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 22, 2017

1 Exercise I

1.a

fdss

1.b

(ii) fsds

2 Exercise II

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

$$\boldsymbol{A} = \begin{vmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & a_{34} \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & a_{55} \end{vmatrix} x_t^l = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\} \tag{1}$$

$$\tag{2}$$

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

$$p(D_1 ... D_N | \theta) \tag{3}$$

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

$$\tag{5}$$

Because

$$p(D_1 | \theta) \tag{6}$$

$$= p(x_1^l ... x_T^l | \theta) \tag{7}$$

$$= p(x_1^l) \prod_{t=2}^{T} p(x_t^l | x_{t-1}^l) \tag{8}$$

$$= \prod_{k=1}^{5} \pi_k^{x_{1k}^l} \prod_{T=2}^{T} \prod_{j=1}^{5} \prod_{k=1}^{5} a_{jk}^{x_{t-1j}^l x_{tk}^l} \tag{9}$$

$$ \tag{10} $$

so

$$ p(D_1 ... D_N|\theta) \tag{11} $$

$$ = \prod_{l=1}^{N} \prod_{k=1}^{5} \pi_k^{x_{1k}^l} \prod_{T=2}^{T} \prod_{j=1}^{5} \prod_{k=1}^{5} a_{jk}^{x_{t-1j}^l x_{tk}^l} \tag{12} $$

$$ \tag{13} $$

Take the log

$$ \ln(p(D_1 ... D_N|\theta)) \tag{14} $$

$$ = \sum_{l=1}^{N} \sum_{k=1}^{5} x_{1k}^l \ln(\pi_k) + \sum_{l=1}^{N} \sum_{T=2}^{T} \sum_{j=1}^{5} \sum_{k=1}^{5} x_{t-1j}^l x_{tk}^l \ln(a_{jk}) \tag{15} $$

$$ \tag{16} $$

Now we take

$$ N_k^l = \sum_{l=1}^{N} x_{1k}^l \tag{17} $$

$$ N_{jk} = \sum_{l=1}^{N} \sum_{T=2}^{T} x_{t-1j}^l x_{tk}^l \tag{18} $$

$$ \tag{19} $$

so

$$ \ln(p(D_1 ... D_N|\theta)) \tag{20} $$

$$ = \sum_{l=1}^{N} N_k^l \ln(\pi_k) + \sum_{j=1}^{5} \sum_{k=1}^{5} N_{jk} \ln(a_{jk}) \tag{21} $$

$$ \tag{22} $$

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

$$ \sum_{k=1}^{K} \pi_k = 1 \tag{23} $$

$$ \sum_{k=1}^{K} a_{jk} = 1 \tag{24} $$

$$ \tag{25} $$

So

$$ \boldsymbol{L}(\boldsymbol{\pi}, \boldsymbol{A}) = \sum_{l=1}^{N} N_k^l \ln(\pi_k) + \sum_{j=1}^{5} \sum_{k=1}^{5} N_{jk} \ln(a_{jk}) - \lambda(\sum_{k=1}^{K} \pi_k - 1) - \gamma(\sum_{k=1}^{K} a_{jk} - 1) \tag{26} $$

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

$$\frac{\partial \boldsymbol{L}(\boldsymbol{\pi}, \boldsymbol{A})}{\partial \pi_k} = \frac{N_k^l}{\pi_k} - \lambda \tag{27}$$

$$\frac{\partial \boldsymbol{L}(\boldsymbol{\pi}, \boldsymbol{A})}{\partial \pi_k} = 0 \tag{28}$$

$$\lambda = \frac{N_k^l}{\pi_k} \tag{29}$$

$$\sum_{k=1}^{K} \pi_k = 1 = \sum_{k=1}^{K} \frac{N_k^l}{\lambda} \tag{30}$$

$$\lambda = \sum_{k=1}^{K} N_k^l \tag{31}$$

$$\pi_k = \frac{N_k^l}{\sum_{k=1}^{K} N_k^l} \tag{32}$$

$$\frac{\partial \boldsymbol{L}(\boldsymbol{\pi}, \boldsymbol{A})}{\partial a_{jk}} = \frac{N_{jk}}{a_{jk}} - \gamma \tag{33}$$

$$\gamma = \sum_{k=1}^{K} N_{jk} \tag{34}$$

$$a_{jk} = \frac{N_{jk}}{\sum_{k=1}^{K} N_{jk}} \tag{35}$$

$$\tag{36}$$

Given our matrix $\boldsymbol{A}$, we know that:

$$a_{12} = 1 \tag{37}$$

$$a_{22} = \frac{N_{22}}{\sum_{k=1}^{K} N_{2k}} \tag{38}$$

$$a_{23} = 1 - a_{22} = \frac{N_{22}}{\sum_{k=1}^{K} N_{2k}} \tag{39}$$

$$a_{33} = \frac{N_{33}}{\sum_{k=1}^{K} N_{3k}} \tag{40}$$

$$a_{34} = \frac{N_{34}}{\sum_{k=1}^{K} N_{3k}} \tag{41}$$

$$a_{35} = 1 - a_{34} - a_{33} = 1 - \frac{N_{33} + N_{34}}{\sum_{k=1}^{K} N_{3k}} \tag{42}$$

$$a_{44} = \frac{N_{44}}{\sum_{k=1}^{K} N_{4k}} \tag{43}$$

$$a_{45} = 1 - a_{44} = 1 - \frac{N_{44}}{\sum_{k=1}^{K} N_{4k}} \tag{44}$$

$$a_{55} = 1 \tag{45}$$

$$\pi_k = \frac{N_k^l}{\sum_{k=1}^{K} N_k^l} \tag{46}$$

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

$$p(z_1^l|\pi) = \prod_{c=1}^{K} \pi_k^{z_{1c}^l} \tag{47}$$

$$p(z_t^l|z_{t-1}^l, A) = \prod_{j=1}^{K}\prod_{k=1}^{K} a_{jk}^{z_{t-1j}^l z_{tk}^l} \tag{48}$$

$$p(z_{1k}^l = 1) = \pi_k \tag{49}$$

$$\boldsymbol{\pi} = \{\pi_1 ... \pi_5\} \tag{50}$$

$$\boldsymbol{A} = [a_{ij}] \tag{51}$$

$$p(x_t^l|z_t^l) = \prod_{j=1}^{5}\prod_{k=1}^{K} b_{kj}^{x_{t-1j}^l x_{tk}^l} \tag{52}$$

$$Estimate \quad \theta = \{\boldsymbol{\pi}, A, \boldsymbol{B}\} \tag{53}$$

$$for \quad l = 1..N : D_l = \left\{x_1^l ... x_T^l\right\} \tag{54}$$

$$\tag{55}$$

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

$$x_t^l = \left\{\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\right\} \tag{56}$$

$$p(D_1 ... D_N, Z_1 ... Z_N|\theta) = \prod_{l=1}^{N} p(x_1^l ... x_T^l, z_1^l ... z_T^l|\theta) \tag{57}$$

$$= \prod_{l=1}^{N}\prod_{t=1}^{T} p(x_t^l|z_t^l) p(z_1^l) \prod_{t=2}^{T} p(z_t^l|z_{t-1}^l) \tag{58}$$

$$= \prod_{l=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{5}\prod_{k=1}^{K} b_{kj}^{x_{tj}^l z_{tk}^l} \prod_{k=1}^{K} \pi_k^{z_{1k}^l} \prod_{t=2}^{T}\prod_{j=1}^{K}\prod_{k=1}^{K} a_{jk}^{z_{t-1j}^l z_{tk}^l} \tag{59}$$

$$\tag{60}$$

Take the log of this posterior

$$\ln(p(D_1 ... D_N, Z_1 ... Z_N|\theta))$$

$$\tag{61}$$

$$= \sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{5}\sum_{k=1}^{K} x_{tj}^{l} z_{tk}^{l} \ln(b_{kj}) + \sum_{l=1}^{N}\sum_{k=1}^{K} z_{1k}^{l} \ln(\pi_k) + \sum_{l=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} z_{t-1j}^{l} z_{tk}^{l} \ln(a_{jk}) \tag{62}$$

(63)

Take expectation with respect to posterior

$$E_{[z_{tk}^{l}]}[\ln(p(D,Z|\theta))] = \sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{5}\sum_{k=1}^{K} x_{tj}^{l} E[z_{tk}^{l}] \ln(b_{kj}) \tag{64}$$

$$+ \sum_{l=1}^{N}\sum_{k=1}^{K} E[z_{1k}^{l}] \ln(\pi_k) + \sum_{l=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} E[z_{t-1j}^{l} z_{tk}^{l}] \ln(a_{jk}) \tag{65}$$

To maximise it we need to put constraints

$$\sum_{k=1}^{5} b_{jk} = 1 \tag{66}$$

$$\sum_{k=1}^{5} \pi_k = 1 \tag{67}$$

$$\sum_{k=1}^{K} a_{jk} = 1 \tag{68}$$

So

$$\boldsymbol{L}(\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B}) = \sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{5}\sum_{k=1}^{K} x_{tj}^{l} E[z_{tk}^{l}] \ln(b_{kj}) + \sum_{l=1}^{N}\sum_{k=1}^{K} E[z_{1k}^{l}] \ln(\pi_k) \tag{69}$$

$$+ \sum_{l=1}^{N}\sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} E[z_{t-1j}^{l} z_{tk}^{l}] \ln(a_{jk}) - \lambda(\sum_{k=1}^{5} b_{jk} - 1) - \gamma(\sum_{k=1}^{5} \pi_k - 1) - \mu(\sum_{k=1}^{K} a_{jk} - 1) \tag{70}$$

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

$$\frac{\partial \boldsymbol{L}(\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B})}{\partial b_{jk}} = \frac{\sum_{l=1}^{N}\sum_{t=1}^{T} E[z_{tk}^{l}]}{b_{jk}} - \lambda = 0 \tag{71}$$

$$\lambda = \frac{\sum_{l=1}^{N}\sum_{t=1}^{T} E[z_{tk}^{l}]}{b_{jk}} \tag{72}$$

$$b_{jk} = \frac{\sum_{l=1}^{N}\sum_{t=1}^{T} x_{tj}^{l} E[z_{tk}^{l}]}{\sum_{l=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{K} E[z_{tk}^{l}]} \tag{73}$$

$$\frac{\partial \boldsymbol{L}(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})}{\partial \pi_k} = \frac{\sum_{l=1}^N \sum_{k=1}^K E[z_{1k}^l]}{\pi_k} - \gamma = 0 \tag{74}$$

$$\gamma = \frac{\sum_{l=1}^N \sum_{k=1}^K E[z_{1k}^l]}{\pi_k} \tag{75}$$

$$\pi_k = \frac{\sum_{l=1}^N E[z_{1k}^l]}{\sum_{l=1}^N \sum_{r=1}^K E[z_{1r}^l]} \tag{76}$$

$$\frac{\partial \boldsymbol{L}(\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})}{\partial a_{jk}} = \frac{\sum_{l=1}^N \sum_{t=2}^T E[z_{t-1j}^l z_{tk}^l]}{a_{jk}} - \mu = 0 \tag{77}$$

$$a_{jk} = \frac{\sum_{l=1}^N \sum_{t=2}^T E[z_{t-1j}^l z_{tk}^l]}{\sum_{l=1}^N \sum_{t=2}^T \sum_{r=1}^K E[z_{t-1j}^l z_{tr}^l]} \tag{78}$$

Answer is

$$b_{jk} = \frac{\sum_{l=1}^N \sum_{t=1}^T x_{tj}^l E[z_{tk}^l]}{\sum_{l=1}^N \sum_{t=1}^T \sum_{j=1}^K E[z_{tk}^l]} \tag{79}$$

$$\pi_k = \frac{\sum_{l=1}^N E[z_{1k}^l]}{\sum_{l=1}^N \sum_{r=1}^K E[z_{1r}^l]} \tag{80}$$

$$a_{jk} = \frac{\sum_{l=1}^N \sum_{t=2}^T E[z_{t-1j}^l z_{tk}^l]}{\sum_{l=1}^N \sum_{t=2}^T \sum_{r=1}^K E[z_{t-1j}^l z_{tr}^l]} \tag{81}$$