Student Name: Pengcheng Xu

Collaboration Statement:

Total hours spent: 5 hrs

I discussed ideas with these individuals:

- I did it on my own
- ...

I consulted the following resources:

- Course website
- Course slides
- ...

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: 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

Next, we'll show the derivation step-by-step and also provide the comments alongside.

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

/* By Bayes rule */

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

/* By product rule */

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

/* By assumption B */

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

$$= \frac{p(z_{t+1}, z_t)}{p(z_t)}$$

/* By product rule */

$$= \frac{p(z_t)p(z_{t+1}|z_t)}{p(z_t)}$$

$$= p(z_{t+1}|z_t)$$

Thus, we demonstrated that $p(z_{t+1}|x_t, z_t) = p(z_{t+1}|z_t)$.

1b: Problem Statement

Prove the following property under a Hidden Markov Model.

$$p(x_{t+1}|x_{1:t}, z_{1:t}) = p(x_{t+1}|z_t) \tag{2}$$

1b: Solution

Next, we'll show the derivation step-by-step and also provide the comments alongside.

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

/* By sum rule, sum over all possible values of z_{t+1} */

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

/* By product rule */

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

/* By assumption A and result from 1a, the 1st item could be simplified to $p(z_{t+1}|z_t)$; By assumption B, the 2nd item could be simplified to $p(x_{t+1}|z_{t+1})$ */

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

/* By the definition of conditional probability */

$$= p(x_{t+1}|z_t)$$

Consequently, we've proved that $p(x_{t+1}|x_{1:t}, z_{1:t}) = p(x_{t+1}|z_t)$.

2a: Problem Statement

Write out an expression for the expected complete log likelihood:

$$\mathbb{E}_{q(z_{1:T}|s)}\left[\log p(z_{1:T}, x_{1:T}|\theta)\right] \tag{3}$$

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

Next, we'll show the derivation step-by-step and also provide the comments alongside.

First, let's simplify the complete log likelihood.

$$\log p(z_{1:T}, x_{1:T}|\theta)$$

/* Using HMM probabilistic model defined in the question, and expand it according to the log law */

$$= \log p(z_{1:T}|\pi, A) + \log p(x_{1:T}|z_{1:T}, \varphi)$$

/* Expanding items based on the given definition equations in the question */

$$= \sum_{k=1}^{K} \delta(z_1, k) \log \pi_k + \sum_{t=2}^{T} \sum_{j=1}^{K} \sum_{k=1}^{K} \delta(z_{t-1}, j)\delta(z_t, k) \log A_{jk}$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{k=1}^{K} \delta(z_t, k) \log BernPMF(x_{td}|\varphi_{kd})$$

Then, we'll do the expectation operation on the complete log likelihood, and use its linearity property.

$$E_{q(z_{1:T}|s)} \log p(z_{1:T}, x_{1:T}|\theta)$$

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

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{k=1}^{K} E_{q(z_{1:T}|s)}[\delta(z_t, k)] \log BernPMF(x_{td}|\varphi_{kd})$$

/* Using the result given in the question */

$$= \sum_{k=1}^{K} r_{tk}(s) \log \pi_k + \sum_{t=2}^{T} \sum_{j=1}^{K} \sum_{k=1}^{K} s_{tjk} \log A_{jk}$$

$$+ \sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{k=1}^{K} r_{tk}(s) \log BernPMF(x_{td}|\varphi_{kd})$$

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:

$$\phi_{kd} = \frac{\sum_{t=1}^{T} r_{tk} x_{td}}{\sum_{t=1}^{T} r_{tk}} \tag{4}$$

2b: Solution

Since from the 2a, we know that φ_{kd} only occurs in the last term, the optimal φ_{kd} is equivalent to solving:

$$\arg\max_{\varphi_{kd}} \sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{k=1}^{K} r_{tk}(s) \log BernPMF(x_{td}|\varphi_{kd}) \qquad (*)$$

We solve this problem by taking the first-order derivative in terms of φ_{kd}, and set it equals to zero.

$$\frac{\partial}{\partial \varphi_{kd}} (\sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{k=1}^{K} r_{tk}(s) \log BernPMF(x_{td}|\varphi_{kd})) = 0$$

/* Only one item is related to φ_{kd}. That is, when k = k, d = d. */

$$\frac{\partial}{\partial \varphi_{kd}} (\sum_{t=1}^{T} r_{tk}(s) \log BernPMF(x_{td}|\varphi_{kd})) = 0$$

/* Using definition of Bernoulli PMF, log property, and linearity of derivative */

$$\frac{\partial}{\partial \varphi_{kd}} (\sum_{t=1}^{T} r_{tk}(s) \log(\varphi_{kd}^{x_{td}} \cdot (1 - \varphi_{kd})^{1-x_{td}})) = 0$$

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

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

/* Multiply by $\varphi_{kd}(1-\varphi_{kd})$ on both sides */

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

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

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

/* Then, we could solve for φ_{kd} */

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

Thus, we've proved equation (4).

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:

$$A_{jk} = \frac{\sum_{t=1}^{T-1} s_{tjk}}{\sum_{t=1}^{T-1} \sum_{k=1}^{K} s_{tjk}}, \quad \text{for } j \in \{1, \ldots, K\}, k \in \{1, \ldots, K\} \tag{5}$$

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

2c: Solution

We could interpret the numerator and the denominator like the following:

The numerator means: across all time steps, the number of times (or, the total probability, since s_{tjk} is a probability) we transfer from 'j' to 'k'.

The denominator means: across all time steps, the number of times (or, the total probability, since s_{tjk} is a probability) we transfer from 'j' to all other states.