

5. (1 point) What combination of feature extraction (binary/non-binary) and statistical model (NB/LR) is good for this dataset? Non-binary, LR

   (1 point) Is evaluating on a test set a good way to do a model selection? Why/Why not? No, you could end up overfitting to the test set

6. a. (8 points) Fill up the table below with your average accuracies across 10-folds:

Max features	Binary (True/False)	Average Accuracy
1000	T	74%
1000	F	73.8%
2000	T	75.1%
2000	F	74.9%
3000	T	75.4%
3000	F	75.2%
4000	T	75.5%
4000	F	75.3%

b. (2 points) Accuracy on the test set (when training on the whole train data with the best combination of max_features and binary/non-binary count based off cross-validation) is:

78%

7. c. (6 points) Accuracy on the test set is: 65%

   (1 point) Does dense feature representation computed this way improve the accuracy of your classifier? No

8. a. (3 points) What is the model you choose and why is it appropriate for text features (briefly compare to logistic regression):

   We could use a neural network in a similar way to how we used logistic regression. Vectors can be constructed in the same way as in logistic regression and then given to a neural network. It will take longer to train a neural network but it is theoretically capable of learning more intricate connections between features than logistic regression alone.

   b. (3 points) Describe at least one method to address class imbalance in one of the methods you have considered so far:

   In logistic regression class imbalance can be addressed with class weights, which means modifying the loss function to more heavily penalize incorrect classifications where the true label is the minority class.

9. (Bonus: 6 points) Fill up the table below with your average accuracies across 10-folds:

Max features	Binary (True/False)	Average Accuracy
1000	T	74.6%
1000	F	74.5%
2000	T	76%
2000	F	75.8%
3000	T	76.4%
3000	F	76.3%

4000	T	76.7%
4000	F	76.6%

(Bonus: 2 points) Accuracy on the test set (when training on the whole train data with the best combination of max_features and binary/non-binary count based off cross-validation) is: 79%

(Bonus: 2 points) Does having a huge amount of training data allow a simple classifier such as NB classifier with bag-of-words features to perform even better on the test set? Yes, there was a slight increase in accuracy when using the larger training set. 79% is the best accuracy that was achieved by any model so far.

Problem 3.

1. a. (1 point) What do you think each of the 1000 features you created represent? I.e., what do they correspond to?

They correspond to words that were seen in the training data such as “good”, “bad”

1. b. (3 points) Examine the coefficients that correspond to the word “good” and “bad”. How do they compare to each other and in which way do they contribute to the prediction?

Good: .84
Bad: -.99

The presence of the word bad contributes more to a negative classification while the presences of good contributes to a positive classification

2. a. (2 points) Explain gradient tree boosting here: …

Gradient tree boosting involves an ensemble of weak prediction models such as decision trees. For predictions, the average prediction of the ensemble models is taken.

2. b. The model that I choose is (catboost/xgboost/lightgbm): catboost

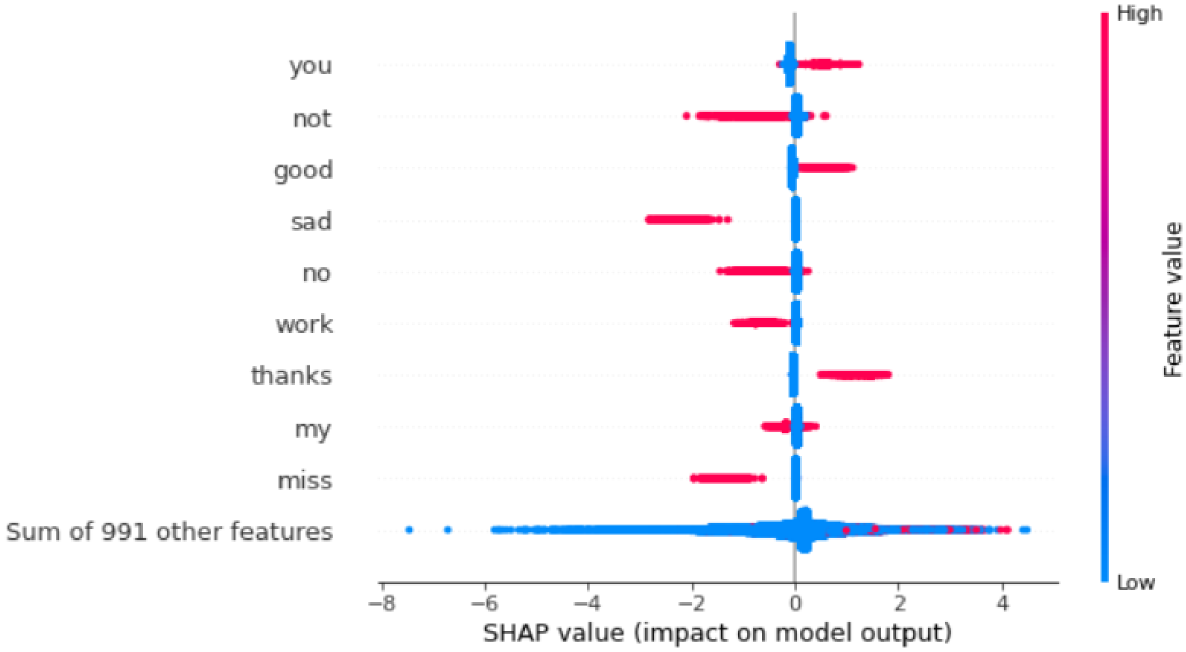
(2 points) Accuracy on train is: 79%

(2 points) Accuracy on test is: 77%

2. c. (3 points) Explain SHAP values here: …

Shap values indicate the contribution of each feature's weight to the overall prediction of the model.

2. d. (1 point) Put your plot below:



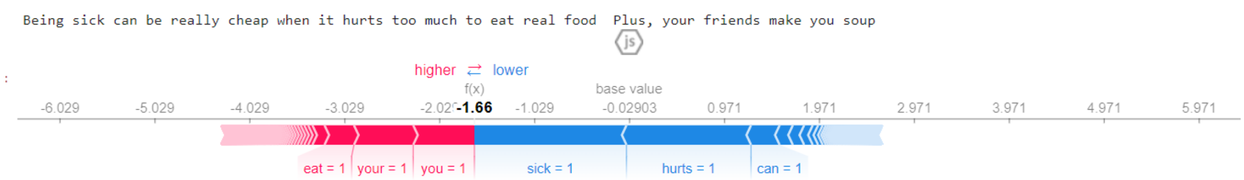
(1 point) describe what axis/colors/legend/and dots mean: …

The blue is the original value of a parameter's weight, so they are all initially 0. The red dots are the shap values of each feature for different iterations, and the clusters tell you the rough impact of a feature on the prediction.

(2 points) Give examples to words that lead to positive and negative sentiment predictions: …

Sad leads to negative predictions,
Thanks leads to positive

2. e. (3 points) Using shap.plots.text, examine one interesting example, print the actual tweet and discuss what the plot tells you and significant words that led to the score:



The plot shows that eat, your, and you were words contributing in a positive direction, and that sick, hurts and can were words contributing in a negative direction.

What does the bold number indicate?

This tweet received a score of -1.66, a negative tweet.