

Solution

1 EM for updating the Markov chain parameters for an HMM

1.1 Updating the initial state distribution $\{\pi\}$ for an HMM

Consider the problem of maximizing the following function

$$Q_{\pi}(\pi; \Psi^{(q)}) = \sum_{k=1}^K \tau_{1k}^{(q)} \log \pi_k$$

with respect to the initial state distribution $\pi = (\pi_1, \dots, \pi_K)$ subject to the constraint $\sum_{k=1}^K \pi_k = 1$, where $\tau_{1k}^{(q)}$ are the posterior probabilities of the initial state k at the q th iteration of EM.

- To perform this constrained maximization, introduce the Lagrange multiplier λ and derive the resulting unconstrained maximization problem (the Lagrangian function).
- To maximize the Lagrangian with respect to π_k ($k = 1, \dots, K$), first set the derivative of the Lagrangian with respect to π_k to zero, determine the Lagrange multiplier λ , and then the resulting value $\pi_k^{(q+1)}$ ($k = 1, \dots, K$) that corresponds to the maximum (the updating formula for the initial state distribution π_k ($k = 1, \dots, K$))

Solution

To perform this constrained maximization, we introduce the Lagrange multiplier λ ; the Lagrangian is then given by:

$$L(\pi_1, \dots, \pi_K) = \sum_{k=1}^K \tau_{1k}^{(q)} \log \pi_k + \lambda \left(\sum_{k=1}^K \pi_k - 1 \right). \quad (1)$$

We then take derivative of the Lagrangian with respect to π_k we obtain:

$$\frac{\partial L(\pi_1, \dots, \pi_K)}{\partial \pi_k} = \frac{\tau_{1k}^{(q)}}{\pi_k} - \lambda, \quad \forall k \in \{1, \dots, K\}. \quad (2)$$

Then, setting these derivatives to zero yields:

$$\frac{\tau_{1k}^{(q)}}{\pi_k} = \lambda, \quad \forall k \in \{1, \dots, K\}. \quad (3)$$

By multiplying each hand side of (3) by π_k ($k = 1, \dots, K$) and summing over k we get

$$\sum_{k=1}^K \frac{\pi_k \times \tau_{1k}^{(q)}}{\pi_k} = \sum_{k=1}^K \lambda \times \pi_k \quad (4)$$

which implies that $\lambda = 1$. Finally, from (3) we get the updating formula for the mixing proportions π_k 's, that is

$$\pi_k^{(q+1)} = \frac{\tau_{1k}^{(q)}}{\lambda} = \tau_{1k}^{(q)}, \quad \forall k \in \{1, \dots, K\}. \quad (5)$$

1.2 Updating the transition probabilities (transition matrix) \mathbf{A} for an HMM

Now consider the problem of maximizing the following function

$$Q_{\mathbf{A}}(\mathbf{A}; \Psi^{(q)}) = \sum_{t=2}^n \sum_{k=1}^K \sum_{l=1}^K \xi_{tlk}^{(q)} \log \mathbf{A}_{lk}$$

with respect to the transition probabilities \mathbf{A}_{lk} subject to the constraint $\sum_{k=1}^K \mathbf{A}_{lk} = 1$, where $\tau_{tk}^{(q)}$ (resp. $\xi_{tlk}^{(q)}$) are the posterior probabilities (resp. the joint posterior probabilities) at the q th iteration of EM.

- To perform this constrained maximization, introduce the Lagrange multiplier λ and derive the resulting unconstrained maximization problem (the Lagrangian function).
- To maximize the Lagrangian with respect to \mathbf{A}_{lk} ($l, k = 1, \dots, K$), first set the derivative of the Lagrangian with respect to \mathbf{A}_{lk} to zero, determine the Lagrange multiplier λ , and then the resulting value $\mathbf{A}_{lk}^{(q+1)}$ ($l, k = 1, \dots, K$) that corresponds to the maximum (the updating formula for the transition matrix).

Solution

To perform this constrained maximization, we follow the same steps as previously. As we have K constraints, we introduce K Lagrange multipliers λ_l for $l = 1, \dots, K$. The Lagrangian is therefore given by:

$$L(\mathbf{A}) = \sum_{t=2}^n \sum_{k=1}^K \sum_{l=1}^K \xi_{tlk}^{(q)} \log \mathbf{A}_{lk} + \sum_{l=1}^K \lambda_l \left(\sum_{k=1}^K \mathbf{A}_{lk} - 1 \right). \quad (6)$$

We then take derivative of the Lagrangian with respect to \mathbf{A}_{lk} we obtain:

$$\frac{\partial L(\mathbf{A})}{\partial \mathbf{A}_{lk}} = \frac{\sum_{t=2}^n \xi_{tlk}^{(q)}}{\mathbf{A}_{lk}} + \lambda_l. \quad (7)$$

Then, setting these derivatives to zero yields:

$$\lambda_l = - \frac{\sum_{t=2}^n \xi_{tlk}^{(q)}}{\mathbf{A}_{lk}} \quad (8)$$

By multiplying each hand side of (8) by \mathbf{A}_{lk} and summing over k we get

$$\sum_{k=1}^K \lambda_l \times \mathbf{A}_{lk} = - \sum_{k=1}^K \sum_{t=2}^n \xi_{tlk}^{(q)} \quad (9)$$

which implies that

$$\lambda_l = - \sum_{k=1}^K \sum_{t=2}^n \xi_{tlk}^{(q)}$$

Finally, from (8) we get the updating formula for the transition probabilities \mathbf{A}_{lk} 's, that is

$$\mathbf{A}_{lk}^{(q+1)} = \frac{- \sum_{t=2}^n \xi_{tlk}^{(q)}}{\lambda_l} = \frac{\sum_{t=2}^n \xi_{tlk}^{(q)}}{\sum_{k=1}^K \sum_{t=2}^n \xi_{tlk}^{(q)}}, \quad \forall l, k = 1, \dots, K. \quad (10)$$

This formula can also be expressed as

$$\mathbf{A}_{lk}^{(q+1)} = \frac{\sum_{t=2}^n \xi_{tlk}^{(q)}}{\sum_{k=1}^K \sum_{t=2}^n \xi_{tlk}^{(q)}} = \frac{\sum_{t=2}^n \xi_{tkl}^{(q)}}{\sum_{t=2}^n \tau_{t\ell}^{(q)}}, \quad \forall l, k = 1, \dots, K. \quad (11)$$

2 Hidden Markov model with discrete observations

Here we consider a hidden Markov model having discrete observations $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ governed by a multivariate Bernoulli distribution. Consider the case where the HMM outputs are multiple binary variables (\mathbf{x}_t is a binary vector in $0, 1^d$); each variable is governed by a Bernoulli conditional distribution.

For the vector \mathbf{x}_t , whose d components are binary For example, for $d = 5$, we can have $\mathbf{x}_t = (1, 0, 0, 1, 0, 1)^T$. Each variable $x_{tj}, j = 1 \dots, d$ is therefore binary and governed by a Bernoulli conditional distribution.

We recall that a binary variable x has a Bernoulli distribution x means

$$p(x) = \begin{cases} \mu & \text{if } x = 1, \\ 1 - \mu & \text{if } x = 0, \end{cases} \quad (12)$$

or equivalently $p(x) = \mu^x (1 - \mu)^{1-x}, x \in \{0, 1\}$

1. by assuming that the variables of each vector \mathbf{x}_t are independent, give the conditional distribution of the observed data $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ given the hidden states at iteration q of the EM algorithm: $\sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(q)} \log p(\mathbf{x}_t | \boldsymbol{\mu}_k)$ where $\mathbf{x}_t = (x_{t1}, \dots, x_{tj}, \dots, x_{td})$ and $\boldsymbol{\mu}_k = (\mu_{k1}, \dots, \mu_{kj}, \dots, \mu_{kd})$ is the parameter of state k
2. give the corresponding M-step updating formula for maximum likelihood solutions of $\{\mu_{kj}\}$

Solution

1.

$$\begin{aligned}
Q(\boldsymbol{\mu}) &= \sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(q)} \log p(\mathbf{x}_t | \boldsymbol{\mu}_k) \\
&= \sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(q)} \log \prod_{j=1}^d p(x_{tj} | \mu_{kj}) \\
&= \sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(q)} \log \prod_{j=1}^d \mu_{kj}^{x_{tj}} (1 - \mu_{kj})^{1-x_{tj}} \\
&= \sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(q)} \sum_{j=1}^d \log \left(\mu_{kj}^{x_{tj}} (1 - \mu_{kj})^{1-x_{tj}} \right) \\
&= \sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(q)} \sum_{j=1}^d (x_{tj} \log \mu_{kj} + (1 - x_{tj}) \log(1 - \mu_{kj}))
\end{aligned} \tag{13}$$

2. By getting the derivative of this function to zero we obtain

$$\frac{\partial Q(\boldsymbol{\mu})}{\partial \mu_{kj}} = \frac{\partial \sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(q)} \sum_{j=1}^d (x_{tj} \log \mu_{kj} + (1 - x_{tj}) \log(1 - \mu_{kj}))}{\partial \mu_{kj}} \tag{14}$$

$$= \sum_{t=1}^n \tau_{tk}^{(q)} \left(\frac{x_{tj}}{\mu_{kj}} - \frac{1 - x_{tj}}{1 - \mu_{kj}} \right) \tag{15}$$

$$= \sum_{t=1}^n \tau_{tk}^{(q)} \left(\frac{x_{tj}(1 - \mu_{kj})}{\mu_{kj}(1 - \mu_{kj})} - \frac{\mu_{kj}(1 - x_{tj})}{\mu_{kj}(1 - \mu_{kj})} \right) \tag{16}$$

$$= \frac{\sum_{t=1}^n \tau_{tk}^{(q)} x_{tj} - \sum_{t=1}^n \tau_{tk}^{(q)} \mu_{kj}}{\mu_{kj}(1 - \mu_{kj})} \tag{17}$$

$$\tag{18}$$

Setting this to zero and solving for μ_{kj} , we get

$$\begin{aligned}
\sum_{t=1}^n \tau_{tk}^{(q)} x_{tj} &= \sum_{t=1}^n \tau_{tk}^{(q)} \mu_{kj} \\
\mu_{kj} &= \frac{\sum_{t=1}^n \tau_{tk}^{(q)} x_{tj}}{\sum_{t=1}^n \tau_{tk}^{(q)}}.
\end{aligned} \tag{19}$$