Student Name: Guangyu Lin

Collaboration Statement:

Total hours spent: 4 hours

I discussed ideas with these individuals:

- Mingyang Wu
- Xiaohui Chen
- ...

I consulted the following resources:

- Bishop's textbook
- TODO
- ...

By submitting this assignment, I affirm this is my own original work that abides by the course collaboration policy.

Links: [HW5 instructions] [collab. policy]

Contents

1a: Solution 2
1b: Solution 3
2a: Solution 4
2b: Solution 6
2c: Solution 8

1a: Problem Statement

Prove the following property under a Hidden Markov Model.

$$p(z_{t+1}|x_t, z_t) = p(z_{t+1}|z_t) \tag{1}$$

1a: Solution

We can first use product rule to write $p(z_{t+1}|x_t, z_t)$ as

$$\frac{p(z_{t+1}, x_t, z_t)}{p(x_t, z_t)} \tag{2}$$

The numerator term $p(z_{t+1}, x_t, z_t)$ can be written by chain rule as

$$p(z_t)p(z_{t+1}|z_t)p(x_t|z_{t+1}, z_t) \tag{3}$$

The denominator term $p(x_t, z_t)$ can be written by product rule as $p(x_t|z_t)p(z_t)$
By the HMM assumption B, we can regard the $p(x_t|z_{t+1}, z_t)$ as $p(x_t|z_t)$
So we have

$$p(z_{t+1}|x_t, z_t) = \frac{p(z_t)p(z_{t+1}|z_t)p(x_t|z_t)}{p(x_t|z_t)p(z_t)} \tag{4}$$

Now we can cancel the $p(x_t|z_t)p(z_t)$ in numerator and denominator term and get $p(z_{t+1}|z_t)$

1b: Problem Statement

Prove the following property under a Hidden Markov Model.

$$p(x_{t+1}|x_{1:t}, z_{1:t}) = p(x_{t+1}|z_t) \tag{5}$$

1b: Solution

The right side of the statement $p(x_{t+1}|x_{1:t}, z_{1:t})$can be written by sum rule as:

$$\sum_{z_{t+1}} p(x_{t+1}, z_{t+1}|x_{1:t}, z_{1:t}) \tag{6}$$

and expand this term by product rule we can get

$$\sum_{z_{t+1}} \frac{p(x_{t+1}, z_{t+1}, x_{1:t}, z_{1:t})}{p(x_{1:t}, z_{1:t})} \tag{7}$$

now by chain rule the numerator term can be expanded as

$$p(x_{t+1}|x_{1:t}, z_{1:t}, z_{t+1})p(z_{t+1}|x_{1:t}, z_{1:t})p(x_{1:t}, z_{1:t}) \tag{8}$$

from the HMM assumption A we know $p(z_{t+1}|x_{1:t}, z_{1:t})$ is just $p(z_{t+1}|z_t)$ and by HMM assumption B we know $p(x_{t+1}|z_{t+1})$ now we have

$$p(x_{t+1}|z_{t+1})p(z_{t+1}|z_t)p(x_{1:t}, z_{1:t}) \tag{9}$$

Therefore, we can cancel out the same term in both numerator and denominator side and get

$$\sum_{z_{t+1}} p(z_{t+1}|z_t)p(x_{t+1}|z_{t+1}) \tag{10}$$

now we can rewrite the formula as

$$\sum_{z_{t+1}} p(z_{t+1}|z_{t+1}, z_t)p(x_{t+1}|z_{t+1}) \tag{11}$$

and simplify this formula we can get

$$\sum_{z_{t+1}} p(x_{t+1}, z_{t+1}|z_t) = p(x_{t+1}|z_t) \tag{12}$$

2a: Problem Statement

Write out an expression for the expected complete log likelihood:

$$\mathbb{E}_{q(z_{1:T}|s)}\left[\log p(z_{1:T}, x_{1:T}|\theta)\right] \tag{13}$$

Use the HMM probabilistic model $p(z_{1:T}, x_{1:T}|\theta)$ and the approximate posterior $q(z_{1:T}|s)$ defined above.

Your answer should be a function of the data x, the local sequence parameters s and $r(s)$, as well as the HMM parameters π, A, ϕ.

2a: Solution

By substituting the definition of $p(z_{1:T}, x_{1:T}|\theta)$

$$\mathbb{E}_{q(z_{1:T}|s)}[\log p(z_{1:T}|\pi, A) + \log p(x_{1:T}|z_{1:T}, \phi)] \tag{14}$$

now we can substituting with the definition of $p(z_{1:T}|\pi, A)$ and $p(x_{1:T}|z_{1:T}, \phi)$ and we can get

$$\mathbb{E}_{q(z_{1:T}|s)}[\log(\prod_{k=1}^{K}\pi_k^{\delta(z_1,k)}\prod_{t=2}^{T}\prod_{j=1}^{K}\prod_{k=1}^{K}A_{jk}^{\delta(z_{t-1},j)\delta(z_t,k)})+\log(\prod_{t=1}^{T}\prod_{d=1}^{D}\prod_{k=1}^{K}BernPMF(x_{td}|\phi_{kd}^{\delta(z_t,k)})] \tag{15}$$

now just expand this formula we can get

$$\sum_{k=1}^{K}\mathbb{E}_{q(z_{1:T}|s)}[\delta(z_1,k)\log\pi(k)] + \sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K}\mathbb{E}_{q(z_{1:T}|s)}[\delta(z_{t-1},j)\delta(z_t,k))\log A_{jk}] \tag{16}$$

now by the definition of the expectation of $q(Z_{1:T}|S)$ we can rewrite the above first term as

$$\sum_{k=1}^{K} r_1 k(s)\log\pi(k) + \sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} S_{tjk}\log A_{jk} \tag{17}$$

the second term can be written as

$$\sum_{t=1}^{T}\sum_{d=1}^{D}\sum_{k=1}^{K}\mathbb{E}_{q(z_{1:T}|s)}[\delta(z_t,k)(x_{td}\log(\phi_{kd}) + (1-x_{td})\log(1-\phi_{kd})] \tag{18}$$

plug in the definition of the expectation and we get

$$\sum_{t=1}^{T}\sum_{d=1}^{D}\sum_{k=1}^{K}[r_{tk}(s)(x_{td}\log(\phi_{kd})+(1-x_{td})\log(1-\phi_{kd})] \tag{19}$$

so finally we have

$$\sum_{k=1}^{K} r_1 k(s)\log\pi(k) \tag{20}$$

$$+\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} S_{tjk}\log A_{jk} \tag{21}$$

$$+\sum_{t=1}^{T}\sum_{d=1}^{D}\sum_{k=1}^{K}[r_{tk}(s)(x_{td}\log(\phi_{kd})+(1-x_{td})\log(1-\phi_{kd})] \tag{22}$$

2b: Problem Statement

Using your objective function from 2a above, show that for the M-step optimal update to the Bernoulli parameters ϕ_{kd}, the optimal update is given by:

$$\phi_{kd} = \frac{\sum_{t=1}^{T} r_{tk} x_{td}}{\sum_{t=1}^{T} r_{tk}} \tag{23}$$

2b: Solution

Based on 2a we want to get the optimal ϕ_{kd} so we have an objective function like this

$$\underset{\phi_{kd}}{\text{argmax}} \sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{k=1}^{K} [r_{tk}(s)(x_{td} \log(\phi_{kd}) + (1 - x_{td}) \log(1 - \phi_{kd})] \tag{24}$$

To find the optimal value of the objective function, we can do gradient to it

$$\sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{k=1}^{K} [r_{tk}(s) \underset{\phi_{kd}}{\nabla} (x_{td} \log(\phi_{kd}) + (1 - x_{td}) \log(1 - \phi_{kd})] \tag{25}$$

by simplify this formula we can get

$$\sum_{t=1}^{T} [r_{tk}(s)(\frac{x_{td}}{\phi_{kd}} - \frac{1 - x_{td}}{1 - \phi_{kd}})] = 0 \tag{26}$$

by expand the formula we can get

$$\sum_{t=1}^{T} [r_{tk}(s)(\frac{x_{td} - x_{td}\phi_{kd} - \phi_{kd} + \phi_{kd} x_{td}}{\phi_{kd}(1 - \phi_{kd})}] = 0 \tag{27}$$

by simplify the form and multiplied by the denominator for both side we can get

$$\sum_{t=1}^{T} [r_{tk}(s)(x_{td} - \phi_{kd}] = 0 \tag{28}$$

continue simplify we get

$$\sum_{t=1}^{T} r_{tk}(s) x_{td} = \sum_{t=1}^{T} r_{tk}(s) \phi_{kd} \tag{29}$$

now we can get

$$\frac{\sum_{t=1}^{T} r_{tk}(s) x_{td}}{\sum_{t=1}^{T} r_{tk}(s)} = \phi_{kd} \tag{30}$$

2c: Problem Statement

You can find out (by looking up in your textbook) that the optimal update for each entry of A is given by:

$$A_{jk} = \frac{\sum_{t=1}^{T-1} s_{tjk}}{\sum_{t=1}^{T-1} \sum_{k=1}^{K} s_{tjk}}, \quad \text{for } j \in \{1, \ldots, K\}, k \in \{1, \ldots, K\} \tag{31}$$

Provide a short plain English summary of the update for A. How should we interpret the numerator? The denominator?

2c: Solution

The numerator term shows the number of times that expectation of doing transition from state j to specific state k over T time steps. The denominator term shows the number of times that expectation of doing transition from state j to any state (include state k) over T time steps. And the overall updating of A is normalizing the number of times that doing transition from state j to state k over T time steps which allowed A_j can be summed to one.