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

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

Solution A.:

$$P(x|y) = \prod_{i=1}^{D} p(x_j|x_{i,..,j-1}, y = c) = \prod_{j=1}^{D} \theta_{xjc}$$

Thus, we need to store D different θ_{xjc}'s, each with 2^{j-1} parameters since the probability depends on $j-1$ binary features (x's) and one y that can take on C different values. So

$$\sum_{j=1}^{D} C2^{j-1} = C(2^D - 1) = O(C2^D)$$

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

Solution B.: *We would need to store the estimates for all possible combinations of $P(x|y = c)$. Since there are D binary x features, and one class feature (y) with C classes, there are a total of $C2^D$ possible combinations of $x|y$. Thus, the space complexity is $O(C2^D)$, which is the same as last part.*

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

Solution C.: *For a small training set, the full model will not have enough data points to cover all its probability estimates (as shown in part A and B, this is a large number), thus many of its estimates will be inaccurate, leading to overfitting and high testing error. On the other hand, the naive model will be more accurate because it only requires the probability of observing one x feature given y's value (way easier to obtain from a small set of data), and thus not likely to overfit to the data.*

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

Solution D.: *For a large training set, the naive model will not be accurate since it has strong assumption on independence, which leads to underfitting. The full model is now able to capture more of its estimates accurately and thus perform better on the test set.*

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

Solution E.:
Naive Bayes:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} = \frac{P(y)}{P(x)} \prod_{j=1}^{D} P(x_j|y=c)$$

Since we have a uniform class prior, calculating $P(y)$ is just retrieving a uniform probability, which is $O(1)$. Since we have already estimated each $P(x_j|y=c)$ and stored them, calculating each probability is like retrieving a stored value, which is $O(1)$. And we have D such probabilities, so calculating $\prod_{j=1}^{D} P(x_j|y=c)$ is $O(D)$. For $P(x)$, we know that the probability of observing a specific set of features is equal to the sum of the probability of observing that set of features under each classification, and the probability of observing a specific set of features given a class is equal to the product of the probability of observing each feature given that class (independence assumption). In equation form, that is:

$$P(x) = \sum_{i=1}^{C} P(x|y=C_i) = \sum_{i=1}^{C} \prod_{j=1}^{D} P(x_j|y=C_i)$$

And calculating the above requires $O(CD)$. Thus, the total computation complexity is $1 + D + CD = D(C+1) + 1 = O(CD)$.

Full Model:
With the full model, $P(x|y) = \prod_{j=1}^{D} P(x_j|y=c)$ and $P(y)$ still take the same computation complexity–$O(D)$ and $O(1)$. However, for $P(x) = \sum_{i=1}^{C} P(x|y=C_i)$, we only need to convert the D-bit vector to an array index, and from there we can just lookup the values in the table of probabilities that we computed before. This operation requires $C + D$ computations. With the safe assumption that $C < D$, the total complexity is $D + 1 + C + D = O(D)$

2 Sequence Prediction [75 Points]

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

```
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
7213413164553611226 7          1031031000031033310 0
4733667771450051060253041      2221000003222223103222223

File #2:
Emission Sequence              Max Probability State Sequence
######################################################################
60622                          11111
4687981156                     2100202111
815833657775062                021011111111111
2131022251596350501 5          020201111111111102 1
6503199452571274006320025      1110202111111102021110211

File #3:
Emission Sequence              Max Probability State Sequence
######################################################################
13661                          00021
2102213421                     3131310213
166066262165133                133333133133100
5316466211216263415 6          200000213131310021 33
1523541005123230226306256      1310021333133133313133133

File #4:
Emission Sequence              Max Probability State Sequence
######################################################################
23664                          01124
3630535602                     0111201112
350201162150142                011244012441112
0021400540201514636 2          112011124124440111 12
2111266524665143562534450      2012012424124011112411124

File #5:
Emission Sequence              Max Probability State Sequence
######################################################################
68535                          10111
4546566636                     1111111111
638436858181213                110111010000011
1324033830844451468 8          000100000001111111 00
0111664434441382533632626      2111111111111100111110101
```

Solution A.:

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

Solution B: *Sorry, I am just gonna take a screen shot of the results instead of copy-pasting a thousand time. The forward and backward algorithms produced the same results (as intended), so I will just post the results from one of the algorithms.*

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

Solution C.: *The learned state transition and output emission matrices are the "Transition Matrix" and "Observation Matrix" below, respectively:*

```
##########################################################################
                    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
```

Problem D [15 points]: Learned State Transition and Output Emission Matrices of Unsupervised Hidden Markov Model (use the seeds specified in the Piazza post)

Solution D.:

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

Solution E.:

Problem F [5 points]: Generating Emission Sequences

Solution F.:

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

Solution G.:

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

Solution H.:

Problem I [5 points]: Analyzing Visualization of State

Solution I.: