Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

Problem 1

Part A

$$P(x|y) = \sum_c P(x|y=c)P(y=c)$$

i)

We have for some c

$P(x|y=c) = P(x_1, x_2, \dots, x_D|y=c)$

$= P(x_2, \dots, x_D|x_1, y=c)P(x_1|y=c)$ (note that x_1 can take two values -> 2 permutations)

$= P(x_3, \dots, x_D)P(x_2|x_1, y=c)P(x_1, y=c)$ (note that x_1, x_2 can each take two values -> 4 permutations)

As we can see if we continues this process we will get the following series to represent the permutations that must be considered:

$$2 + 2^2 + 2^3 + \cdots + 2^D$$

$$= \frac{2^{D+1} - 1}{1}$$

$$\Rightarrow O(2^D)$$

We must do this for all c so altogether we have:

$$O(C \cdot 2^D)$$

ii)

We know that

$$p(x\,|y=c) = \prod_{j=1}^{D} P\left(x_j\,|y=c\right)$$

Each x_j can take on 1 of 2 values
So each such calculation is $O(2^D)$
We need to do this $\forall c$ of which there are C

$$\Rightarrow O(C \cdot 2^D)$$

They are the same.

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

Part B

With small N, I would expect naïve bayes to perform better on the testing set because will far more parameters than naïve, full bayes would almost definitely over fit drastically leading to a low training error but a higher testing error, whereas naïve would probably have a higher training error than full but a lower testing error.

Part C

With large N, I would expect full bayes to out perform naïve bayes. This is because full bayes has more paramters which gives the model more flexibility to fit the underlying nature of the data, while a large N would ideally give a good representation of that nature to the point where the model would not suffer from overfitting.

Part D

We want

$$P(y=c|x)=\frac{P(x\mid y=c)P(y=c)}{P(x)}$$

$$P(x|y=c)$$

Can be computed in $O(D)$ since there are D x's

$$P(y=c)$$

Can be computed in $O(1)$ since we already have the value and we have a uniform class prior and it is a simple lookup.

We also have:

$$P(x)=\sum_c P(x\,|y=c)P(y=c)$$

Which can be computed in $O(C)$ since there are C elements to the summation.

Altogether we have:

$$O(CD)$$

Computing the prediction for full bayes would also be a $O(CD)$ computation
We want

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

$$P(x|y=c) = \prod_{j=1}^{D} P\left(x_j \,\middle|\, x_{j-1}, \dots, x_1, y=c\right) = \prod_{j=1}^{D} \theta_{xjc}$$

This information can be stored in a 3 nested array and you can get the first index as given in $O(D)$

To get probabilities for a given j we need $O(D)$,
To do so for all possible y would give $O(CD)$.

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

Problem 2

Part A

```
######################################################################
                     Running Code For Question 2A
######################################################################


File #0:
Emission Sequence              Max Probability State Sequence
######################################################################
25421                          31033
01232367534                    22222100310
5452674261527433               1031003103222222
7226213164512267255            1310331000033100310
0247120602352051010255241      2222222222222222222222103


File #1:
Emission Sequence              Max Probability State Sequence
######################################################################
77550                          22222
7224523677                     2222221000
505767442426747                222100003310031
72134131645536112267           10310310000310333100
4733667771450051060253041      2221000003222223103222223


File #2:
Emission Sequence              Max Probability State Sequence
######################################################################
60622                          11111
4687981156                     2100202111
815833657775062                021011111111111
21310222515963505015           02020111111111111021
6503199452571274006320025      1110202111111102021110211


File #3:
Emission Sequence              Max Probability State Sequence
######################################################################
13661                          00021
2102213421                     3131310213
166066262165133                133333133133100
53164662112162634156           20000021313131002133
1523541005123230226306256      1310021333133133313133133


File #4:
Emission Sequence              Max Probability State Sequence
######################################################################
23664                          01124
3630535602                     0111201112
350201162150142                011244012441112
00214005402015146362           11201112412444011112
2111266524665143562534450      2012012424124011112411124


File #5:
Emission Sequence              Max Probability State Sequence
######################################################################
68535                          10111
4546566636                     1111111111
638436858181213                110111010000011
13240338308444514688           00010000000111111100
0111664434441382533632626      2111111111111100111110101
```

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

Part Bi

```
######################################################################
                    Running Code For Question 2Bi
######################################################################


File #0:
Emission Sequence             Probability of Emitting Sequence
######################################################################
25421                         4.537e-05
01232367534                   1.620e-11
5452674261527433              4.348e-15
7226213164512267255           4.739e-18
0247120602352051010255241     9.365e-24


File #1:
Emission Sequence             Probability of Emitting Sequence
######################################################################
77550                         1.181e-04
7224523677                    2.033e-09
505767442426747               2.477e-13
72134131645536112267          8.871e-20
4733667771450051060253041     3.740e-24


File #2:
Emission Sequence             Probability of Emitting Sequence
######################################################################
60622                         2.088e-05
4687981156                    5.181e-11
815833657775062               3.315e-15
21310222515963505015          5.126e-20
6503199452571274006320025     1.297e-25


File #3:
Emission Sequence             Probability of Emitting Sequence
######################################################################
13661                         1.732e-04
2102213421                    8.285e-09
166066262165133               1.642e-12
53164662112162634156          1.063e-16
1523541005123230226306256     4.535e-22


File #4:
Emission Sequence             Probability of Emitting Sequence
######################################################################
23664                         1.141e-04
3630535602                    4.326e-09
350201162150142               9.793e-14
00214005402015146362          4.740e-18
2111266524665143562534450     5.618e-22


File #5:
Emission Sequence             Probability of Emitting Sequence
######################################################################
68535                         1.322e-05
4546566636                    2.867e-09
638436858181213               4.323e-14
13240338308444514688          4.629e-18
0111664434441382533632626     1.440e-22
```

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

Part Bii

```
######################################################################
                    Running Code For Question 2Bii
######################################################################


File #0:
Emission Sequence              Probability of Emitting Sequence
######################################################################
25421                          4.537e-05
01232367534                    1.620e-11
5452674261527433               4.348e-15
7226213164512267255            4.739e-18
0247120602352051010255241      9.365e-24


File #1:
Emission Sequence              Probability of Emitting Sequence
######################################################################
77550                          1.181e-04
7224523677                     2.033e-09
505767442426747                2.477e-13
72134131645536112267           8.871e-20
4733667771450051060253041      3.740e-24


File #2:
Emission Sequence              Probability of Emitting Sequence
######################################################################
60622                          2.088e-05
4687981156                     5.181e-11
815833657775062                3.315e-15
21310222515963505015           5.126e-20
6503199452571274006320025      1.297e-25


File #3:
Emission Sequence              Probability of Emitting Sequence
######################################################################
13661                          1.732e-04
2102213421                     8.285e-09
166066262165133                1.642e-12
53164662112162634156           1.063e-16
1523541005123230226306256      4.535e-22


File #4:
Emission Sequence              Probability of Emitting Sequence
######################################################################
23664                          1.141e-04
3630535602                     4.326e-09
350201162150142                9.793e-14
00214005402015146362           4.740e-18
2111266524665143562534450      5.618e-22


File #5:
Emission Sequence              Probability of Emitting Sequence
######################################################################
68535                          1.322e-05
4546566636                     2.867e-09
638436858181213                4.323e-14
13240338308444514688           4.629e-18
0111664434441382533632626      1.440e-22
```

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

Part C

```
#######################################################################
                    Running Code For Question 2C
#######################################################################


Transition Matrix:
#######################################################################
2.833e-01   4.714e-01   1.310e-01   1.143e-01
2.321e-01   3.810e-01   2.940e-01   9.284e-02
1.040e-01   9.760e-02   3.696e-01   4.288e-01
1.883e-01   9.903e-02   3.052e-01   4.075e-01


Observation Matrix:
#######################################################################
1.486e-01   2.288e-01   1.533e-01   1.179e-01   4.717e-02   5.189e-02   2.830e-02   1.297e-01   9.198e-02   2.358e-03
1.062e-01   9.653e-03   1.931e-02   3.089e-02   1.699e-01   4.633e-02   1.409e-01   2.394e-01   1.371e-01   1.004e-01
1.194e-01   4.299e-02   6.529e-02   9.076e-02   1.768e-01   2.022e-01   4.618e-02   5.096e-02   7.803e-02   1.274e-01
1.694e-01   3.871e-02   1.468e-01   1.823e-01   4.839e-02   6.290e-02   9.032e-02   2.581e-02   2.161e-01   1.935e-02
```

Part D

```
#######################################################################
                    Running Code For Question 2D
#######################################################################


Transition Matrix:
#######################################################################
4.695e-21   9.707e-02   4.454e-01   4.575e-01
2.223e-01   3.357e-01   4.420e-01   1.048e-12
1.886e-07   9.071e-02   5.227e-01   3.866e-01
4.612e-01   4.550e-01   1.188e-12   8.381e-02


Observation Matrix:
#######################################################################
3.971e-04   2.496e-02   4.739e-25   2.299e-04   3.216e-01   4.569e-04   1.426e-18   2.890e-01   1.039e-01   2.594e-01
1.147e-01   5.540e-19   1.262e-10   3.904e-01   7.338e-02   2.853e-01   1.350e-01   1.160e-03   1.316e-06   4.881e-26
2.235e-01   1.379e-01   1.511e-01   4.230e-02   9.365e-02   4.842e-02   1.118e-01   5.696e-15   1.914e-01   1.267e-19
1.139e-01   6.457e-02   1.656e-01   1.249e-08   4.343e-02   4.909e-02   1.994e-02   2.412e-01   1.985e-01   1.038e-01
```

Part E

The transition matrices from 2C appear more accurate for a few reasons. For starters, it can't be seen on the above images but I ran part D multiple times and the transition matrix values were highly variable which seems to indicate that the unsupervised algorithm doesn't have as accurate convergence behavior and is encountering various local minima. Intuitively, it makes sense that the data with labels yields more consistent and accurate training.

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

Part F

```
########################################################################
                     Running Code For Question 2F
########################################################################


File #0:
Generated Emission
########################################################################
427661071701412455242
131731467734055575442
344400540742514765773
135254777552254057517
467423417745623201234


File #1:
Generated Emission
########################################################################
067764044216476645240
663662542506222551424
612707566524276407020
214725477755040752075
742263765052117377424


File #2:
Generated Emission
########################################################################
286545092525265966101
267250186090612527917
337921051161687611260
463447658678178392873
731576891319898272767


File #3:
Generated Emission
########################################################################
540422214555353351601
505226262212413366045
634200566366113336415
651522106421121166602
131246114332542452002


File #4:
Generated Emission
########################################################################
152360612642465666146
331626365343605603465
420346266241533561666
100026303066052030110
623163031011300022232


File #5:
Generated Emission
########################################################################
033314146403836321743
316421818863548821538
488868118840856831421
316344645864843302838
500264815484263381116
```

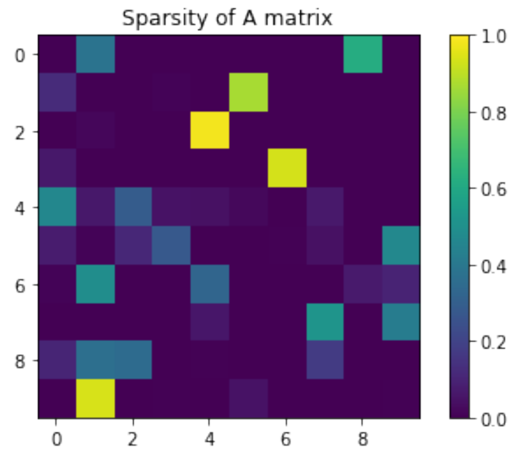
Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

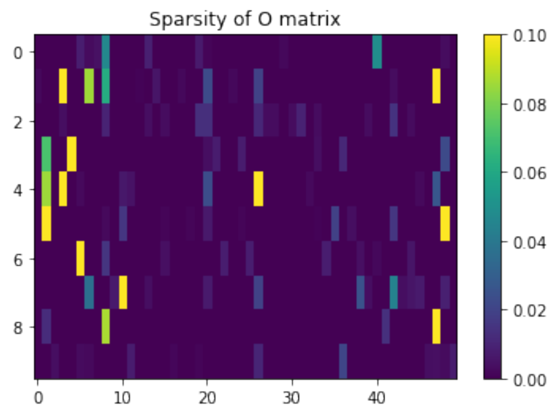
Part G

The transition matrix is rather sparse with mostly values that are close to or at 0.

There are handful of entries that have high values which indicates a high probability of transition from one state to another. This indicates that the state transitions are fairly determined in that there isn't a ton of variability in which state comes next given a current state.

We have somewhat similar sparsity in the O matrix. A strong majority of the entries appear to be 0. In each row there are generally no more than a few non-zero entries which indicates that a state is only liable to emit a very specific subset of emissions.





Part H

In the special case of only 1 hidden state we know that we are essentially choosing words randomly. The transition matrix would be 1 by 1 and the observation matrix would be 1 by D. Since there is only one state we can also say that the observation matrix would be uniform. Thus, this special case corresponds to choosing words at complete random.

We know that as we increase the number of hidden states we are increasing the likelihood of the training data since we are allowing more parameters with which the model can fit the training data.

We also notice an increase in subsets of the emission that are semblances of grammatically correct clauses in English. These occur more consistently with more hidden states.

Eli Pinkus
CS 155
Set 6
I AM USING MY REMAINING 46 EXTENSION HOURS

```
hmm1 = unsupervised_HMM(obs, 1, 100)
print('\nSample Sentence:\n====================')
print(sample_sentence(hmm1, obs_map, n_words=25))
```

```
Sample Sentence:
====================
Powers of power offices to law to two in holding them to of unanimous such of eight the enum
eration rules the of president absent profit before...
```

```
hmm2 = unsupervised_HMM(obs, 2, 100)
print('\nSample Sentence:\n====================')
print(sample_sentence(hmm2, obs_map, n_words=25))
```

```
Sample Sentence:
====================
Follows on subject for shall to as v the seventh or oath state president of to and the numbe
r not amendment of up no house part...
```

```
hmm4 = unsupervised_HMM(obs, 4, 100)
print('\nSample Sentence:\n====================')
print(sample_sentence(hmm4, obs_map, n_words=25))
```

```
Sample Sentence:
====================
Shall other the the by the the to convened of the against provide and the united the the of
the shall enumeration who extend he the...
```

```
hmm16 = unsupervised_HMM(obs, 16, 100)
print('\nSample Sentence:\n====================')
print(sample_sentence(hmm16, obs_map, n_words=25))
```

```
Sample Sentence:
====================
Uniform other and but the united adjourn congress alliance or of the office of the a shall r
eserving up chuse them and like choice of the...
```

Part I

I find state 2 to be semantically meaningful in that the words have somewhat similar properties. The major components consist of words like Law, Power, Service, Public which are related in that they are nouns that do well to describe situations in public policies.

This differs from others in that a lot of the other word clouds in that the prominent words are related and could occur in similar contexts in sentences and they are also non-filler words.

State 2