STAT 534

Lecture 9

HMM Part II, ML Estimation, and Viterbi Algorithm

May 7, 2019

mmp@stat.washington.edu

Scribes: Michael Pearce, Shelly Luo, Sierra Adibi

1 Hidden Markov Models (HMM) Part II

Assume:

1. $q_{1:T}$ hidden $\in \{1, \dots, N\}$,
2. $O_{1:T}$ observed $\in \{1, \dots, M\}$,
3. Model A, B, π known.

Then

$$\alpha_t(i) = P[O_{1:t}, q_t = i]$$

$$\beta_t(i) = P[O_{t+1:T} | q_t = i]$$

$$\gamma_t(i) = P[q_t = i | O_{1:T}] = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i'=1}^{N} \alpha_t(i')\beta_t(i')}$$

$$\begin{aligned} \xi_t(i,j) &= P[q_t = i, q_{t+1} = j | O_{1:T}] \\ &= \frac{\alpha_t(i) \times a_{ij} \times \beta_{t+1}(i) \times b_{j,O_{t+1}}}{P[O_{1:T}]} \end{aligned}$$

The three things we want to do are in the following list:

1. $P[O_{1:T}]$ (in previous lecture notes)
2. ML Estimation of A, B, π, given $O_{1:T}$ (in Section 2 below)
3. Most likely sequence $q_{1:T}$ (started in Section 3 below; to be finished next lecture)

2 Maximum Likelihood (ML) Estimation for Markov Chain

Problem: Given $q_{1:T}$, we wish to estimate A and π.

First, note that

$$\begin{aligned} P[q_{1:T}|\pi, A] &= \pi_{q_1} \Pi_{t=1}^{T-1} a_{q_t,q_{t+1}} \\ &= pi_0 \Pi_{i,j=1}^{N} a_{ij}^{n_{ij}} \end{aligned}$$

where $n_{ij} =$ is the number of transitions from i to j, and $\sum_{i,j=1}^{N} n_{ij} = T - 1$.
We can solve this equation by finding

$$\max_A \sum_{i,j=1}^{N} n_{ij} \log(a_{ij}) + \log(\pi_{q_1})$$

such that

$$\sum_{j=1}^{N} a_{ij} = 1 \text{ for all } i$$

The general solution is given by:

$$\hat{A}_{ij} = \frac{n_{ij}}{\sum_{j'} n_{ij'}}$$

Now, we illustrate the above equations with an example: Suppose we have the sequence $\{0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0\}$, $T = 11$.
Then,

$$\hat{\pi_i} = \begin{cases} 1, & \text{if } q_1 = i, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we know

$$P[q_{1:T}|\pi, A] = \pi_0 a_{01}^2 a_{11}^3 a_{00}^3 a_{10}^2$$

which leads us to the solution

$$\hat{a}_{00} = 3/5, \hat{a}_{01} = 2/5, \hat{a}_{10} = 2/5, \hat{a}_{11} = 3/5$$

Thus the transition matrix is:

$$\hat{A} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{3}{5} \end{bmatrix}$$

Furthermore, note the following formula for the ML estimation of $B|q_{1:T}$:

$$B|q_{1:T} = \max_B \sum_{i=1}^{N} \sum_{k=1}^{M} n'_{ik} \log(b_{ik})$$

such that

$$\sum_{k=1}^{K} b_{ik} = 1 \text{ for all } i$$

which implies that

$$\hat{B}_{ik} = \frac{n'_{ik}}{\sum_{k'} n'_{ik'}}$$

2.1 ML for HMM

Another problem we may have is: Given $O_{1:T}$ and known N, we wish to estimate A, B, π by ML.

Start by writing

$$\begin{aligned}
P[O_{1:T}|A, B, \pi] &= \sum_{q_1} \sum_{q_2} \cdots \sum_{q_T} P[q_{1:T}|A, \pi] P[O_{1:T}|q_{1:T}, B] \\
&= \sum_{q_1} \sum_{q_2} \cdots \sum_{q_T} P[q_{1:T}|A, \pi] \Pi_{t=1}^{T} b_{q_t, O_t} \\
&= \sum_{q_1} \sum_{q_2} \cdots \sum_{q_T} P[q_{1:T}|A, \pi] \Pi_{i=1}^{N} \Pi_{k=1}^{M} b_{ik}^{n'_{ik}}
\end{aligned}$$

So we have formulae for the terms within each sum now. However, calculating the sum is harder. This will be done with the **Baum-Welch Algorithm** (which is an EM Algorithm).

2.2 Baum-Welch Algorithm

1. Initialize: A, B, π

2. E Step: Estimate $E[n_{ij}], E[n'_{ij}] \mid O_{1:T}$

   Note that

$$\begin{aligned}
E[n'_{ik}] &= \sum_{t:O_t=k} P[q_t = i|A, B, \pi, O_{1:T}] \\
&= \sum_{t:O_t=k} \gamma_t(i) \\
E[n_{ij}] &= \sum_{t=1}^{T-1} P[q_t = i, q_{t+1} = j|A, B, \pi, O_{1:T}] \\
&= \sum_{t=1}^{T-1} \xi_t(i, j)
\end{aligned}$$

3. M Step: $\hat{a}_{ij} = \frac{E[n_{ij}]}{E[\sum_{j'} n_{ij'}]}$, $\hat{b}_{ik} = \frac{E[n'_{ik}]}{E[\sum_{k'} n'_{ik}]}$, $\hat{\pi} = P[q_1 = i]$

   Notation: $\Gamma(i) = \sum_{t=1}^{T} \gamma_t(i)$, $\sum_{i=1}^{N} \Gamma(i) = T$.

Note that

$$
\begin{aligned}
\hat{a}_{ij} &= \frac{E[n_{ij}]}{E[\sum_{j'} n_{ij'}]} \\
&= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\
\hat{b}_{ik} &= \frac{E[n'_{ik}]}{E[\sum_{k'} n'_{ik}]} \\
&= \frac{\sum_{t:O_t=k} \gamma_t(i)}{\Gamma(i)} \\
\hat{\pi} &= P[q_1 = i] \\
&= \gamma_1(i)
\end{aligned}
$$

4. Iterate until convergence.

Theorem: Baum-Welch Algorithm converges to a local maximum of the likelihood.

3 Most likely sequence $q_{1:T}$

In this problem, we are given A, B, pi, and $O_{1:T}$. We will use the Viterbi Algorithm (Dynamic Programming) to solve.

3.1 Viterbi Algorithm

$$Pr[q_1 = i | A, B, \pi, O_{1:T}] = \gamma_1(i)$$

For a sequence of length 1, the solution is:

$$q_1^* = argmax_i \gamma_1(i)$$

The algorithm (continued next lecture) should help us find:

$$
\begin{aligned}
&\delta_t(i) = P[q^*_{1:t-1}, q_t = i | A, B, \pi, O_{1:t}] \\
&\psi_t(i)
\end{aligned}
$$

given

$$
\begin{aligned}
\delta_{t-1}(j), &j = 1 : N \\
\psi_{t-1}(j), &j = 1 : N
\end{aligned}
$$