1 Class-Conditional Densities for Binary Data

Problem A

Let us use the chain rule of probability to factorize $p(x|y)$. Let $\theta_{xjc} = p(x_j|x_{1,...,j-1}, y = c)$ (such that $\theta_{x1c} = p(x_1|y = c)$).

$$p(x|y = c) = p(x_1, x_2, ..., x_D|y = c)$$
$$= p(x_2, x_3, ..., x_D|x_1, y = c) \cdot p(x_1|y = c)$$
$$= p(x_3, x_4, ..., x_D|x_1, x_2, y = c) \cdot p(x_2|x_1, y = c) \cdot p(x_1|y = c)$$
$$= p(x_4, x_5, ..., x_D|x_1, x_2, x_3, y = c) \cdot p(x_3|x_1, x_2, y = c) \cdot p(x_2|x_1, y = c) \cdot p(x_1|y = c)$$
$$...$$
$$= \theta_{xDc} \cdot \theta_{x(D-1)c} \cdot \theta_{x(D-2)c} \cdot ... \cdot \theta_{x1c}$$
$$= \Pi_{j=1}^{D} \theta_{xjc}$$

Assuming we store each θ_{xjc}, the number of parameters needed to represent this factorization can be expressed by $O(2^D \cdot C)$. This is because we are working with binary features, so for a single class, we have to consider permutations that scale to 2^D. Thus, for C classes, we consider $O(2^D \cdot C)$ parameters for this factorization.

Problem B

Let us consider if we just used the joint probability $p(x|y = c)$ with arbitrary x (length D) and c (C total classes). For the input x, there are 2^D possibilities since each x_j is binary. So, for all c, the number of parameters needed to be estimated for this computation is expressed by $O(2^D \cdot C)$ which is the same as *problem A*.

Problem C

Let us assume that the number of features D is fixed and let there be N training cases. If the sample size N is very small, the Naive Bayes model is likely to give lower test set error because the full model is likely to overfit with small N since it has more parameters. In other words, the full model would likely have lower training error and higher testing error in this case.

Problem D

For large sample size N, the full model is likely to give lower test set error because it is less likely that it will overfit in this case, thus the full model is more likely to have a more accurate performance on the test set relative to the Naive Bayes model considering the difference in the number of parameters and the fully dependent features of the full model. In this case, it is also more likely that the Naive Bayes model would underfit the data.

Problem E

2 Sequence Prediction

Question 2 Code

Problem A

```
File #0:
Emission Sequence             Max Probability State Sequence
######################################################################
25421                         31033
01232367534                   22222100310
5452674261527433              1031003103222222
72262131645122672 55          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
```

Problem B

```
File #0:
Emission Sequence             Probability of Emitting Sequence
######################################################################
25421                         4.537e-05
01232367534                   1.620e-11
5452674261527433              4.348e-15
722621316451226725 5          4.739e-18
02471206023520510102552 41    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
```


```
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
```


Problem C

```
Transition Matrix:
######################################################################
2.830e-01   4.670e-01   1.344e-01   1.156e-01
2.336e-01   3.803e-01   2.934e-01   9.266e-02
1.051e-01   9.873e-02   3.678e-01   4.283e-01
1.887e-01   9.839e-02   3.032e-01   4.097e-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
```

Problem D

```
Transition Matrix:
######################################################################
5.075e-01   4.596e-01   6.533e-09   3.292e-02
3.127e-03   2.107e-04   9.964e-01   2.733e-04
1.195e-09   6.886e-02   9.686e-16   9.311e-01
6.203e-01   3.796e-01   1.555e-05   1.579e-04


Observation Matrix:
######################################################################
1.117e-01   1.525e-01   7.740e-02   1.975e-02   1.594e-01   4.574e-13   3.556e-16   2.475e-01   1.139e-01   1.180e-01
1.205e-01   2.548e-15   1.103e-01   1.751e-01   3.656e-04   2.190e-01   1.002e-01   6.178e-02   1.323e-01   8.053e-02
1.276e-01   2.665e-02   5.788e-02   1.682e-01   1.700e-01   6.969e-02   1.254e-01   3.940e-02   1.627e-01   5.244e-02
1.918e-01   8.206e-02   1.376e-01   8.725e-02   1.152e-01   1.209e-01   1.033e-01   3.101e-02   1.308e-01   5.847e-38
```

Problem E

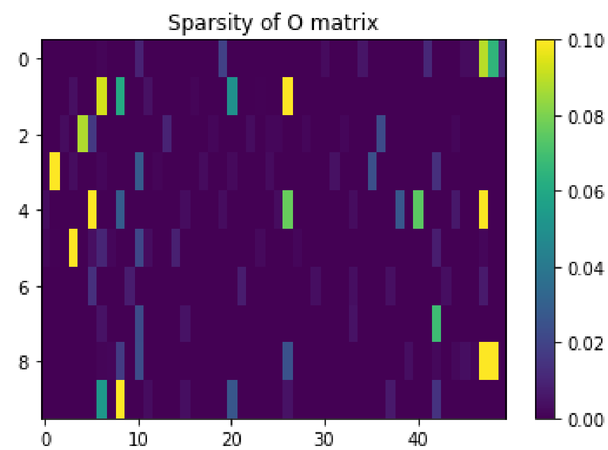
For the transition matrices, it seems like the results from 2C are consistent and larger than the values we see from 2D. More specifically, we see that the smallest order of magnitude in the 2C transition matrix is -2 compared to -16 in the 2D transition matrix. Overall, the transition matrices are quite different from each other considering the greater range of values and and significantly smaller numbers in the matrix from 2D. Furthermore, these trends generally for the observation matrices as well, but it seems like the values are slightly more similar besides the outliers (in 2D) with orders of magnitude as small as -15, -16, -38. Considering these differences, it seems like 2C provides a more accurate representation of Ron's moods and how they affect his moods especially knowing that supervised learning was performed in 2C. This makes sense when we consider that 2C technically uses better data in this sense which seems to help explain the greater consistency and behavior with the resulting matrices.

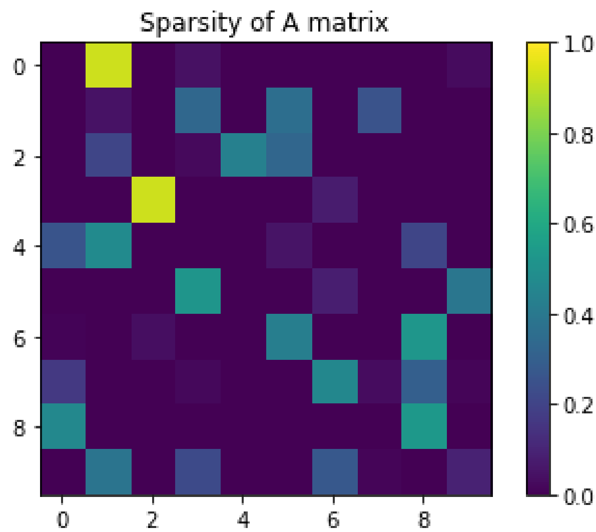
Problem F

```
File #0:
Generated Emission
######################################################################
25457577435433345774
45252137727277572171
02000356747475542452
46452565155154041065
61724341566151764220

File #1:
Generated Emission
######################################################################
60757142001354034547
77213724732517516110
51670575502721512114
24300021407445445277
42167772111611657742

File #2:
Generated Emission
######################################################################
31776532596052487570
53975355193602879761
98269462219799733991
12537762732177220175
78031952085091265614
```

```
File #3:
Generated Emission
######################################################################
01566233364032432166
64605506453443660046
01621664523164325111
10161512266111622505
05331016212162656646

File #4:
Generated Emission
######################################################################
55206544666053110442
10245004242305110466
10656664216663131045
63235465060234442044
45502113533263163446

File #5:
Generated Emission
######################################################################
61332166087454348035
01445615331813636834
70333516061334180842
04586833104655644112
48316363206346316403
```

Problem G





Considering the sparsity of the trained A and O matrices, we see a large majority of values very close to zero (dark purple) for both matrices. With a deeper comparison, we might say that O is more sparse than A since A has a higher concentration of values around .4 (lightish blue). However, both matrices are similar in that that there are only a handful of tiles/states are not values close to zero (not purple). This sparsity would determine the probabilities of transitioning from a certain state to another such that most of the states essentially have probabilities of 0 for transitioning to the next state. On the other hand, we can clearly see the entries that have a high probability of transitioning as indicated by the brighter colorings. Considering the observation behavior, this sparsity would similarly tell us the probability of observing a particular emission such that there is limited behavior as there is a small subset of emissions that are likely to be observed. Overall, the strong sparsity with the trained A and O matrices essentially shows us the limited behaviors for each state.

Problem H

It seems that as the number of hidden states increases, the emission sentences make more sense (grammatically and logically) and sound more like sentences from the constitution. With 1 hidden state, it seems like the words are basically random and the sentence does not really make any sense. This corresponds to what we know about the transition and observation matrices in this case which basically results in randomly selected words independent from each other. This further explains the trend we see since with more hidden states, we are increasing the likelihood of the training data.

Problem I

The state I found semantically meaningful was state 2. Its wordcloud contains several meaningful nouns that are core terms in the constitution such as united, congress, state, president, senate, and even constitution itself. To me, this state represents the core subjects/nouns of the sentences in the constitution, generally. This state differs from others such that other states do not appear to contain a lot of nouns and other meaningful political words. To note, state 6 seems to be similar to state 2 in this sense.