Imperial College London

COURSEWORK 2 - HIDDEN MARKOV MODELS

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Advanced Statistical Machine Learning and Pattern Recognition - C495

Author:
Georgios Vasilakis (CID: 01445958)

Date: February 27, 2018

1 Exercise I

In this exercise we implemented the Expectation Maximisation algorithm for a Hidden Markov Model for both continuous and discrete variables.
For the continuous (Gaussian) variables our algorithm converged after 51 iterations and the results for 2 hidden states were as following:

Number of iterations: 51
A: [[0.38832547 0.38608991] [0.61167453 0.61391009]]
pi: [0.33392947 0.66607053]
Means: [[0.06975463] [4.97302249]]
Variances: [[0.37220458] [0.91312145]]

For the discrete variables, our algorithm ran for all 100 iterations, and the results we got for 2 hidden states were:

Number of iterations: 100
B : [[0.10979842 0.30528693 0.03692295 0.02734931 0.0801274 0.44051499] [0.13567859 0.0709345 0.15867929 0.15413939 0.14505279 0.33551544]]
pi: [0.23097902 0.76902098]
A: [[0.26247072 0.24292709] [0.73752928 0.75707291]]

2 Exercise II

2.1 Question (i)

In this question we are going to apply the Maximum Likelihood method for the given automaton, which represents a Markov Chain for the word ”kid”.
In this Markov Chain we have 5 states and as a result our parameter π will be a vector of size 5 and our transition matrix A will be a $5x5$ matrix.
We denote our set of parameters as $\theta = \{\pi, A\}$.

Given a set of observations $D_l = \{x_1^l, ..., x_T^l\}$, $l = 1, ..., N$ we want to find the parameters that maximise: $p(D_1, ..., D_N|\theta)$.

$$p(D_1, ..., D_N|\theta = \prod_{l=1}^{N} p(D_l|\theta) \tag{1}$$

$$p(D_l|\theta) = p(x_1^l, ..., x_T^l|\theta) = \tag{2}$$

$$p(x_1^l) \prod_{t=2}^{T} p(x_t^l|x_{t-1}^l = \tag{3}$$

$$\prod_{k=1}^{5} \pi_k^{x_{1k}^l} \prod_{t=2}^{T} \prod_{j=1}^{5} \prod_{k=1}^{5} a_{jk}^{x_{t-1\ j}^l x_{tk}^l}. \tag{4}$$

So, we get that:

$$p(D_1, ..., D_N|\theta) = \prod_{l=1}^{N} \prod_{k=1}^{5} \pi_k^{x_{1k}^l} \prod_{t=2}^{T} \prod_{j=1}^{5} \prod_{k=1}^{5} a_{jk}^{x_{t-1\ j}^l x_{tk}^l} \leftrightarrow \tag{5}$$

$$lnp(\theta) = \sum_{l=1}^{N} \sum_{k=1}^{5} ln\pi_k x_{1k}^l + \sum_{t=2}^{T} \sum_{j=1}^{5} \sum_{k=1}^{5} x_{t-1\ j}^l x_{tk}^l lna_{jk} = \tag{6}$$

$$\sum_{k=1}^{5} (\sum_{l=1}^{N} x_{1k}^l) ln\pi_k \sum_{j=1}^{5} \sum_{k=1}^{5} \sum_{k=1}^{5} (\sum_{l=1}^{N} \sum_{t=2}^{T} x_{t-1\ j}^l x_{tk}^l) lna_{jk} \tag{7}$$

We define the counts as:

$$N_k^1 = \sum_{l=1}^{N} X_{1k}^l \ and \ N_{jk} = \sum_{l=1}^{N} \sum_{t=2}^{T} x_{t-1\ j}^l x_{tk}^l \tag{8}$$

So, equation (4) transforms to:

$$\sum_{k=1}^{5} N_k^1 ln\pi_k \sum_{j=1}^{5} \sum_{k=1}^{5} \sum_{k=1}^{5} N_{jk} lna_{jk} \tag{9}$$

We solve the above problem subject to the following constraints:

$$\sum_{k=1}^{5} \pi_k = 1 \ and \ \sum_{k=1}^{5} a_{jk} = 1 \tag{10}$$

As a result, the Lagrangian we get is:

$$L(\pi, A) = \sum_{k=1}^{5} N_k^1 ln\pi_k + \sum_{j=1}^{5} \sum_{k=1}^{5} N_{jk} lna_{jk} - \lambda(\sum_{k=1}^{5} \pi_{k-1}) - \gamma(\sum_{k=1}^{5} a_{jk-1}) \tag{11}$$

, which finally gives us:

$$\pi_k = \frac{N_k^1}{\sum_{k=1}^{5} N_k^1}, \ a_{jk} = \frac{N_{jk}}{\sum_{k=1}^{5} N_{jk}} \tag{12}$$

2.2 Question (ii)

In this exercise we assume that the number of hidden states is K. Given the rest information, we can infer that our set of parameters consists of the following:

- π, which is an array of size K and contains the prior probabilities of each hidden state
- A, which is a KxK matrix and is our transition matrix
- B, which is a Kx5 matrix, that contains the emission probabilities of each of the 5 discrete variables happening in each hidden state

We consider x to be a set of one-hot vectors as following:

$$x = \{\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\} \tag{13}$$

z is the variable corresponding to the hidden states. So, we have:

$$p(z_1|\pi) = \prod_{c=1}^{K} \pi_c^{z_{1c}} \tag{14}$$

$$p(z_t|z_{t-1}, A) = \prod_{j=1}^{5} \prod_{k=1}^{K} a_{jk}^{z_{t-1\ j} z_{tk}} \tag{15}$$

$$p(x_t|z_t) = \prod_{j=1}^{5} \prod_{k=1}^{K} b_{kj}^{x_{tj} z_{tk}} \tag{16}$$

We want to maximise the probability :

$$p(D_1, D_2, \ldots, D_N, Z_1, \ldots, Z_N|\theta) = \tag{17}$$

$$\prod_{l=1}^{N} p(x_1^l, x_2^l, \ldots, z_1^l, z_2^l, \ldots, z_T^l|\theta) = \tag{18}$$

$$\prod_{l=1}^{N} \prod_{t=1}^{T} \prod_{j=1}^{5} \prod_{k=1}^{K} b_{jk}^{x_{tj}^1 z_{tk}^l} \prod_{k=1}^{K} \pi_k^{z_{1k}^l} \prod_{t=2}^{T} \prod_{j=1}^{K} \prod_{k=1}^{K} a_{jk}^{z_{t-1\ j}^l z_{tk}^l} = (ln) \tag{19}$$

$$\sum_{l=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{5} \sum_{k=1}^{K} x_{tj}^1 z_{tk}^l ln b_{jk} + \tag{20}$$

$$\sum_{l=1}^{N} \sum_{k=1}^{K} z_k^l ln \pi_k + \tag{21}$$

$$\sum_{l=1}^{N} \sum_{t=2}^{T} \sum_{j=1}^{K} \sum_{k=1}^{K} z_{t-1\ j}^l z_{tk}^l ln a_{jk} \tag{22}$$

Taking the expectations with regards to the posterior, equation (20) becomes:

$$\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{5}\sum_{k=1}^{K} x_{tj}^{1} E[z_{tk}^{l}] ln b_{jk} + \tag{23}$$

$$\sum_{l=1}^{N}\sum_{k=1}^{K} E[z_{k}^{l}] ln \pi_{k} + \tag{24}$$

$$\sum_{l=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} E[z_{t-1\ j}^{l} z_{tk}^{l}] ln a_{jk} \tag{25}$$

Now we are going to implement the maximisation step.
The Lagrangian equation with respect to B is :

$$\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{5}\sum_{k=1}^{K} x_{tj}^{1} E[z_{tk}^{l}] ln b_{jk} + \tag{26}$$

$$\sum_{l=1}^{N}\sum_{k=1}^{K} E[z_{k}^{l}] ln \pi_{k} + \tag{27}$$

$$\sum_{l=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} E[z_{t-1\ j}^{l} z_{tk}^{l}] ln a_{jk} + \lambda(\sum_{j=1}^{5} b_{jk} - 1) \tag{28}$$

$$b_{jk} = \frac{\sum_{l=1}^{N}\sum_{t=1}^{T} E[z_{tk}^{l}] x_{tj}^{l}}{\sum_{l=1}^{N}\sum_{t=1}^{T} E[z_{tk}^{l}]} \tag{29}$$

The Lagrangian equation with respect to π is:

$$\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{5}\sum_{k=1}^{K} x_{tj}^{1} E[z_{tk}^{l}] ln b_{jk} + \tag{30}$$

$$\sum_{l=1}^{N}\sum_{k=1}^{K} E[z_{k}^{l}] ln \pi_{k} + \tag{31}$$

$$\sum_{l=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} E[z_{t-1\ j}^{l} z_{tk}^{l}] ln a_{jk} + \lambda(\sum_{k=1}^{K} \pi_{k} - 1) \tag{32}$$

$$\pi_{k} = \frac{\sum_{l=1}^{n} E[z_{1k}^{l}]}{\sum_{l=1}^{N}\sum_{r=1}^{K} E[z_{1r}^{l}]} \tag{33}$$

The Lagrangian equation with respect to A is:

$$\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{5}\sum_{k=1}^{K} x_{tj}^{l} E[z_{tk}^{l}] ln b_{jk} + \tag{34}$$

$$\sum_{l=1}^{N}\sum_{k=1}^{K}E[z_k^l]ln\pi_k+ \tag{35}$$

$$\sum_{l=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K}E[z_{t-1\ j}^l z_{tk}^l]lna_{jk}+\lambda(\sum_{k=1}^{K}a_{jk}-1) \tag{36}$$

$$a_{jk}=\frac{\sum_{l=1}^{N}\sum_{t=2}^{T}E[z_{t-1\ j}^l z_{tk}^l]}{\sum_{r=1}^{K}\sum_{l=1}^{N}\sum_{t=2}^{T}E[z_{t-1\ j}^l z_{tr}^l]} \tag{37}$$