Assignment 5

Maxime CHAMBREUIL
McGill ID: 260067572
maxime.chambreuil@mail.mcgill.ca

Contents

1 Exercise 1 : Expectation Maximization (EM)

1.1 Notations

Assume that the data are sorted with missing data at the beginning and that we have:

- Number of appearances of d_k in the data: m_k
- Number of appearances of s_{il} in the data: m_{il}

1.2 Likelihood

$$
\begin{aligned}
L(data/parameters) &= \prod_{j=1}^{m} a_k^j b_{1lk}^k \cdots b_{nlk}^k \\
&= \prod_{k=1}^{K-1} a_k \times \left(1 - \sum_{k=1}^{K-1} a_k\right)^{m_K} \times \prod_{i=1}^{n} \left[\prod_{l=1}^{L-1} b_{ilk}^{m_{il}} \times \left(1 - \sum_{l=1}^{L-1} b_{ilk}\right)^{m_{iL}}\right]
\end{aligned}
$$

We have gathered the different factor but the last one is expressed with the others to ensure the summation to one, as we are dealing with probabilities.

1.3 Maximum likelihood of the parameters

$$logL = \sum_{k=1}^{K-1} m_k log\left(a_k\right) - m_K log\left(\sum_{k=1}^{K-1} a_k\right) + \sum_{i=1}^{n}\left[\sum_{l=1}^{L-1} m_{il} log\left(b_{ilk}\right) - m_{iL} log\left(\sum_{l=1}^{L-1} b_{ilk}\right)\right]$$

$$\frac{\partial logL}{\partial a_k} = \frac{m_k}{a_k} - \frac{m_K}{\sum_{k=1}^{K-1} a_k} = 0 \Rightarrow \frac{m_k}{a_k} = \frac{m_K}{\sum_{k=1}^{K-1} a_k}$$

$$\frac{\partial logL}{\partial b_{ilk}} = \frac{m_{il}}{b_{ilk}} - \frac{m_{iL}}{\sum_{l=1}^{L-1} b_{ilk}} = 0 \Rightarrow \frac{m_{il}}{b_{ilk}} = \frac{m_{iL}}{\sum_{l=1}^{L-1} b_{ilk}}$$

1.4 Marginal log likelihood

Before computing, the missing data has been put at the beginning of the data, that's why we will multiply from 1 to p:

$$L(data) = \sum_{j=1}^{p}\left[\sum_{k=1}^{K} p(D^j = d_k / S_1^j = s_{1l}, \cdots, S_n^j = s_{nl})\right]$$

$$L(data) = \sum_{j=1}^{p}\sum_{k=1}^{K} \frac{a_k \times b_{ilk}}{p(S_1^j = s_{1l}, \cdots, S_n^j = s_{nl})}$$

1.5 Expectation step

$$\begin{aligned} c_k^j &= p(d_k^j / s_1^j \cdots s_n^j) \\ &= \frac{p(d_k^j) p(s_1^j \cdots s_n^j / d_k)}{p(s_1^j \cdots s_n^j)} \\ &= \frac{a_k \sum_{l=1}^{L} p(s_{1l}^j \cdots s_{nl}^j / d_k^j)}{p(s_1^j \cdots s_n^j)} \\ &= \frac{a_k \sum_{l=1}^{L} b_{ilk}}{p(s_1^j \cdots s_n^j)} \end{aligned}$$

1.6 Expected complete log likelihood

We have basically have to add the log likelihood of the data that were completed with the complete data. Let us first compute the log likelihood of the completed data:

$$\begin{aligned} L(\text{ incomplete data }) &= \sum_{j=1}^{p}\sum_{k=1}^{K} \frac{a_k \times b_{ilk}}{p(S_1^j = s_{1l}, \cdots, S_n^j = s_{nl})} \\ &= \sum_{j=1}^{p}\sum_{k=1}^{K} \frac{a_k \times b_{ilk} \times c_k^j}{p(d_k, S_1^j = s_{1l}, \cdots, S_n^j = s_{nl})} \\ &= \sum_{j=1}^{p}\sum_{k=1}^{K} \frac{a_k \times b_{ilk} \times c_k^j}{a_k \times b_{ilk}} \\ &= \sum_{j=1}^{p}\sum_{k=1}^{K} c_k^j \\ &= \sum_{j=1}^{p}\sum_{k=1}^{K} \frac{a_k \sum_{l=1}^{L} b_{ilk}}{p(s_1^j \cdots s_n^j)} \end{aligned}$$

So the expected complete log likelihood is:

$$logL(\text{ all data }) = \sum_{k=1}^{K-1} m_k log\left(a_k\right) - m_K log\left(\sum_{k=1}^{K-1} a_k\right) + \sum_{i=1}^{n}\left[\sum_{l=1}^{L-1} m_{il} log\left(b_{ilk}\right) - m_{iL} log\left(\sum_{l=1}^{L-1} b_{ilk}\right)\right]$$

$$+log\left[\sum_{j=1}^{p}\sum_{k=1}^{K}\frac{a_k \sum_{l=1}^{L} b_{ilk}}{p(s_1^j \cdots s_n^j)}\right]$$

1.7 Maximization step

1.8 Algorithm

2 Exercise 2 : Hidden Markov Models (HMM)

$$\begin{aligned}
\alpha(S_{t+1}) &= p(S_{t+1}/y_0,\cdots,y_{t+1}) \\
&= \frac{p(S_{t+1},y_0,\cdots,y_{t+1})}{p(y_0,\cdots,y_{t+1})} \\
&= \frac{\sum_{S_t} p(y_0,\cdots,S_t,y_t,S_{t+1},y_{t+1})}{p(y_0,\cdots,y_{t+1})} \\
&= \frac{\sum_{S_t} p(y_{t+1}/S_{t+1})p(S_{t+1}/S_t,y_0,\cdots,y_t)p(S_t,y_0,\cdots,y_t)}{p(y_0,\cdots,y_{t+1})} \\
&= \frac{\sum_{S_t} p(y_{t+1}/S_{t+1})p(S_{t+1}/S_t)p(S_t,y_0,\cdots,y_t)}{p(y_0,\cdots,y_{t+1})} \\
&= \frac{\sum_{S_t} p(y_{t+1}/S_{t+1})p(S_{t+1}/S_t)p(S_t/y_0,\cdots,y_t)p(y_0,\cdots,y_t)}{p(y_{t+1}/y_0,\cdots,y_t)\times p(y_0,\cdots,y_t)} \\
&= \frac{p(y_{t+1}/S_{t+1})\times p(y_0,\cdots,y_t)\times \sum_{S_t} p(S_{t+1}/S_t)\times\alpha(S_t)}{p(y_{t+1}/y_0,\cdots,y_t)\times p(y_0,\cdots,y_t)} \\
&= \frac{p(y_{t+1}/S_{t+1})\times \sum_{S_t} p(S_{t+1}/S_t)\times\alpha(S_t)}{\sum_{S_{t+1}} p(S_{t+1},y_{t+1}/y_0,\cdots,y_t)} \\
&= \frac{p(y_{t+1}/S_{t+1})\times \sum_{S_t} p(S_{t+1}/S_t)\times\alpha(S_t)}{\sum_{S_{t+1}} p(y_{t+1}/S_{t+1})} \\
\alpha(S_{t+1}) &= \frac{p(y_{t+1}/S_{t+1})\times \sum_{S_t} p(S_{t+1}/S_t)\times\alpha(S_t)}{p(y_{t+1})}
\end{aligned}$$

3 Exercise 3 : Problem formulation

- State S_t: position on the map at time t
- Observation y_t: Result given by the sensor at time t, binary variable (Hallway / Intersection).

This problem can be formulated with an Hidden Markov Model (HMM), as we know the state through observations that can be false. As Robby wants to know where it is at time t knowing everything it has observed before (p (S_t / $y_0,\cdots,y_t$)), I would use filtering to solve the inference problem.

Considering the map as a grid, I assume Robby can move to any of 8 states around him with an equal probability:

$$P_{SS'} = \frac{1}{8}$$

Concerning the observation:

$$p(y_t/S_t) = \begin{cases} 0,9 & \text{if the sensor is right} \\ 0,1 & \text{otherwise} \end{cases}$$

4 Exercise 4 : Structural EM

When EM will complete the data, it has no reason to arrange any connection between H and any other variables X. So each network with an additional arc between H and another variable will have the same score as other net with an arc between H and another X. During the remove step, this arc would be removed as it would be the least relevant arc and the best network is the one without connection.

In fact there is 2 choices for the best net: no connection between H and Xs, and all Xs connected to H. The relationship between H and Xs are the same and constant. If this constant is strong enough to be represented by an arc, then the best nets has all arcs; if not the no-connection net remains the best.