The aim of this practical work is to show how we learn the HMM parameters with the EM algorithm for sequential data.

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

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

with respect to the initial state distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ 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 qth 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 $(k = 1, \ldots, K)$, first set the derivative of the Lagrangian with respect to π_k to zero, determine the Lagrange multiplier λ, and then the resulting value $\pi_k^{(q+1)}$ $(k = 1, \ldots, K)$ that corresponds to the maximum (the updating formula for the initial state distribution π_k $(k = 1, \ldots, K)$)

1.2 Updating the transition probabilities (transition matrix) A for an HMM

Now consider the problem of maximizing the following function

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

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. $\xi_{t\ell k}^{(q)}$) are the posterior probabilities (resp. the joint posterior probabilities) at the qth 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 $\mathbf{A}_{lk}$ $(l, k = 1, \ldots, K)$, 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 $\mathbf{A}_{lk}^{(q+1)}$ $(l, k = 1, \ldots, K)$ that corresponds to the maximum (the updating formula for the transition matrix.

2 Discrete HMM

Here we consider a hidden Markov model having discrete observations $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ 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 $\mathbb{R}^d$); 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 $\mathbf{x}_t = (0, 1, 0, 1, 0, 1)^T$. Each variable $x_{tj}, j = 1 \ldots, d$ is therefore binary and governed by a Bernoulli conditional distribution.

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

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

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 $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ 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 $\mathbf{x}_t = (x_{t1}, .., x_{tj}, ...x_{td})$ and $\boldsymbol{\mu}_k = (\mu_{k1}, ..., \mu_{kj}, ..., \mu_{kd})$ is the parameter of state k

2. give the corresponding M-step updating formula for maximum likelihood solutions of $\{\mu_{kj}\}$