1.1a Probability of Owning a Computer (OC)
P(OC)= 0.8,Owns a computer, 0.2 Does not own a computer

1.2a Probability of a Computer-Related Purchase (CRP | OC)
P(CRP | OC)= 0.1, owns a computer, 0.01, dont own

1.3a Probability of Fraud (FT | CT)
P(FT | CT)= 0.01,Cardholder is traveling, 0.004,Cardholder is not traveling

1.4a Probability of a Foreign Purchase (FP | FT, CT)
P(FP | FT,CT) = 0.9 Fraudulent and Traveling
0.1 Fraudulent and Not Traveling
0.9 Legitimate and Traveling
0.01 Legitimate and Not Traveling

1.5a Probability of an Internet Purchase (IP | FT, OC)
P(IP | FT,OC)= 0.15, Fraudulent and Owns a Computer
0.051, Fraudulent and Does Not Own a Computer
0.10, Legitimate and Owns a Computer
0.001,Legitimate and Does Not Own a Computer

1.6a Probability of Traveling (CT)
P(CT)= 0.05,Traveling
0.95,Not Traveling

1.1b What is the prior probability (i.e., before we search for previous computer related purchases and before we verify whether it is a foreign and/or an internet purchase) that the current transaction is a fraud?
Conversion: $P(FT) = P(FT \mid CT) \cdot P(CT) + P(FT \mid \text{Not } CT) \cdot P(\text{Not } CT)$
$P(FT) = (0.01)(0.05)+(0.004)(0.95)$
Value obtained from:
$P(FT \mid CT)$ from 1.3a
$P(CT)$ from 1.6a
$P(FT \mid \text{Not } CT)$ from 1.3a
$P(\text{Not } CT)$ from 1.6a
1.2b What is the probability that the current transaction is a fraud once we have verified that it is a foreign transaction, but not an internet purchase and that the card holder purchased computer related accessories in the past week?
Conversion:What is P(FT = | FP, Not IP, CRP)?
$P(FT \mid FP,\text{Not IP},CRP) = [P(FP,\text{Not IP},CRP \mid FT) \cdot P(FT)] / P(FP,\text{Not IP},CRP)$
$P(FP \mid FT,CT)P(\text{Not IP} \mid FT,OC)P(CRP \mid OC)P(OC)P(CT)P(FT \mid CT)] /$
$P(FP \mid FT,CT)P(IP \mid FT,OC)P(CRP \mid OC)P(OC)P(CT)P(FT \mid CT)]$

Values obtained from:
P(FP | FT,CT) from 1.4a
P(Not IP | FT,OC) from 1.5a
P(CRP | OC) from 1.3a
P(OC) from 1.1a
P(CT) from 1.6a
P(FT | CT) from 1.3a
P(IP | FT,OC) from 1.5a

1.3c After computing those probabilities, the fraud detection system raises a flag and recommends that the card holder be called to confirm the transaction. An agent calls at the domicile of the card holder but she is not home. Her spouse confirms that she is currently out of town on a business trip. How does the probability of a fraud change based on this new piece of information? Follow the same instructions as for B

Bayes rule:
P(FT | FP,Not IP,CRP,CT)= P(FT,CT,FP,Not IP,CRP)/P(FP | FT,CT)P(IP | FT,OC)P(CRP | OC)P(OC)P(CT)P(FT | CT)]

The numerator for the calculations changed, but the denominator remains the same

2. Yes it is good features because there is a high classification ratio which is what we sought to do as shown by the output, it was pretty clear different in classifying which the model was able to do from the high accuracy and testing accuracy, and since these are uncommon words that doesn't appear that often, it makes them significant features to consider. For example, when a word like graphic appears, there's a high probability based on our model accuracy that it will appear in the comp.graphics newspaper article which got converted into binary files.

```
Naive Bayes - Train Accuracy: 0.99, Test Accuracy: 0.96

Top 10 Discriminative Words:
graphics: 21.2663
atheism: 20.6133
religion: 20.5656
moral: 20.4897
evidence: 20.4897
atheists: 20.4630
bible: 20.3784
christian: 20.3486
religious: 20.3486
islam: 20.3178
```

3.

Hidden Size	Optimizer	Epochs	Train Acc (%)	Test Acc (%)
100	adam	50	94.59	72.44
100	adam	100	98.46	72.75
100	sgd	50	93.13	72.06
100	sgd	100	95.07	71.37
200	adam	50	98.40	74.12
200	adam	100	98.85	74.29
200	sgd	50	96.41	72.24
200	sgd	100	100.00	74.55

When the hidden sizes are 200, the test accuracy is better than 100 while when comparing epochs, the 100 epoche is usually better than 50 epoche in training accuracy with a 5% difference. Lastly, with the optimizer, adam performs better at 50 epochs while sgd performs better than adam while when they are compared with 100 epochs, This shows that sgd has better scaling while adam has better initial performance.

4.(Bonus)

```
C:\Users\danvs\anaconda3\lib\site-packages\sklearn\clust
ble OMP_NUM_THREADS=4
  warnings.warn(
KMeans Accuracy: 75.96%
MiniBatchKMeans Accuracy: 79.41%
```

For clustering, I choose to use K means and MiniBatchKMeans because they are good at dealing with large datasets like the ones provided to me. I didn't realize it was a giant dataset until I tried the other clustering methods such as aggro clustering which slowed my pc down by a lot. With the results of the accuracy shown.