UNIVERSITY
OF WARSAW

Faculty of Mathematics, Informatics and Mechanics
University of Warsaw

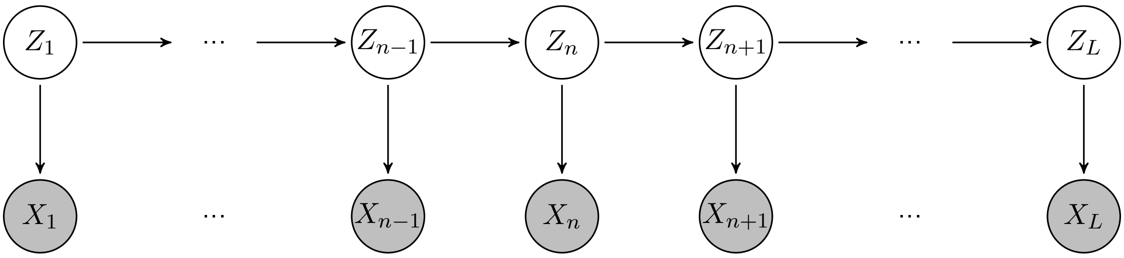
Statistical Data Analysis 2

Exercises # 3

Ewa Szczurek
Lukasz Kozlowski

September 19, 2018

Problem 1: HMMs and EM: Expected Hidden Log Likelihood



Consider the HMM represented by the above graph, where X_n are observed variables and Z_n are latent variables. The Z_n are K-dimensional binary random vectors satisfying $\sum_{k=1}^{K} z_{nk} = 1$ and $z_{nk} \in \{0, 1\}$. The parameters of the HMM are $\theta = \{\pi, A, E\}$. The *initial state probabilities* are given by

$$P(Z_1 = z_1) = \prod_{k=1}^{K} \pi_k^{z_{1k}},$$

where $\sum_k \pi_k = 1$. For the *transition probabilities* with $A = (a_{jk}) \in \mathbb{R}^{K \times K}$ with $0 \leq a_{jk} \leq 1$ and $\sum_k a_{jk} = 1$ we have

$$P(Z_n = z_n \mid Z_{n-1} = z_{n-1}, A) = \prod_{k=1}^{K} \prod_{j=1}^{K} a_{jk}^{z_{n-1,j} z_{nk}}.$$

The *emission probabilities* $P(X_n = x_n \mid Z_n = z_n, E)$ obey a probability distribution parametrized by E, specified later. Given a set of successive observations $x_1, \ldots, x_L$, the EM algorithm can be used to approximate maximum likelihood estimators of all the parameters θ.

Let $\mathrm{E}(\ell_{hid}(\theta))$ shortly denote $\mathrm{E}_{Z|X,\theta}(\ell_{hid}(\theta))$, i.e., the expectation is computed with respect to the posterior probability $P(Z_1, ..., Z_L | X_1, ..., X_L, \theta)$. Write down the factorization for the joint probability distribution $P(X, Z)$ of the HMM and derive that the expected hidden log likelihood $\mathrm{E}(\ell_{hid}(\theta))$ can be written as

$$\mathrm{E}(\ell_{hid}(\theta)) = \left(\sum_{n=1}^{L} \sum_{k=1}^{K} \mathrm{E}(z_{nk}) \log(P(X_n = x_n \mid Z_n = e_k, E)) \right) + \left(\sum_{l=1}^{K} \mathrm{E}(z_{1l}) \log(\pi_l) \right)$$
$$+ \sum_{n=2}^{L} \sum_{k=1}^{K} \sum_{j=1}^{K} \mathrm{E}(z_{n-1,j} z_{nk}) \log(a_{jk}),$$

where e_k is the unit vector with the k th entry equal to 1.

Problem 2: HMMs and EM: The E-Step

Consider the Hidden Markov model defined above. Show that $\mathrm{E}(z_{nk}) = P(Z_n = e_k \mid X_1 = x_1, ..., X_L = x_L, \theta)$ and $\mathrm{E}(z_{n-1,j}z_{nk}) = P(Z_{n-1} = e_j, Z_n = e_k \mid X_1 = x_1, \ldots, X_L = x_L, \theta)$ by exploiting the binary character of Z_n.

Problem 3: HMMs and EM: The M-Step

Consider the Hidden Markov model defined above. In the M-step, maximizing $\mathrm{E}(\ell_{hid}(\theta))$ with respect to π and A (while assuming that $\mathrm{E}(z_{nk})$ and $\mathrm{E}(z_{n-1,j}z_{nk})$ are constant) we get

$$\pi_k = \frac{\mathrm{E}(z_{1k})}{\sum\limits_{k=1}^{K} \mathrm{E}(z_{1k})} \quad \text{and} \quad a_{jk} = \frac{\sum\limits_{n=2}^{L} \mathrm{E}(z_{n-1,j}z_{nk})}{\sum\limits_{n=2}^{L}\sum\limits_{l=1}^{K} \mathrm{E}(z_{n-1,j}z_{nl})}.$$

Show that if any elements of the parameters π or A are initially set to zero, then those elements will remain zero in all subsequent updates of the EM algorithm.
Hint: Given that the posterior probability is

$$P(Z_1, ..., Z_L \mid X_1, ..., X_L, \theta) = \frac{\left(\prod\limits_{n=1}^{L} P(X_n \mid Z_n)\right) P(Z_1) \prod\limits_{n=2}^{L} P(Z_n \mid Z_{n-1})}{C},$$

(where C is a normalization constant), derive expressions for $P(Z_1 \mid X_1, ..., X_L, \theta)$ and $P(Z_{u-1}, Z_u \mid X_1, ..., X_L, \theta)$.

Problem 4: HMMs and EM: Multinomial Emission Probabilities

Consider the Hidden Markov model defined above. If the emission probabilities $E = (p_{jk})$ are *multinomial*, i.e.

$$f_{X_n|Z_n}(x_n \mid Z_n = z_n) = \prod_{j=1}^{m}\prod_{k=1}^{K} p_{jk}^{x_{nj}z_{nk}}$$

with parameters $p_{jk} \in [0, 1]$, show that the maximization step yields

$$p_{jk} = \frac{\sum\limits_{n=1}^{L} \mathrm{E}(z_{nk})x_{nj}}{\sum\limits_{n=1}^{L} \mathrm{E}(z_{nk})}.$$

Problem 5: Modelling the duration of hidden Markov processes

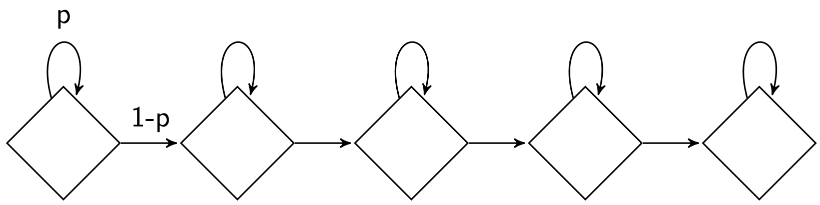
Consider again the hidden Markov model encoded by the graph of Problem 1. Assume that $P(Z_{i+1} = \xi \mid Z_n = \xi) = p$ for a fixed state ξ of the hidden variable. Thus, after entering the state ξ there is a probability of $1 - p$ of leaving it.

(a) Show that the probability of staying in state ξ for l time steps is given by

$$P(Z_i = \xi, Z_{i+1} = \xi, ..., Z_{i+l-1} = \xi, Z_{i+l} \neq \xi) = (1 - p)p^{l-1}.$$

This exponentially decaying distribution on lengths (l) is called *geometric distribution*.

(b) Geometric distribution on lengths can be inappropriate in some applications, where the distribution of lengths is important and significantly different from geometric. A way to circumvent this issue is to model the state ξ with an array of $N \in \mathbb{L}$ identical states in which the transition probabilities are as in the following figure:



Show that for any given path of length l through the model the probability of all its transitions is given by $p^{l-N}(1-p)^L$. Moreover, show that the total probability summed over all possible paths is

$$P(l) = \binom{l-1}{N-1} p^{l-N}(1-p)^L.$$