CMSC 25400 Assignment 4

Rachel Hwang

March 4, 2014

First, let us prove a few useful properties.

1. Suppose X, Y, Z are random variables and $X \perp\!\!\!\perp Y | Z$. We have
$$p(x|y,z) = \frac{p(x,y,z)}{p(y,z)} = \frac{p(x,y|z)p(z)}{p(y,z)} = \frac{p(x|z)p(y|z)p(z)}{p(y,z)} = \frac{p(x|z)p(y,z)}{p(y,z)} = p(x|z).$$
2. Given random variables X, Y, Z, we have
$$p(x,y|z) = \frac{p(x,y,z)}{p(z)} = \frac{p(x|y,z)p(y,z)}{p(z)} = \frac{p(x|y,z)p(y|z)p(z)}{p(z)} = p(x|y,z)p(y|z).$$

(a) If we imagine that y_m has been removed from the model, the expressions for $\alpha_t(i)$ and $\beta_t(i)$, but removing the probabilistic information given by y_m (set $p(y_i|y_m) = 1, \forall i)$), that is

$$\alpha_t(x_t) = \begin{cases} p(y_t|x_t) \sum_{x_{t-1}} p(x_t|x_{t-1})\alpha_{t-1}(x_{t-1}) & \text{if } t \neq m \\ \sum_{x_{t-1}} p(x_t|x_{t-1})\alpha_{t-1}(x_{t-1}) & \text{if } t = m \end{cases}$$

$$\beta_t(x_t) = \begin{cases} \sum_{x_{t+1}} p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1})\beta_{t+1}(x_{t+1}) & \text{if } t+1 \neq m \\ \sum_{x_{t+1}} p(y_{t+1}|x_{t+1})\beta_{t+1}(x_{t+1}) & \text{if } t+1 = m \end{cases}$$

Now, by the law of total probability, we have the equivalence on the first line. The fact that (if $(A \perp\!\!\!\perp B)|C$ then $p(A,B|C) = p(A|C)p(B|C)$) gives the second line. By d-separation (X_m d-separates y_m from the other ys) and the property derived from the definition of conditional independence in 2., we have the next line. This expresses the probability of seeing a particular y_m at time m.

$$\begin{aligned} p(y_m|y_0,\ldots,y_{m-1},y_{m+1},\ldots,y_T) &= \sum_{x_m} p(y_m, x_m|y_0,\ldots,y_{m-1},y_{m+1},\ldots,y_T) \\ &= \sum_{x_m} p(y_m|x_m, y_0,\ldots,y_{m-1},y_{m+1},\ldots,y_T)p(x_m|y_0,\ldots,y_{m-1},y_{m+1},\ldots,y_T) \\ &= \sum_{x_m} p(y_m|x_m)p(x_m|y_0,\ldots,y_{m-1},y_{m+1},\ldots,y_T) \end{aligned}$$

(b) We can derive the desired expression using a handful of properties about independence. First, using that property that $p(A,B|C) = p(A|B,C)p(B|C)$ then the definition of $\gamma_t(x_t)$, we have

$$\begin{aligned} p(x_t, x_{t+1}|y_0,\ldots,y_T) &= p(x_t|y_0,\ldots,y_T)p(x_{t+1}|x_t,y_0,\ldots,y_T) \\ &= \gamma_t(x_t)p(x_{t+1}|x_t,y_{t+1},\ldots,y_T) \end{aligned}$$

We can then apply Bayes' Theorem to derive

$$= \gamma_t(x_t)\frac{p(x_t,y_{t+1},\ldots,y_T|x_{t+1})p(x_{t+1})}{p(x_t,y_{t+1},\ldots,y_T)}$$

By applying the definition of conditional probability, $p(A|B) = p(A,B)/p(B)$, then the definition of $\beta_t(x_t)$, this is equivalent to

$$\begin{aligned} &= \gamma_t(x_t)\frac{p(x_t,y_{t+1},\ldots,y_T|x_{t+1})p(x_{t+1})}{p(y_{t+1},\ldots,y_T|x_t)p(x_t)} \\ &= \gamma_t(x_t)\frac{p(x_t|x_{t+1})p(y_{t+1},\ldots,y_T|x_{t+1})p(x_{t+1})}{\beta_t(x_t)p(x_t)} \end{aligned}$$

Applying the definition of conditional probability twice more, we have

$$= \gamma_t(x_t)\frac{p(y_{t+1},\dots,y_T|x_{t+1})p(x_t,x_{t+1})}{\beta_t(x_t)p(x_t)}$$
$$= \gamma_t(x_t)\frac{p(y_{t+1},\dots,y_T|x_{t+1})p(x_{t+1}|x_t)}{\beta_t(x_t)}$$

Finally, d-separation tells us that $(y_t \perp\!\!\!\perp (y_{t+1},\dots,y_T))|x_{t+1}$. We know that if $(A \perp\!\!\!\perp B)|C$ then $p(A,B|C) = p(A|C)p(B|C)$, which gives us

$$= \gamma_t(x_t)\frac{p(y_{t+2},\dots,y_T|x_{t+1})p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{\beta_t(x_t)}$$
$$= \gamma_t(x_t)\frac{\beta_{t+1}(x_{t+1})p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{\beta_t(x_t)} =: \xi_t(x_t,x_{t+1})$$

(c) The intuitive meaning of the update expression for π_i^{new}, is that we update the probability of starting in state i with the value we calculated that represents the probability that given the observed sequence, we are in state i at time 0. More concisely, this is the likelihood of being in state i at time 0.

The intuitive meaning of the update expression for $\omega_{i,j}^{new}$, the probability that we emit a character j at time i, is the sum(over all times t where the emitted character was j) of the probability that we are in state i over the total sum of the probability we are in state i. This expression means, when we are in state i, what are the odds that we emit j.

The intuitive meaning of the update expression for $\theta_{i,j}^{new}$, the probability that we transition from state i to state j is sum(over times 0-$(T-1)$) of the probability that we transition from i to j given the observed sequence, divided by the sum(over all t) that we are in state i. This expression means, when we are in state i, what are the odds that we are in state j at the next time step.

(d) My implementation is called "hw4/HMM.py". See README for run instructions.

(e) Finding that I was having issues with code runtime (next time, I'll use C), trained my HMM on only a segment of "alice.txt", about 3000 characters. I found that I did need to store my probabilities in log form, and even so occasionally had some overflow errors (very small probabilities -¿ very large negative numbers in log form).
It seems that the likihood of the training data (therefore the accuracy of the model) increases with the number of states, as shown below. 20 states consistently produced the best overall probability.

Plot at the end of the document.

(f) Using the expression from part a, I generated predictions for the corrupted text. Comparing my resulting "corrected" text to the proper Delcaration text, I was able to reduce the number of errors from 87 to 50, using a model where $N_h = 20$, trained for 30 iterations. This is a prediction accuracy of 43%, which is reasonably satisfying considrering that given a random prediction generator, we'd expect to get only about 3/87 correct. I'm certain that my HMM could be trained to do better given a larger training set and many more iterations, but unfortunately I struggled with my implementation, and only had time to gather the included data. My corrected text is as follows.

My error logs, should you care to see them, are below. Note that the HMM was generally good about predicting vowels where there should be vowels, and consonants where there should be consonants.

"when in the course of human evente it beyomes neyessare for one people to dessolve the po itical bands which have connected ther with another and to assume among the powers of the earth the separate and equal station to which the laws of nature and of nature s god ontitle them a decent respect to the opinions of mankind reqlires that they should declare the caures which impel them to the separation we hold these truths to be self avident that all men are created eiual that they are end wed by their creator winh certain nalienable rights that among these are life lilerty and the pirsuit of tappiness that to secure these rights governmeats are instituted among man deriving their jutt powers from the consent of the governed that whenevir any form of government becomes des ructive of these ends it is the right of the people to alter or ta abolish it and to institute new goveynment laying its soundation on such principles and orgalizing its powers in such form as to them shall seem most likely to effect their salety and happiness prudence indeed will dictate that governments long establ shed should not be changed for linht and transient causes and accordingly ale experience hath shewn that mankind are more disposed to suffer thile evils are sufferable than to right themselves by ab lishing the forms to whinh they are accustomed but then a long train of abuses and uturpations pursuang invariable the same object evinces a design to reduce them under absolute despotism it is their right at is their duty to throw off such tovernment and to provide new guarts for their fature security such has been theepatient sufferance of these colenies and such is now the neyessity which constrains them to alter their foumer systems of gover ment the tistory of the present king of great britain is a histore of repeated injuries and usurpations all having in direct object the establishment of an absolute tyrainy over these states to prove this let facts be submitted to a tandid world"

Here is the list of inserted characters: ['e', ' ', 'e', 'y', 'y', 'e', 'e', ' ', ' ', 'r', 'e', 'e', 's', 'h', ' ', 'o', 'h', ' ', 'l', 'r', 's', 'a', 'i', ' ', 'n', ' ', 'l', 'i', 't', 'a', 'a', 't', 'i', ' ', ' ', 'a', 'y', 's', 'l', ' ', 'l', ' ', ' ', 'h', 'h', 'n', 'e', 'h', 'a', 't', ' ', ' ', 'e', ' ', ' ', 'n', ' ', 't', 't', 'a', 'e', ' ', 'e', ' ', 'a', 't', ' ', 'i', 't', 'a', 'e', ' ', 'e', 'y', 'e', 'u', ' ', ' ', 't', 'i', 'n', 'e', ' ', 'e', ' ', 'i', 'h', 't']

(My apologies: I didn't have the time to format them more nicely.)

Error! idx = 33 got e want s
Error! idx = 40 got y want c
Error! idx = 48 got y want c
Error! idx = 54 got e want y
Error! idx = 75 got e want i
Error! idx = 89 got want l
Error! idx = 127 got r want m
Error! idx = 265 got o want e
Error! idx = 325 got l want u
Error! idx = 363 got r want s
Error! idx = 434 got a want e
Error! idx = 468 got i want q
Error! idx = 490 got want o
Error! idx = 514 got n want t
Error! idx = 525 got want i
Error! idx = 572 got l want b
Error! idx = 587 got i want u
Error! idx = 597 got t want h
Error! idx = 643 got a want n
Error! idx = 669 got a want e
Error! idx = 689 got t want s
Error! idx = 743 got i want e
Error! idx = 780 got want t
Error! idx = 846 got a want o
Error! idx = 884 got y want r
Error! idx = 902 got s want f
Error! idx = 940 got l want n
Error! idx = 1023 got l want f
Error! idx = 1099 got want i
Error! idx = 1133 got n want g
Error! idx = 1176 got e want l
Error! idx = 1241 got t want w
Error! idx = 1298 got want o
Error! idx = 1323 got n want c
Error! idx = 1350 got t want w
Error! idx = 1383 got t want s
Error! idx = 1399 got a want i
Error! idx = 1505 got a want i
Error! idx = 1540 got t want g
Error! idx = 1574 got t want d
Error! idx = 1588 got a want u
Error! idx = 1620 got e want
Error! idx = 1652 got e want o
Error! idx = 1680 got y want c
Error! idx = 1727 got u want r

Error! idx = 1748 got want n
Error! idx = 1758 got t want h
Error! idx = 1814 got e want y
Error! idx = 1918 got i want n
Error! idx = 1982 got t want c
ERRORS = 50

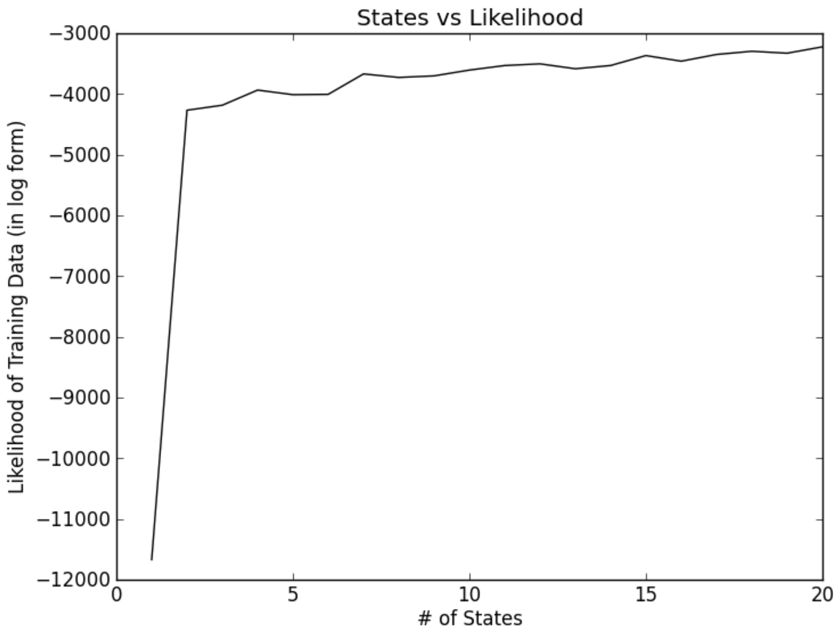
States vs Likelihood
−3000
−4000
−5000
−6000
−7000
−8000
−9000
−10000
−11000
−12000
Likelihood of Training Data (in log form)
0
5
10
15
20
# of States

Figure 1: