

COURSEWORK 2

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

495 - Advanced Statistical Machine Learning and Pattern Recognition Coursework 2

Author:

Jiahao Lin (CID: 00837321)

Date: February 25, 2017

1 Exercise I

- (i) `EM_HMM_continuous_E.m`: This is function for continuous observations Expectation step
`EM_HMM_continuous_M.m`: This is function for continuous observations Maximization step
`EM_HMM_discrete_E.m`: This is function for discrete observations Expectation step
`EM_HMM_discrete_M.m`: This is function for discrete observations Maximization step
`continuous_filtering.m`: This is function for continuous observations Filtering step to calculate alphas and scaling factor
`continuous_smoothing.m`: This is function for continuous observations Smoothing step to calculate betas
`discrete_filtering.m`: This is function for discrete observations Filtering step to calculate alphas and scaling factor
`discrete_smoothing.m`: This is function for discrete observations Smoothing step to calculate betas
To run either the discrete or continuous HMM algorithm, just change folder path to the directory containing these files and then click run in one of the following files:
`example_continuous_observations.m`
`example_discrete_observations.m`
To decide what data is used to initialise the algorithm, you can comment out one setting of parameters like A_e, p_{i_e}, E_e , and uncomment the other set, this mean you can use the original data used to generate observations to initialise the algorithm or other manually set data.
After the algorithm finish, the result will be logged into the command window.
- (ii) `EM_HMM_discrete_viterbi.m`: This is the function for discrete viterbi algorithm
`EM_HMM_continuous_viterbi.m`: This is the function for continuous viterbi algorithm
The viterbi algorithm will automatically be run after corresponding EM algorithm finish, the result accuracy will be logged after running.

2 Exercise II

- (i) N sequences, T observations per sequence, 5 states
Estimate $\theta = \{\pi, A\}$
 $\pi = \{\pi_1 \dots \pi_5\}$
for $l = 1..N : D_l = \{x_1^l \dots x_T^l\}$
 $A = [a_{ij}]$

$$A = \begin{bmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & a_{34} \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & a_{55} \end{bmatrix} \quad (1)$$

$$x_t^l = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\} \quad (2)$$

Maximise $p(D_1 \dots D_N | \theta)$

$$p(D_1 \dots D_N | \theta) \quad (3)$$

$$= \prod_{l=1}^N p(D_l | \theta) \quad (4)$$

Because

$$p(D_1 | \theta) \quad (5)$$

$$= p(x_1^l \dots x_T^l | \theta) \quad (6)$$

$$= p(x_1^l) \prod_{t=2}^T p(x_t^l | x_{t-1}^l) \quad (7)$$

$$= \prod_{k=1}^5 \pi_k^{x_{1k}^l} \prod_{T=2}^T \prod_{j=1}^5 \prod_{k=1}^5 a_{jk}^{x_{t-1j}^l x_{tk}^l} \quad (8)$$

so

$$p(D_1 \dots D_N | \theta) \quad (9)$$

$$= \prod_{l=1}^N \prod_{k=1}^5 \pi_k^{x_{1k}^l} \prod_{T=2}^T \prod_{j=1}^5 \prod_{k=1}^5 a_{jk}^{x_{t-1j}^l x_{tk}^l} \quad (10)$$

Take the log

$$\ln(p(D_1 \dots D_N | \theta)) \quad (11)$$

$$= \sum_{l=1}^N \sum_{k=1}^5 x_{1k}^l \ln(\pi_k) + \sum_{l=1}^N \sum_{T=2}^T \sum_{j=1}^5 \sum_{k=1}^5 x_{t-1j}^l x_{tk}^l \ln(a_{jk}) \quad (12)$$

Now we take

$$N_k^l = \sum_{l=1}^N x_{1k}^l \quad (13)$$

$$N_{jk} = \sum_{l=1}^N \sum_{T=2}^T x_{t-1j}^l x_{tk}^l \quad (14)$$

so

$$\ln(p(D_1 \dots D_N | \theta)) \quad (15)$$

$$= \sum_{l=1}^N N_k^l \ln(\pi_k) + \sum_{j=1}^5 \sum_{k=1}^5 N_{jk} \ln(a_{jk}) \quad (16)$$

When maximising $\ln(p(D_1 \dots D_N | \theta))$, we put constraints:

$$\sum_{k=1}^K \pi_k = 1 \quad (17)$$

$$\sum_{k=1}^K a_{jk} = 1 \quad (18)$$

So

$$L(\pi, A) = \sum_{l=1}^N N_k^l \ln(\pi_k) + \sum_{j=1}^5 \sum_{k=1}^5 N_{jk} \ln(a_{jk}) - \lambda \left(\sum_{k=1}^K \pi_k - 1 \right) - \gamma \left(\sum_{k=1}^K a_{jk} - 1 \right) \quad (19)$$

Maximise $L(\pi, A)$:

$$\frac{\partial L(\pi, A)}{\partial \pi_k} = \frac{N_k^l}{\pi_k} - \lambda \quad (20)$$

$$\frac{\partial L(\pi, A)}{\partial \pi_k} = 0 \quad (21)$$

$$\lambda = \frac{N_k^l}{\pi_k} \quad (22)$$

$$\sum_{k=1}^K \pi_k = 1 = \sum_{k=1}^K \frac{N_k^l}{\lambda} \quad (23)$$

$$\lambda = \sum_{k=1}^K N_k^l \quad (24)$$

$$\pi_k = \frac{N_k^l}{\sum_{k=1}^K N_k^l} \quad (25)$$

$$\frac{\partial L(\pi, A)}{\partial a_{jk}} = \frac{N_{jk}}{a_{jk}} - \gamma \quad (26)$$

$$\gamma = \sum_{k=1}^K N_{jk} \quad (27)$$

$$a_{jk} = \frac{N_{jk}}{\sum_{k=1}^K N_{jk}} \quad (28)$$

Given our matrix A , we know that:

$$a_{12} = 1 \quad (29)$$

$$a_{22} = \frac{N_{22}}{\sum_{k=1}^K N_{2k}} \quad (30)$$

$$a_{23} = 1 - a_{22} = \frac{N_{22}}{\sum_{k=1}^K N_{2k}} \quad (31)$$

$$a_{33} = \frac{N_{33}}{\sum_{k=1}^K N_{3k}} \quad (32)$$

$$a_{34} = \frac{N_{34}}{\sum_{k=1}^K N_{3k}} \quad (33)$$

$$a_{35} = 1 - a_{34} - a_{33} = 1 - \frac{N_{33} + N_{34}}{\sum_{k=1}^K N_{3k}} \quad (34)$$

$$a_{44} = \frac{N_{44}}{\sum_{k=1}^K N_{4k}} \quad (35)$$

$$a_{45} = 1 - a_{44} = 1 - \frac{N_{44}}{\sum_{k=1}^K N_{4k}} \quad (36)$$

$$a_{55} = 1 \quad (37)$$

$$\pi_k = \frac{N_k^l}{\sum_{k=1}^K N_k^l} \quad (38)$$

(ii) K latent states, N sequences, T observations per sequence, 5 Observation state

$$p(z_1^l | \pi) = \prod_{c=1}^K \pi_k^{z_{1c}^l} \quad (39)$$

$$p(z_t^l | z_{t-1}^l, A) = \prod_{j=1}^K \prod_{k=1}^K a_{jk}^{z_{t-1j}^l z_{tk}^l} \quad (40)$$

$$p(z_{1k}^l = 1) = \pi_k \quad (41)$$

$$\pi = \{\pi_1 \dots \pi_5\} \quad (42)$$

$$A = [a_{ij}] \quad (43)$$

$$p(x_t^l | z_t^l) = \prod_{j=1}^5 \prod_{k=1}^K b_{kj}^{x_{t-1j}^l x_{tk}^l} \quad (44)$$

$$\text{Estimate } \theta = \{\pi, A, B\} \quad (45)$$

$$\text{for } l = 1..N : D_l = \{x_1^l \dots x_T^l\} \quad (46)$$

Maximise $p(D_1 \dots D_N, Z_1 \dots Z_N | \theta)$

$$x_t^l = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\} \quad (47)$$

$$p(D_1 \dots D_N, Z_1 \dots Z_N | \theta) = \prod_{l=1}^N p(x_1^l \dots x_T^l, z_1^l \dots z_T^l | \theta) \quad (48)$$

$$= \prod_{l=1}^N \prod_{t=1}^T p(x_t^l | z_t^l) p(z_1^l) \prod_{t=2}^T p(z_t^l | z_{t-1}^l) \quad (49)$$

$$= \prod_{l=1}^N \prod_{t=1}^T \prod_{j=1}^5 \prod_{k=1}^K b_{kj}^{x_{tj}^l z_{tk}^l} \prod_{k=1}^K \pi_k^{z_{1k}^l} \prod_{t=2}^T \prod_{j=1}^K \prod_{k=1}^K a_{jk}^{z_{t-1j}^l z_{tk}^l} \quad (50)$$

Take the log of this posterior

$$\ln(p(D_1 \dots D_N, Z_1 \dots Z_N | \theta)) \quad (51)$$

$$= \sum_{l=1}^N \sum_{t=1}^T \sum_{j=1}^5 \sum_{k=1}^K x_{tj}^l z_{tk}^l \ln(b_{kj}) + \sum_{l=1}^N \sum_{k=1}^K z_{1k}^l \ln(\pi_k) + \sum_{l=1}^N \sum_{t=2}^T \sum_{j=1}^K \sum_{k=1}^K z_{t-1j}^l z_{tk}^l \ln(a_{jk}) \quad (52)$$

Take expectation with respect to posterior

$$E_{[z_{tk}^l]}[\ln(p(D, Z | \theta))] = \sum_{l=1}^N \sum_{t=1}^T \sum_{j=1}^5 \sum_{k=1}^K x_{tj}^l E[z_{tk}^l] \ln(b_{kj}) \quad (53)$$

$$+ \sum_{l=1}^N \sum_{k=1}^K E[z_{1k}^l] \ln(\pi_k) + \sum_{l=1}^N \sum_{t=2}^T \sum_{j=1}^K \sum_{k=1}^K E[z_{t-1j}^l z_{tk}^l] \ln(a_{jk}) \quad (54)$$

To maximise it we need to put constraints

$$\sum_{j=1}^5 b_{jk} = 1 \quad (55)$$

$$\sum_{k=1}^5 \pi_k = 1 \quad (56)$$

$$\sum_{k=1}^K a_{jk} = 1 \quad (57)$$

So

$$L(\pi, \mathbf{A}, \mathbf{B}) = \sum_{l=1}^N \sum_{t=1}^T \sum_{j=1}^5 \sum_{k=1}^K x_{tj}^l E[z_{tk}^l] \ln(b_{jk}) + \sum_{l=1}^N \sum_{k=1}^K E[z_{1k}^l] \ln(\pi_k) \quad (58)$$

$$+ \sum_{l=1}^N \sum_{t=2}^T \sum_{j=1}^K \sum_{k=1}^K E[z_{t-1j}^l z_{tk}^l] \ln(a_{jk}) - \lambda \left(\sum_{j=1}^5 b_{jk} - 1 \right) - \gamma \left(\sum_{k=1}^K \pi_k - 1 \right) - \mu \left(\sum_{k=1}^K a_{jk} - 1 \right) \quad (59)$$

Maximise $L(\pi, \mathbf{A}, \mathbf{B})$:

$$\frac{\partial L(\pi, \mathbf{A}, \mathbf{B})}{\partial b_{jk}} = \frac{\sum_{l=1}^N \sum_{t=1}^T x_{tj}^l E[z_{tk}^l]}{b_{jk}} - \lambda = 0 \quad (60)$$

$$\lambda = \frac{\sum_{l=1}^N \sum_{t=1}^T x_{tj}^l E[z_{tk}^l]}{b_{jk}} \quad (61)$$

$$b_{jk} = \frac{\sum_{l=1}^N \sum_{t=1}^T x_{tj}^l E[z_{tk}^l]}{\sum_{l=1}^N \sum_{t=1}^T \sum_{j=1}^5 x_{tj}^l E[z_{tr}^l]} \quad (62)$$

$$\frac{\partial L(\pi, \mathbf{A}, \mathbf{B})}{\partial \pi_k} = \frac{\sum_{l=1}^N \sum_{k=1}^K E[z_{1k}^l]}{\pi_k} - \gamma = 0 \quad (63)$$

$$\gamma = \frac{\sum_{l=1}^N \sum_{k=1}^K E[z_{1k}^l]}{\pi_k} \quad (64)$$

$$\pi_k = \frac{\sum_{l=1}^N E[z_{1k}^l]}{\sum_{l=1}^N \sum_{r=1}^K E[z_{1r}^l]} \quad (65)$$

$$\frac{\partial L(\pi, \mathbf{A}, \mathbf{B})}{\partial a_{jk}} = \frac{\sum_{l=1}^N \sum_{t=2}^T E[z_{t-1j}^l z_{tk}^l]}{a_{jk}} - \mu = 0 \quad (66)$$

$$a_{jk} = \frac{\sum_{l=1}^N \sum_{t=2}^T E[z_{t-1j}^l z_{tk}^l]}{\sum_{l=1}^N \sum_{t=2}^T \sum_{r=1}^K E[z_{t-1j}^l z_{tr}^l]} \quad (67)$$

Answer is

$$b_{jk} = \frac{\sum_{l=1}^N \sum_{t=1}^T x_{tj}^l E[z_{tk}^l]}{\sum_{l=1}^N \sum_{t=1}^T \sum_{j=1}^5 x_{tj}^l E[z_{tr}^l]} \quad (68)$$

$$\pi_k = \frac{\sum_{l=1}^N E[z_{1k}^l]}{\sum_{l=1}^N \sum_{r=1}^K E[z_{1r}^l]} \quad (69)$$

$$a_{jk} = \frac{\sum_{l=1}^N \sum_{t=2}^T E[z_{t-1j}^l z_{tk}^l]}{\sum_{l=1}^N \sum_{t=2}^T \sum_{r=1}^K E[z_{t-1j}^l z_{tr}^l]} \quad (70)$$