CS440/ECE448 Fall 2016

Artificial intelligence
Assignment 3:
Naive Bayes Classification

Work Distribution:
Part 1: Yifei Li
Part 2: Bangqi Wang
13th Nov, 2016

Part 1: Digit classification

The goal of this assignment is to implement Naive Bayes classifiers and to apply it to the task of classifying digit and face data.

General Implementation

The basic idea of applying image classification is to first calculate likelihood and prior probability on the training set and then calculate posteriori probability according to the previous step. There are 28*28 pixels in each training image and '+', '#' indicates it is foreground and ' ' denotes background. For this part, we don't distinguish between two foreground value but will differentiate them later.

In step 1, there are two probabilities that needs to be calculated, Prior and likelihood. The likelihood is defined as **$P(F_{ij}$ | class)** for every pixel location (i, j) and for every digit class. More specifically:

$P(F_{ij} = f$ | class) = (# of times pixel (i, j) has value f in training examples from this class) / (Total # of training examples from this class)

Note that in order to dealing with feature that were never seen or seen too few times, we experiment with few laplace smoothing constant for best accuracy. Finally, the constant k works best when it is equal to 1.

The prior probability is simply obtained by count frequencies of each digit in the example set.

In step 2, now we should use the likelihood and prior we have to predict the class of testing image. This is achieved by performing **maximum a posteriori (MAP)** classification of test images. We calculate:

$$P(\text{class}) \cdot P(f_{1,1} \mid \text{class}) \cdot P(f_{1,2} \mid \text{class}) \cdot \ldots \cdot P(f_{28,28} \mid \text{class})$$

In order to avoid underflow, the log version should be used:

log P(class) + log P($f_{1,1}$ | class) + log P($f_{1,2}$ | class) + ... + log P($f_{28,28}$ | class)
(Note we need to calculate the posteriori probability for class 0-9)

After calculating probability for all classes, we pick the class with largest probability and use it as our prediction.

Classification results

Our model achieves 77.1% overall accuracy on all digit classes. The classification result for each digit is reported as follows:

```
Overall: 0.771
Classification rate for digit 0: 0.844
Classification rate for digit 1: 0.963
Classification rate for digit 2: 0.777
Classification rate for digit 3: 0.79
Classification rate for digit 4: 0.766
Classification rate for digit 5: 0.674
Classification rate for digit 6: 0.758
Classification rate for digit 7: 0.726
Classification rate for digit 8: 0.602
Classification rate for digit 9: 0.8
```

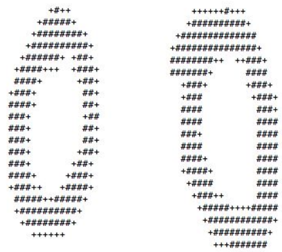
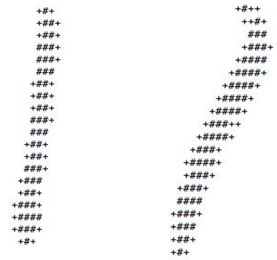
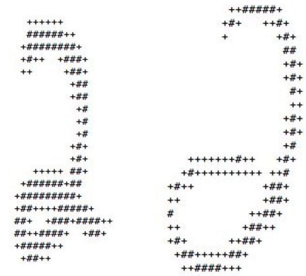
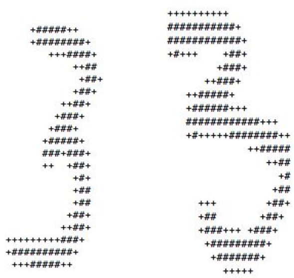
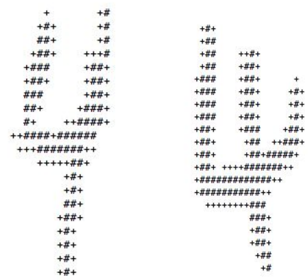
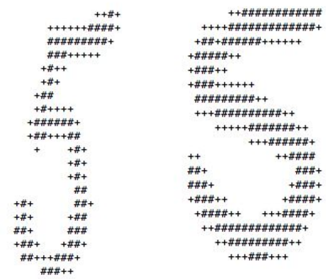
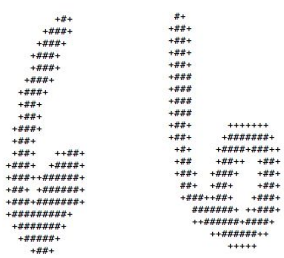
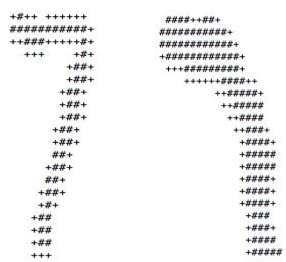
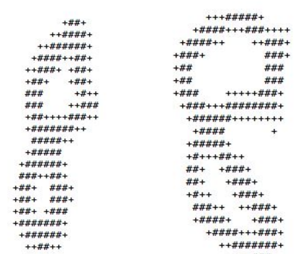
Also, we calculate the confusion matrix and most/least prototypical instance for better understanding of the learning result. (the original confusion matrix from program output is messy...so we manually input the matrix here)

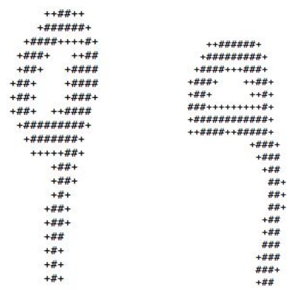
Confusion Matrix

	0	1	2	3	4	5	6	7	8	9
0	0.844	0.0	0.011	0.0	0.011	0.056	0.033	0.0	0.044	0.0
1	0.0	0.963	0.009	0.0	0.0	0.019	0.009	0.0	0.0	0.0
2	0.01	0.029	0.777	0.039	0.01	0.0	0.058	0.01	0.049	0.019
3	0.0	0.02	0.0	0.079	0.0	0.03	0.02	0.06	0.02	0.06
4	0.0	0.009	0.0	0.0	0.766	0.0	0.028	0.009	0.019	0.168
5	0.022	0.022	0.011	0.13	0.033	0.674	0.011	0.011	0.022	0.065
6	0.011	0.066	0.044	0.0	0.044	0.055	0.758	0.0	0.022	0.0
7	0.0	0.057	0.028	0.0	0.028	0.0	0.0	0.726	0.028	0.132
8	0.019	0.01	0.029	0.136	0.019	0.058	0.0	0.01	0.602	0.117
9	0.01	0.01	0.01	0.03	0.09	0.02	0.0	0.02	0.01	0.8

Row represents Truth, Column represents prediction

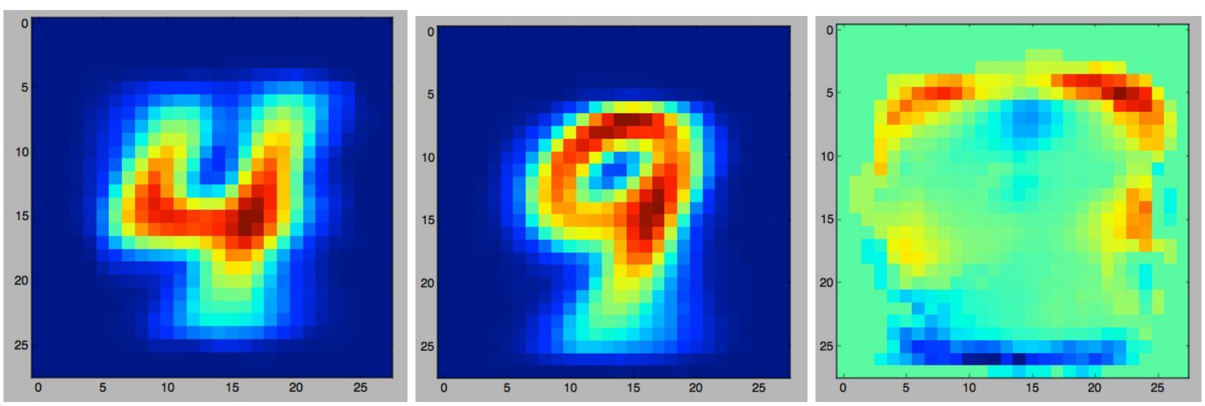
Most/Least prototypical instance

 Most (image: 723) Least (image: 610)	 Most (image: 633) Least (image: 993)	 Most (image: 795) Least (image: 50)
 Most (image: 205) Least (image: 291)	 Most (image: 111) Least (image: 253)	 Most (image: 471) Least (image: 70)
 Most (image: 632) Least (image: 362)	 Most (image: 784) Least (image: 671)	 Most (image: 560) Least (image: 758)

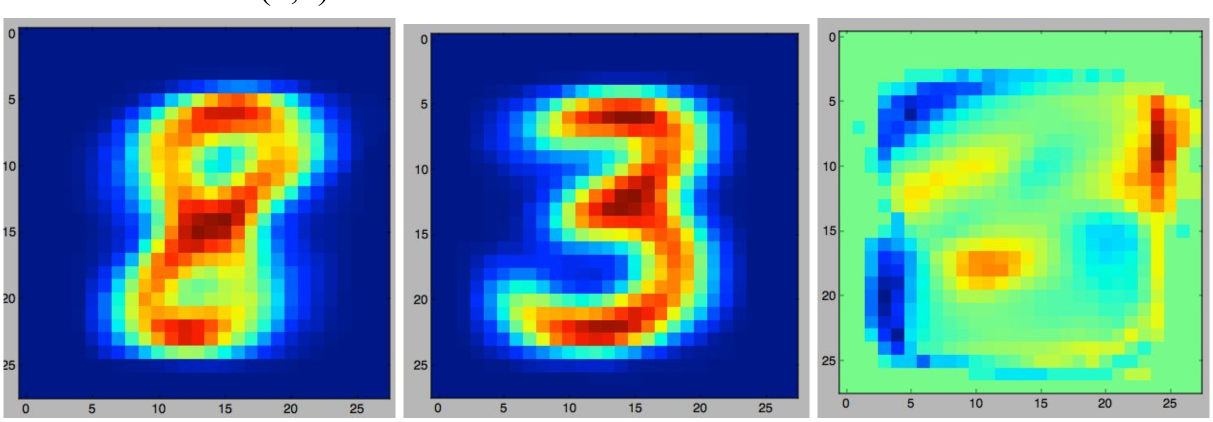
 Most (image: 745) Least (image: 492)	**The images use 0-based index.** **Left is Most prototypical, Right is Least prototypical.**

For the most confusing pairs in the confusion matrix, I have generated odd ratios and displayed results here. Note that the deeper the color the lower probability is. Light color denotes higher odd ratios.

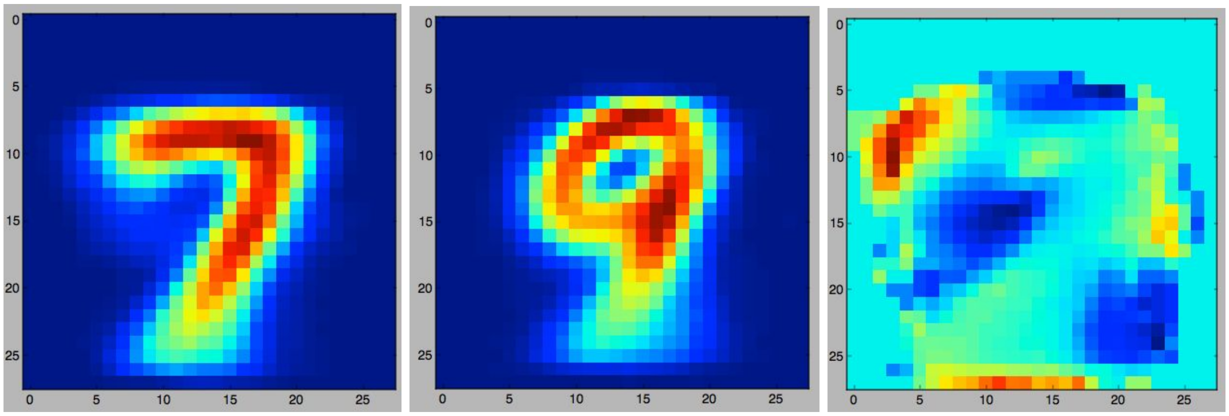
odd ratio: 0.168 (4,9)



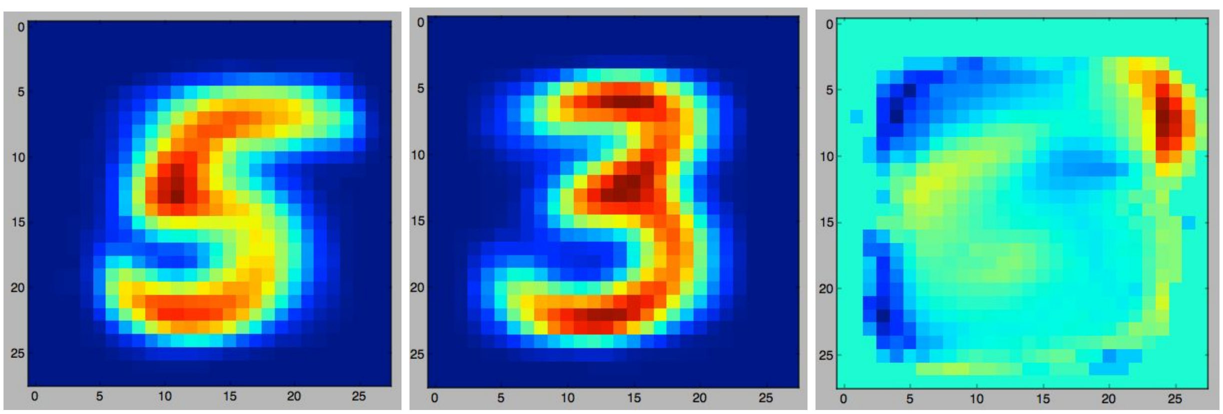
odd ratio: 0.136 (8,3)



odd ratio: 0.132 (7,9)



odd ratio: 0.13 (5,3)



Extra credit

In this part, we will perform classification on ternary features, which means treating '+' and '#' differently. Since we favor '#' more than '+', we will assign more weight when counting foreground. In this MP, we assign 1 to '#' and 0.5 to '+'. The accuracy is thus increased by 0.1%.

For facedata, we did some modification on the original code and achieve the following results. The output is satisfying:

```
Overall: 0.9066666666666666
Classification rate for not face: 0.883
Classification rate for face: 0.932
```

Part 2: Text Document Classification

–Bangqi Wang

This project uses two different Naïve Bayes classifiers, Multinomial Model and Bernoulli Model, to classify the text documents.

General Implementation

The train and test files contains the label and the words in each review or topics. Each record is preprocessed and stored as:

```
label word_1:count_1 word_2:count_2 ... word_n:count_n
```

The basic idea of this project is calculating the conditional probability of each word in each class and the probability of each class. The conditional probability is ***likelihood***, ***P(document | class)***. The probability of each class is ***priors***, ***P(class)***. The likelihood of document is represented by a sequence of words in the document, known as ***bag of words***, $P(w_i \mid class)$.

$$P(document \mid class) = P(w_1, \ldots, w_n \mid class) = \prod_{i=1}^{n} P(w_i \mid class)$$

prior

spam:	0.33
¬spam:	0.67

P(word | spam)

the :	0.0156
to :	0.0153
and :	0.0115
of :	0.0095
you :	0.0093
a :	0.0086
with:	0.0080
from:	0.0075
...	

P(word | ¬spam)

the :	0.0210
to :	0.0133
of :	0.0119
2002:	0.0110
with:	0.0108
from:	0.0107
and :	0.0105
a :	0.0100
...	

The algorithm will calculate the likelihood of document for each class and find the ***maximum likelihood (ML)*** as the predicted label for the documents. The main idea is to calculate the maximum likelihood estimate, but different models have different method to calculate the bag of words likelihood. To improve the accuracy and deal with the word that were never seen or seen too few times. The project uses the ***Laplacian smoothing***. The smoothing method will introduce below for different models.

Then the project assigns the document to the class with the highest posterior and avoid the underflow by using the logs of probabilities.

$$P(class \mid document) \propto P(class)\prod_{i=1}^{n} P(w_i \mid class)$$

$$L(class \mid document) = \log P(class) + \sum_{i=1}^{n} \log P(w_i \mid class)$$

Multinomial Model

The Multinomial Model calculates the likelihood for each word by calculating the times of occurrences. The algorithm smooths the probabilities by pretending have seen every vocabulary one more time than actually did.

$$\text{P(word | class)} = \frac{\#\ of\ occurrences\ of\ this\ word\ in\ docs\ from\ this\ class}{total\ \#\ of\ words\ in\ docs\ from\ this\ class}$$

$$\text{P(word | class)} = \frac{\#\ of\ occurrences\ of\ this\ word\ in\ docs\ from\ this\ class + 1}{total\ \#\ of\ words\ in\ docs\ from\ this\ class + V}$$

(V: total number of unique words)

Bernoulli Model

The Bernoulli model calculates the likelihood for each word by counting whether the word appeared at least once. The algorithm smooths the probabilities by pretending have seen every vocabulary one more times than actually did.

$$\text{P(word | class)} = \frac{\#\ of\ documents\ this\ word\ appeared\ from\ this\ class}{total\ \#\ of\ documents\ from\ this\ class}$$

$$\text{P(word | class)} = \frac{\#\ of\ documents\ this\ word\ appeared\ from\ this\ class + 1}{total\ \#\ of\ documents\ from\ this\ class + 2}$$

Part 2.1: Classification Results

The table below are the classification results for part 2.1.

Accuracy & Confusion Matrix

The two tables contain the output for two different datasets. The overall accuracies on the sentiment analysis of movie review task are around 76% and the accuracies on the topical theme classification task are around 93%.

Sentiment Analysis of Movie Review					
	Accuracy		**Confusion Matrix**		
Multinomial	Overall	76.00%		Negative	Positive
	Negative	75.00%	Negative	75.00%	25.00%
	Positive	77.00%	Positive	23.00%	77.00%
Bernoulli	Overall	76.00%		Negative	Positive
	Negative	72.80%	Negative	72.80%	27.20%
	Positive	79.20%	Positive	20.80%	79.20%

(Table for dataset 1)

Binary conversation topic classification					
	Accuracy		**Confusion Matrix**		
Multinomial	Overall	90.82%		Life Partner	Min Wage
	Life Partner	95.92%	Life Partner	95.92%	4.08%
	Min Wage	85.71%	Min Wage	14.29%	85.71%
Bernoulli	Overall	94.90%		Life Partner	Min Wage
	Life Partner	91.84%	Life Partner	91.84%	8.16%
	Min Wage	97.96%	Min Wage	2.04%	97.96%

(Table for dataset 2)

Top 10 Words with the Highest Likelihood

The tables below contain the words that are appear a lot in each classes.

Sentiment Analysis of Movie Review				
	top 10 words with the highest likelihood			
	Negative		Positive	
Multinomial	Movie	0.009317	Film	0.009054
	Film	0.007300	Movie	0.005952
	Like	0.005251	--	0.004337
	One	0.004610	One	0.003546
	--	0.003782	Like	0.003166
	Bad	0.002817	Story	0.003008
	Story	0.002753	Good	0.002610
	Much	0.002689	Comedy	0.002659
	Time	0.002433	Way	0.002564
	Even	0.002273	Even	0.002438
Bernoulli	Film	0.03891	Film	0.03931
	Movie	0.03069	Movie	0.02509
	One	0.02151	One	0.01533
	Like	0.01891	Like	0.01380
	--	0.01665	--	0.01366
	Story	0.01137	Story	0.01268
	Comedy	0.01110	Comedy	0.01143
	Way	0.01055	Way	0.0101
	Even	0.01000	Even	0.01073
	Good	0.00959	Good	0.01031

Binary conversation topic classification				
	top 10 words with the highest likelihood			
	Life Partner		Min Wage	
Multinomial	Know	0.05446	Know	0.05153
	Yeah	0.04535	Yeah	0.04541
	Uh	0.03034	Like	0.02889
	Like	0.02962	Uh	0.02192
	Um	0.02310	Um	0.01960
	Right	0.01933	Right	0.01848
	Just	0.01798	Don	0.01733
	Think	0.01769	Think	0.01707
	Oh	0.01623	Just	0.01649
	Don	0.01590	Oh	0.01482
Bernoulli	Like	0.03780	Um	0.03907
	Know	0.03780	Think	0.03907
	Just	0.03780	Like	0.03907
	Yeah	0.03763	Know	0.03907
	Think	0.03763	Just	0.03907
	Don	0.03763	Don	0.03907
	Um	0.03720	Yeah	0.03898
	Right	0.03711	People	0.03898
	Oh	0.03702	Oh	0.03881
	Really	0.03685	Right	0.03872

Top 10 Words with the Highest Odds Ration

The tables below contain the words that are more likely to appear in specific class.

Sentiment Analysis of Movie Review				
	top 10 words with the highest odds ratio			
	Negative		Positive	
Multinomial	Flat	15.1695	Disturbing	14.8324
	Stale	14.1582	Refreshingly	10.8771
	Dull	13.6526	Haunting	10.8771
	Tired	12.1356	Grief	10.8771
	Plain	11.1243	Engrossing	10.8771
	Mediocre	11.1243	Refreshing	9.8882
	Unfunny	10.1130	Polished	9.8882
	Poorly	10.1130	Inventive	9.8882
	Pointless	10.1130	Gripping	9.8882
	Generic	10.1130	Gem	9.8882
Bernoulli	Flat	14.7431	Disturbing	14.2439
	Stale	13.7602	Refreshingly	11.1917
	Dull	12.7774	Haunting	11.1917
	Tired	11.7945	Grief	11.1917
	Plain	10.8116	Engrossing	11.1917
	Mediocre	10.8116	Refreshing	10.1742
	Unfunny	9.8288	Polished	10.1742
	Poorly	9.8288	Inventive	10.1742
	Pointless	9.8288	Gripping	10.1742
	Generic	9.8288	Gem	10.1742

Binary conversation topic classification				
	top 10 words with the highest odds ratio			
	Life Partner		Min Wage	
Multinomial	Relationship	198.1665	Wage	150.4568
	Compatibility	119.8154	Minimum	150.3094
	Communication	116.6344	Welfare	128.7356
	Marriage	112.1281	Wages	112.7026
	Partner	108.6063	Inflation	90.5393
	Relationships	96.1350	Waitresses	77.3357
	Friendship	95.4282	Tax	76.3925
	Attracted	94.3679	Waitress	68.8476
	Dating	89.0663	Salary	65.0751
	Compatible	82.7044	Increase	58.4733
Bernoulli	Attracted	62.1980	Wage	79.2308
	Compatibility	60.2543	Wages	70.9990
	Relationship	55.3951	Inflation	63.7962
	Relationships	46.9725	Waitresses	56.5934
	Friendship	46.6485	Minimum	56.5934
	Attraction	46.6485	Waitress	54.5355
	Marriage	44.2189	Welfare	52.4775
	Communication	44.2189	Salary	46.8182
	Dating	43.0851	Retail	41.1589
	Qualities	40.8175	Increase	40.1299

Conclusion

The top 10 words with highest likelihood are almost the same for both classes because the words with the most occurrences are similar to stop words. Those words cannot stand for any class because they have little meaning and they are more likely the general words around the topic. However, the top 10 words with highest odds ratio are high representative. The words with high odds ratio means that the words are more likely to appear in specific class only. The words with high likelihood I discussed above have similar high occurrences in any class. Therefore, the words with high likelihood not necessary have high odds ratio.

Part 2.2: Classification Results

The dataset for this part is pretty large and contains 40 classes. Some classes might have common words but with different frequency. In this case, the multinomial model has more accuracy because the occurrences of words do matter in the topics. E.g. distributed system & database system.

Accuracy

Full 40 Topic Corpus		
	Accuracy	
	Multinomial	**Bernoulli**
Accuracy	83.88%	50.87%

Confusion Matrix

The confusion matrix is too large and I will divide the table into 4 tables and display one by one.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0.84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0.89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0.92	0.02	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0
3	0	0	0	0.84	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0
4	0	0	0	0	0.95	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0.23	0	0	0.54	0	0	0	0.15	0	0	0	0	0	0	0	0	0	0
6	0	0	0.08	0	0	0	0.08	0.08	0	0.58	0	0	0	0	0	0	0	0	0	0
7	0.04	0	0	0	0	0	0	0.86	0	0	0	0	0	0	0	0	0	0	0	0
8	0.03	0	0.06	0	0	0	0	0	0.82	0.03	0	0	0	0	0	0	0	0	0	0
9	0	0.03	0	0	0	0	0	0	0.03	0.88	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0	0	0
11	0	0	0.06	0	0	0	0	0	0.06	0	0	0.83	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
14	0.04	0	0.04	0.04	0	0	0	0	0	0	0	0	0.04	0	0.8	0	0	0	0	0
15	0.04	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0.86	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
17	0	0	0	0.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0
18	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.57	0
19	0	0.08	0	0.08	0	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0	0.67
20	0	0	0.71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
24	0.07	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	*0.04*	0	0	0	0
25	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0.12	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0
31	0.04	0.04	0	0	0	0	0	0	0	0	0	0	0.08	0	0	0.04	0	0	0	0
32	0	0	0.13	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0	0
33	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
0	0	0	0	0.13	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0.04	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04
3	0	0	0	0.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0
4	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0.04	0	0	0	0	0	0	0	0	0.04	0	0.04	0	0	0	0	0	0
8	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0
9	0	0	0	0	0	0	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0.04	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0.14	0	0	0	0	0	0	0	0	0	0	0	0	0.21	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0
20	0.07	0	0	0	0	0	0	0	0	0.21	0	0	0	0	0	0	0	0	0	0
21	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0.9	0	0	0	0.03	0.03	0	0	0	0	0	0	0	0	0.03	0	0
24	0	0	0	0	*0.85*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0.26	0	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0.14	0	0	0.24	0.1	0	0	0	0	0	0	0	0	0	0.52	0	0
27	0	0	0	0.03	0	0	0	0.83	0	0	0	0	0	0	0	0	0	0.07	0	0
28	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0.04	0	0	0	0.04	0	0	0.76	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0.81	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0.04	0	0.08	0.71	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0
35	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0.04	0.85	0.04	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
37	0	0	0	0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0.89	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.96	0
39	0	0	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0.43

(Multinomial Model)

Multinomial Model

Confusion Matrix [0:20][0:20]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0.84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0.89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0.92	0.02	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0
3	0	0	0	0.84	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0
4	0	0	0	0	0.95	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0.23	0	0	0.54	0	0	0	0.15	0	0	0	0	0	0	0	0	0	0
6	0	0	0.08	0	0	0	0.08	0.08	0	0.58	0	0	0	0	0	0	0	0	0	0
7	0.04	0	0	0	0	0	0	0.86	0	0	0	0	0	0	0	0	0	0	0	0
8	0.03	0	0.06	0	0	0	0	0	0.82	0.03	0	0	0	0	0	0	0	0	0	0
9	0	0.03	0	0	0	0	0	0	0.03	0.88	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0	0	0
11	0	0	0.06	0	0	0	0	0	0.06	0	0	0.83	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
14	0.04	0	0.04	0.04	0	0	0	0	0	0	0	0	0.04	0	0.8	0	0	0	0	0
15	0.04	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0.86	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
17	0	0	0	0.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0
18	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.57	0
19	0	0.08	0	0.08	0	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0	0.67

(Multinomial Model)

Confusion Matrix [0:20][20:40]

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
0	0	0	0	0.13	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0.04	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04
3	0	0	0	0.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0
4	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0.04	0	0	0	0	0	0	0	0	0.04	0	0.04	0	0	0	0	0	0
8	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0
9	0	0	0	0	0	0	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0.04	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0.14	0	0	0	0	0	0	0	0	0	0	0	0	0.21	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0

(Multinomial Model)

Confusion Matrix [20:40][0:20]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
20	0	0	0.71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
24	0.07	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0
25	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0.12	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0
31	0.04	0.04	0	0	0	0	0	0	0	0	0	0	0.08	0	0	0.04	0	0	0	0
32	0	0	0.13	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0	0
33	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(Multinomial Model)

Confusion Matrix [20:40][20:40]

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
20	0.07	0	0	0	0	0	0	0	0	0.21	0	0	0	0	0	0	0	0	0	0
21	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0.9	0	0	0	0.03	0.03	0	0	0	0	0	0	0	0	0.03	0	0
24	0	0	0	0	0.85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0.26	0	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0.14	0	0	0.24	0.1	0	0	0	0	0	0	0	0	0	0.52	0	0
27	0	0	0	0.03	0	0	0	0.83	0	0	0	0	0	0	0	0	0	0.07	0	0
28	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0.04	0	0	0	0.04	0	0	0.76	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0.81	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0.04	0	0.08	0.71	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0
35	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0.04	0.85	0.04	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
37	0	0	0	0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0.89	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.96	0
39	0	0	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0.43

(Multinomial Model)

Bernoulli Model

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
0	0.84	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0.91	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0.92	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0.92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0.02	0.02	0.95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0.54	0.38	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0.08	0.83	0	0	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0.04	0.07	0	0.82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0
8	0	0	0.59	0.38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
9	0	0.03	0.12	0.44	0	0	0	0	0	0.38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0
10	0	0	0	0.09	0	0	0	0	0	0	0.86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0	0
11	0	0.06	0	0.83	0	0	0	0	0	0	0	0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0.3	0.39	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0	0
13	0.38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.63	0	0	0	0	0	0	0	0	0	0	0
14	0.04	0	0.04	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0
15	0.07	0.04	0.29	0.43	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0.04	0	0	0	0	0	0.07	0	0	0	0	0	0	0	0.04	0	0	0
16	0	0	0	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0.87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0.57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.43	0	0	0
19	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0.07	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0.83	0	0	0	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0
22	0.03	0	0	0.09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0.03	0.28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.68	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
24	0.3	0.04	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0.26	0	0.26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0.05	0.81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0.03	0.45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.24	0	0	0	0.28	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
29	0	0.03	0.08	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.72	0	0	0	0	0	0	0	0	0	0
30	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0.15	0.42	0.04	0.15	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0.12	0	0	0	0	0	0	0	0
32	0	0	0.92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.94	0	0	0	0	0	0
34	0	0.15	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.71	0	0	0	0	0
35	0	0.04	0.08	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.58	0.04	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
37	0	0	0.04	0.86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0.04	0	0	0	0	0	0	0	0	0	0.04	0	0
38	0	0.04	0.08	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.19	0	0	0	0	0	0	0	0	0.54	0
39	0	0	0.22	0.65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(Bernoulli Model)

Confusion Matrix [0:20][0:20]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0.84	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0.91	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0.92	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0.92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0.02	0.02	0.95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0.54	0.38	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0.08	0.83	0	0	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0
7	0.04	0.07	0	0.82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0.59	0.38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0.03	0.12	0.44	0	0	0	0	0	0.38	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0.09	0	0	0	0	0	0	0.86	0	0	0	0	0	0	0	0	0
11	0	0.06	0	0.83	0	0	0	0	0	0	0	0.11	0	0	0	0	0	0	0	0
12	0	0	0	0.3	0.39	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0
13	0.38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0.04	0	0.04	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0.07	0.04	0.29	0.43	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0
16	0	0	0	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0.87	0	0	0
17	0	0	0	0.69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0.57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(Bernoulli Model)

Confusion Matrix [0:20][20:40]

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
0	0	0	0	0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0.06	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0
10	0	0	0	0	0	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0	0
13	0	0	0	0	0	0	0	0	0.63	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0
15	0	0	0.04	0	0	0	0	0	0.07	0	0	0	0	0	0	0	0.04	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0.31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.43	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(Bernoulli Model)

Confusion Matrix [20:40][0:20]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0.07	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0
22	0.03	0	0	0.09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0.03	0.28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0.3	0.04	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0.26	0	0.26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0.05	0.81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0.03	0.45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0.03	0.08	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0.15	0.42	0.04	0.15	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0.92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0.15	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0.04	0.08	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0.04	0.86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0.04	0.08	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0.22	0.65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(Bernoulli Model)

Confusion Matrix [20:40][20:40]

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0.83	0	0	0	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0.88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0.68	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0.1	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0.24	0	0	0	0.28	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0.72	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0.04	0	0	0	0	0	0	0	0	0.12	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0.94	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.71	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.58	0.04	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
37	0	0	0	0.04	0	0	0	0.04	0	0	0	0	0	0	0	0	0	0.04	0	0
38	0	0	0	0	0	0	0	0	0	0.19	0	0	0	0	0	0	0	0	0.54	0
39	0	0	0	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(Bernoulli Model)

Confused Topic in Multinomial Model

The table below shows the most confused topic for each topic.

Multinomial

The overall accuracy for multinomial model is around 83.88%.

Real Topic	Accuracy	Confused Topic	Percentage
0	84.44%	23	13.33%
1	89.29%	25	7.14%
2	91.84%	39	4.08%
3	83.67%	23	12.24%
4	95.45%	7	2.27%
5	53.85%	2	23.08%
6	58.33%	23	16.67%
7	85.71%	0	3.57%
8	82.35%	2	5.88%
9	88.24%	29	5.88%
10	97.14%	28	2.86%
11	83.33%	2	5.56%
12	100.00%	0	0.00%
13	100.00%	0	0.00%
14	80.00%	0	4.00%
15	85.71%	0	3.57%
16	100.00%	0	0.00%
17	68.75%	3	18.75%
18	57.14%	36	21.43%
19	66.67%	1	8.33%
20	71.43%	29	21.43%
21	100.00%	0	0.00%
22	100.00%	0	0.00%
23	90.00%	8	2.50%
24	85.19%	0	7.41%
25	69.57%	23	26.09%
26	52.38%	26	23.81%
27	82.76%	3	6.90%
28	100.00%	0	0.00%
29	97.22%	1	2.78%
30	76.00%	3	12.00%
31	80.77%	12	7.69%
32	70.83%	2	12.50%
33	96.88%	2	3.13%
34	97.06%	7	2.94%
35	84.62%	16	3.85%
36	100.00%	0	0.00%
37	89.29%	23	10.71%
38	96.15%	1	3.85%
39	43.48%	23	30.43%

Bernoulli

The overall accuracy for Bernoulli model is 50.87%

Real Topic	Accuracy	Confused Topic	Percentage
0	84.44%	23	11.11%
1	91.07%	23	7.14%
2	91.84%	3	8.16%
3	91.84%	23	6.12%
4	95.45%	2	2.27%
5	53.85%	3	38.46%
6	83.33%	2	8.33%
7	82.14%	1	7.14%
8	58.82%	3	38.24%
9	44.12%	9	38.24%
10	85.71%	3	8.57%
11	83.33%	11	11.11%
12	39.13%	3	30.43%
13	62.50%	0	37.50%
14	80.00%	23	8.00%
15	42.86%	2	28.57%
16	86.96%	3	13.04%
17	68.75%	23	31.25%
18	57.14%	36	42.86%
19	100.00%	0	0.00%
20	100.00%	0	0.00%
21	83.33%	3	6.67%
22	87.50%	3	9.38%
23	67.50%	3	27.50%
24	33.33%	3	33.33%
25	47.83%	1	26.09%
26	80.95%	23	9.52%
27	44.83%	27	27.59%
28	100.00%	0	0.00%
29	72.22%	3	16.67%
30	100.00%	0	0.00%
31	42.31%	0	15.38%
32	91.67%	29	8.33%
33	93.75%	3	6.25%
34	70.59%	1	14.71%
35	57.69%	3	23.08%
36	100.00%	0	0.00%
37	85.71%	2	3.57%
38	53.85%	29	19.23%
39	65.22%	2	21.74%