Natural Language Processing
Lab 3.1

This lab sheet is to practice the concepts taught this week so far: Naive Bayes classification and Sentiment Analysis.

1. Train two models, multinomial naive Bayes and binarized naive Bayes, both with add-1 smoothing, on the following document counts for key sentiment words, with positive or negative class assigned as noted.

doc	"good"	"poor"	"great"	(class)
d1.	3	0	3	pos
d2.	0	1	2	pos
d3.	1	3	0	neg
d4.	1	5	2	neg
d5.	0	2	0	neg

   Use both naive Bayes models to assign a class (pos or neg) to this sentence:

   A good, good plot and great characters, but poor acting.

   Do the two models agree or disagree?

Ans: See here for potential solutions:

Multinomial Naive Bayes:

Class prior
$logprior[positive] = \log(2/5) = -0.3979$
$logprior[negative] = \log(3/5) = -0.2218$

loglikelihoods:
$loglikelihood[good, pos] = \log \frac{3+1}{9+3} = -0.4771$
$loglikelihood[good, neg] = \log \frac{2+1}{14+3} = -0.7533$
$loglikelihood[poor, pos] = \log \frac{1+1}{9+3} = -0.7782$
$loglikelihood[poor, neg] = \log \frac{10+1}{14+3} = -0.1891$
$loglikelihood[great, pos] = \log \frac{5+1}{9+3} = -0.3010$
$loglikelihood[great, neg] = \log \frac{2+1}{14+3} = -0.7533$

Next we can compute classifier output

$sum[pos] = -0.3979 - 0.4771 \times 2 - 0.7782 - 0.3010 = -2.4313$
$sum[neg] = -0.2218 - 0.7533 \times 2 - 0.1891 - 0.7533 = -2.6708$

Because $sum[pos] > sum[neg]$, we assert it should be classified as pos.

Binarized Naive Bayes:
loglikelihoods:
$loglikelihood[good, pos] = \log \frac{1+1}{4+3} = - - 0.5441$
$loglikelihood[good, neg] = \log \frac{2+1}{6+3} = -0.4771$
$loglikelihood[poor, pos] = \log \frac{1+1}{4+3} = -0.5441$
$loglikelihood[poor, neg] = \log \frac{3+1}{6+3} = -0.3522$
$loglikelihood[great, pos] = \log \frac{2+1}{4+3} = -0.3680$
$loglikelihood[great, neg] = \log \frac{1+1}{6+3} = -0.6532$

$sum[pos] = -0.3979 - 0.5441 - 0.5441 - 0.3680 = -1.8541$
$sum[neg] = -0.2218 - 0.4771 - 0.3522 - 0.6532 = -1.7043$

Because $sum[pos] < sum[neg]$, we assert it should be classified as neg.

They disagree with each other.

2. Consider an NLP classification task where a model is trained to classify text documents into two categories: pos and neg. After testing the model on a validation set, the following confusion matrix was obtained:

Actual / Predicted	pos	neg
pos	80	20
neg	30	70

Based on the confusion matrix provided above, calculate the accuracy, and the precision, recall and F1 for the pos class. Provide your answers in decimal form rounded to two decimal places

Ans:

- Accuracy: 0.75

- Precision (for Positive class): 0.73
- Recall (for Positive class): 0.80
- F1 Score (for Positive class): 0.76

3. In another NLP task, a model is developed to classify text documents into three categories: Positive (pos), Neutral (neut), and Negative (neg). After deploying the model on a test dataset, the following confusion matrix was recorded:

Actual / Predicted	pos	neut	neg
pos	100	20	10
neut	330	120	20
neg	15	25	95

Using the confusion matrix above, calculate the following metrics: accuracy, and precision, recall, F1 for each class. Then calculate the micro-average precision, recall and F1. Finally calculate the macro-average precision, recall and F1 for the classifier. Provide your answers in decimal form rounded to two decimal places.

Ans: Accuracy: 0.43

Class-wise Metrics:
Positive Class: Precision: 0.22 Recall: 0.76 F1 Score: 0.34
Neutral Class: Precision: 0.73 Recall: 0.26 F1 Score: 0.38
Negative Class: Precision: 0.76 Recall: 0.70 F1 Score: 0.73

Macro-averages: Precision: 0.57 Recall: 0.57 F1 Score: 0.57

Micro-averages: Precision: 0.75 Recall: 0.61 F1 Score: 0.67

4. Briefly explain what the bootstrap method is and why it is appropriate for comparing two classifiers.

Ans: The bootstrap method is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement. It's appropriate for comparing classifiers because it allows us to assess the variability of the performance metric (like accuracy or F1 score) and determine if differences in performance are statistically significant, even with a limited sample size.

5. Describe the steps you would take to apply the bootstrap method to compare Model A and Model B.

Ans:

- Step 1: Randomly sample, with replacement, the same number of instances as in the original test dataset from the test dataset.
- Step 2: Evaluate both Model A and Model B on this resampled dataset and record their performance (e.g., accuracy).
- Step 3: Repeat steps 1 and 2 a large number of times (e.g., 1000 times) to create a distribution of performance measures for each model.
- Step 4: Compare the performance distributions of the two models to assess which model performs better and whether the difference is statistically significant.