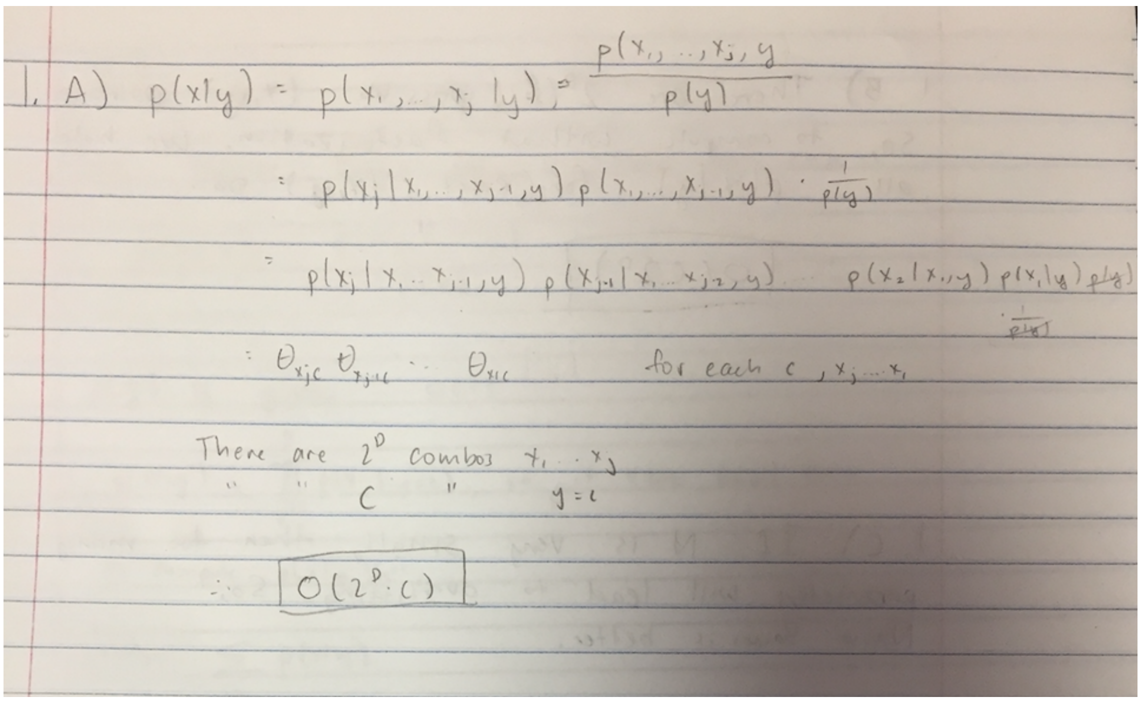
1 Class-Conditional Densities for Binary Data [25 Points]

Problem A [5 points]: Parameters of Full Model with Factorizing

Solution A.:



1. A) $p(x|y) = p(x_1, \dots, x_j | y) = \frac{p(x_1, \dots, x_j, y)}{p(y)}$

$= p(x_j | x_1, \dots, x_{j-1}, y)\, p(x_1, \dots, x_{j-1}, y) \cdot \frac{1}{p(y)}$

$= p(x_j | x_1 \dots x_{j-1}, y)\, p(x_{j-1} | x_1 \dots x_{j-2}, y) \dots p(x_2 | x_1, y)\, p(x_1 | y)\, p(y) \cdot \frac{1}{p(y)}$

$= \theta_{x_j c}\, \theta_{x_{j-1} c} \dots \theta_{x_1 c}$ for each c, $x_j \dots x_1$

There are 2^D combos $x_1 \dots x_j$
" " C " $y = c$

$\therefore \boxed{O(2^D \cdot c)}$

Figure 1

Problem B [5 points]: Parameters of Full Model without Factorizing

Solution B.:

1. B) There are $2^D \cdot C$ possible (x, y) pairs. So, to compute without factorization, we hold all $p(x|y)$ for each (x, y) pair.

$\therefore$ $\boxed{O(C2^D)}$

Figure 2

Problem C [2 points]: Naive Bayes vs. Full Model for Small N

Solution C.:

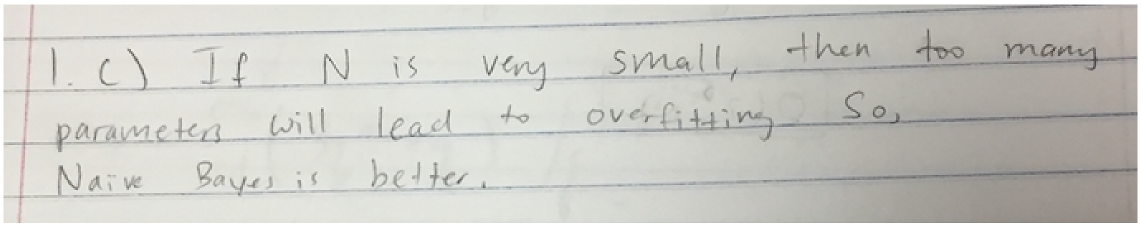
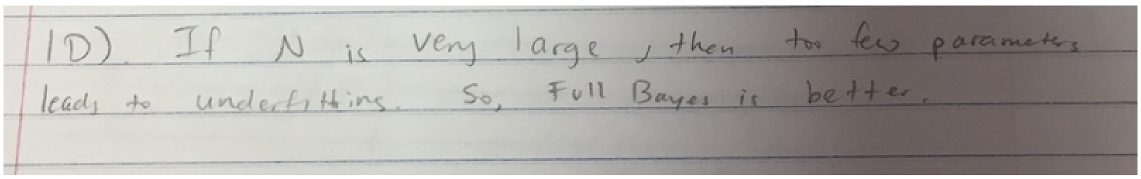

1. c) If N is very small, then too many parameters will lead to overfitting. So, Naive Bayes is better.

Figure 3

Problem D [2 points]: Naive Bayes vs. Full Model for Large N

Solution D.:

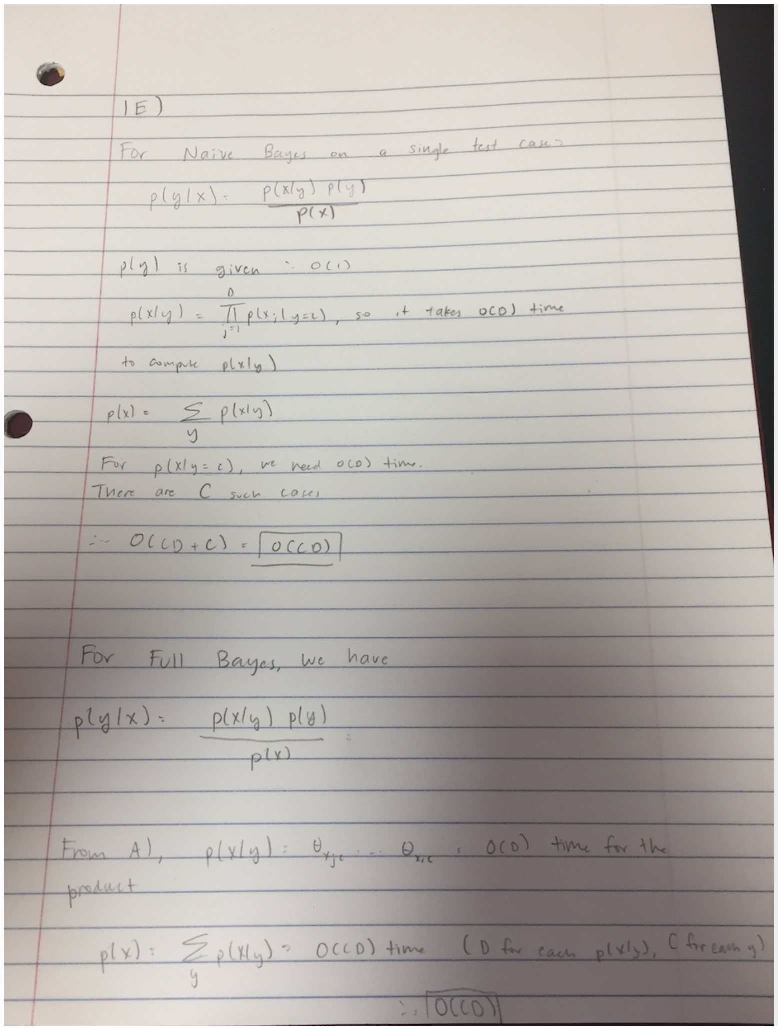


1D). If N is very large, then too few parameters leads to underfitting. So, Full Bayes is better.

Figure 4

Problem E [11 points]: Computational Complexity of Making a Prediction Using Naive Bayes vs Full Model

Solution E.:



1E)

For Naive Bayes on a single test case:

$$p(y|x) = \frac{p(x|y)\,p(y)}{p(x)}$$

$p(y)$ is given $\therefore O(1)$

$p(x|y) = \prod_{j=1}^{D} p(x_j|y=c)$, so it takes $O(D)$ time

to compute $p(x|y)$

$$p(x) = \sum_y p(x|y)$$

For $p(x|y=c)$, we need $O(D)$ time.

There are C such cases

$\therefore O(CD+C) = \boxed{O(CD)}$

For Full Bayes, we have

$$p(y|x) = \frac{p(x|y)\,p(y)}{p(x)}$$

From A), $p(x|y) = \theta_{x_1c} \cdots \theta_{x_Dc}$: $O(D)$ time for the product

$p(x) = \sum_y p(x|y) = O(CD)$ time (D for each $p(x|y)$, C for each y)

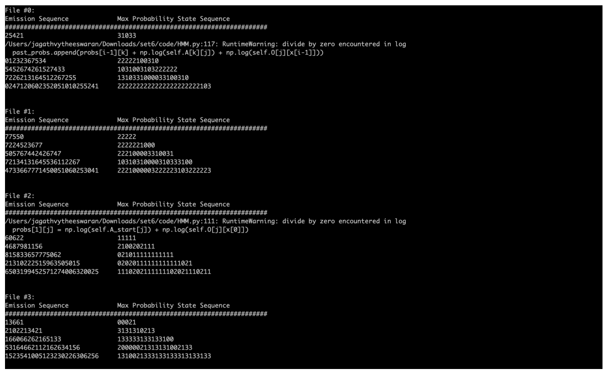
$\therefore \boxed{O(CD)}$

Figure 5

2 Sequence Prediction [75 Points]

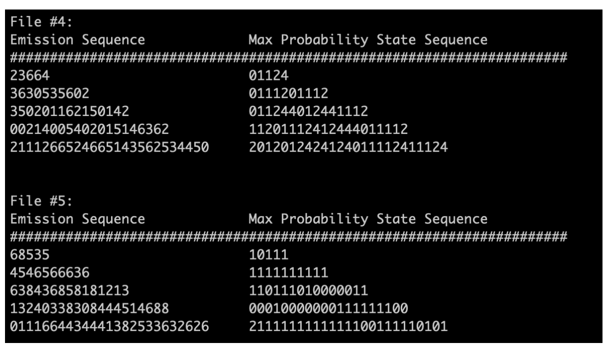
Problem A [10 points]: Max-Probability State Sequences for 6 Trained HMMs

Solution A.:



```
File #0:
Emission Sequence            Max Probability State Sequence
######################################################################
25421                        31033
/Users/jagathvytheeswaran/Downloads/set6/code/HMM.py:117: RuntimeWarning: divide by zero encountered in log
  past_probs.append(probs[i-1][k] + np.log(self.A[k][j]) + np.log(self.O[j][x[i-1]]))
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
######################################################################
/Users/jagathvytheeswaran/Downloads/set6/code/HMM.py:111: RuntimeWarning: divide by zero encountered in log
  probs[1][j] = np.log(self.A_start[j]) + np.log(self.O[j][x[0]])
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
```

Figure 6



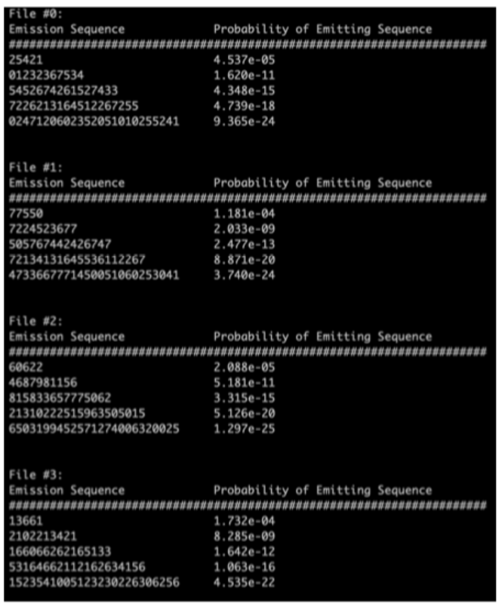
```
File #4:
Emission Sequence            Max Probability State Sequence
######################################################################
23664                        01124
3630535602                   0111201112
350201162150142              011244012441112
00214005402015146362         11201112412444011112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


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

Figure 7

Problem B [17 points]: Probability of Emission Sequence for 6 Trained HMMs

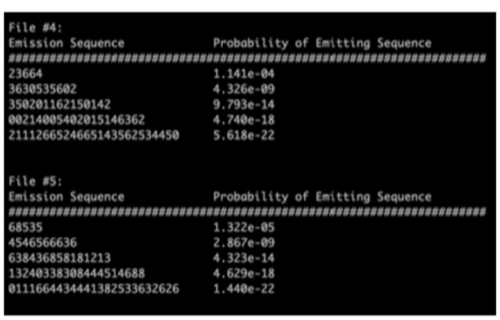
Solution B.:

Alphas on left, Betas on right



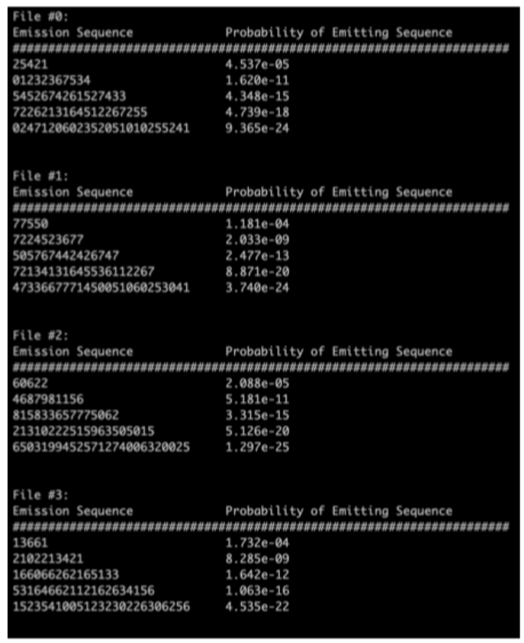
```
File #0:
Emission Sequence             Probability of Emitting Sequence
########################################################################
25421                         4.537e-05
01232367534                   1.620e-11
5452674261527433              4.348e-15
7226213164512267255           4.739e-18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


File #1:
Emission Sequence             Probability of Emitting Sequence
########################################################################
77550                         1.181e-04
7224523677                    2.033e-09
505767442426747               2.477e-13
7213413164553611226 7         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
531646621121626341 56         1.063e-16
1523541005123230226306256     4.535e-22
```

Figure 5



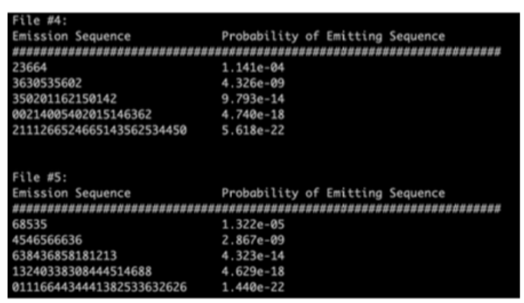
```
File #4:
Emission Sequence             Probability of Emitting Sequence
########################################################################
23664                         1.141e-04
3630535602                    4.326e-09
350201162150142               9.793e-14
00214005402015146362          4.740e-18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

Figure 6




```
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
```


Figure 3




```
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
```


Figure 4

Figure 8

Problem C [10 points]: Learned State Transition and Output Emission Matrices of Supervised Hidden Markov Model

Solution C.:

```
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
```

Figure 9

Problem D [15 points]: Learned State Transition and Output Emission Matrices of Unsupervised Hidden Markov Model

Solution D.:

```
Transition Matrix:
######################################################################
5.413e-06   1.342e-01   8.658e-01   2.379e-08
1.269e-01   3.610e-01   2.221e-02   4.899e-01
3.634e-01   6.366e-01   4.555e-06   3.907e-09
3.501e-02   1.027e-04   3.197e-01   6.452e-01


Observation Matrix:
######################################################################
1.362e-01   7.629e-04   1.634e-01   1.769e-01   6.810e-03   3.249e-01   8.314e-0
3   3.654e-02   9.327e-02   5.301e-02
2.355e-01   1.144e-01   1.697e-01   3.305e-07   1.571e-01   6.108e-15   1.349e-0
1   3.375e-13   1.884e-01   2.590e-05
1.178e-01   6.175e-02   2.302e-41   1.560e-01   1.620e-01   1.034e-01   1.120e-0
1   1.037e-02   1.403e-01   1.363e-01
7.573e-02   6.812e-02   7.632e-02   1.293e-01   8.978e-02   7.933e-02   3.900e-0
2   2.643e-01   1.047e-01   7.342e-02
```

Figure 10: Note: this looks different because I ran it on my friend's computer because mine was performing very very slowly. I also seeded with 2019 (to check with Piazza). I know this seems fishy, but you can confirm the code in my submission, and I have unsupervised runs in the Jupyter notebook as well.

Problem E [5 points]: Compare 2C and 2D

Solution E.: *The variation in the unsupervised matrices is much more extreme; many values are either very close to 0 or decently close to 1 (in a geometric sense; many orders of magnitude larger than others). The variation in the supervised matrices is less extreme.*

Supervised offers a better approximation, as it involves training on observed state-emission pairings.

Unsupervised HMM's performance can be improved by training on additional data.

Problem F [5 points]: Generating Emission Sequences

Solution F.:

```
File #0:
Generated Emission
######################################################################
07325311777124757152
53745057075575217416
13045515722754070475
73045544665376701452
01734233404272007760


File #1:
Generated Emission
######################################################################
14157341770347455754
17214115576711622477
72545435202556554707
76465221562053671444
50274341070441007452


File #2:
Generated Emission
######################################################################
32721226171066715694
33691289911615255999
68230565777551376347
83562633163922007384
65751670269167389530


File #3:
Generated Emission
######################################################################
16224111411616240526
20462010200553040525
35146600612225165310
35026644611056215656
51235135163243411351
```

Figure 11

```
File #4:
Generated Emission
######################################################################
45324205325313104634
61462063126464045101
45633013054143131343
20104103223235150650
40106035636024626061


File #5:
Generated Emission
######################################################################
30684656034184160614
32658642465386882380
66536038883018115844
84438164146884124403
36666044682816316065
```

Figure 12

Problem G [3 points]: Sparsity of Trained A and O Matrices

Solution G.: *A and O seem mostly sparse; that is, each row/column has 0 (or numbers very close to 0) for the majority of its entries.*

This means that, from any state, there are realistically very few other states that are transitioned to, and very few emissions that make sense in the context of this current state.

Problem H [5 points]: Hidden States vs. Sample Emission Sentences from HMM

Solution H.: *In the case where there is one hidden state, the sentence is pretty nonsensical; it's simply a sequence of terms picked from the frequency distribution of words in the document, as with only one state there's no distinguishing between the sampling distribution of words in different positions. As the number of hidden states increases, the sequences of words mirror more closely the kinds of sequences observed in the Constitution. For example, in the most sophisticated model (16 hidden states), we have the phrase "regulate the united at importation declaring," which is pretty sequentially realistic. In the less sophisticated models, the sentences are less and less sequentially logical.*

In general, increasing the number of hidden states will increase the training data likelihood, because it will lead to fitting the training data more closely (and thus mirroring its transitions more closely).

Problem I [5 points]: Analyzing Visualization of State

Solution I.: *State 1 seems to have nouns - more specifically, nouns that represent significant entities in the operation of a government; nouns like "senate," "congress," "state," "office," etc. While other states, like state 9, are also composed primarily of nouns, they're more secondary in nature; state 9, for example, has nouns like "number," "resignation," "consequence," etc.*