Imperial College London

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 21, 2017

1 Exercise I

1.a

fdss

1.b

(ii) fsds

2 Exercise II

(i)

$$p(\boldsymbol{Z}_n) = \prod_{k=1}^{K} \pi_k^{Z_{nk}} \tag{1}$$

$$\begin{vmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 2 & 2 & 2 \end{vmatrix} \tag{2}$$

(ii) If you do longer manipulations, please explain what you are doing: Try to avoid sequences of equations without text breaking up. Here is an example: We consider

$$U_1 = [\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}] \subset \mathbb{R}^4, \quad U_2 = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}] \subset \mathbb{R}^4. \tag{3}$$

===============
N sequences, T observations per sequence, 5 states
Estimate $\theta = \{\boldsymbol{\pi}, A\}$
$\boldsymbol{\pi} = \{\pi_1 ... \pi_5\}$
$for l = 1..N : D_l = \left\{x_1^l ... x_T^l\right\}$
$\boldsymbol{A} = [a_{ij}]$
$x_t^l = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$
Maximise $p(D_1 ... D_N | \theta)$

$$p(D_1 ... D_N | \theta) \tag{4}$$

$$= \prod_{l=1}^{N} p(D_l|\theta) \tag{5}$$

$$\tag{6}$$

Because

$$p(D_1|\theta) \tag{7}$$

$$= p(x_1^l ... x_T^l|\theta) \tag{8}$$

$$= p(x_1^l) \prod_{t=2}^{T} p(x_t^l|x_{t-1}^l) \tag{9}$$

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

$$\tag{11}$$

so

$$p(D_1 ... D_N|\theta) \tag{12}$$

$$= \prod_{l=1}^{N} \prod_{k=1}^{5} \pi_k^{x_{1k}^l} \prod_{T=2}^{T} \prod_{j=1}^{5} \prod_{k=1}^{5} a_{jk}^{x_{t-1j}^l x_{tk}^l} \tag{13}$$

$$\tag{14}$$

Take the log

$$\ln(p(D_1 ... D_N|\theta)) \tag{15}$$

$$= \sum_{l=1}^{N} \sum_{k=1}^{5} x_{1k}^l \pi_k + \sum_{l=1}^{N} \sum_{T=2}^{T} \sum_{j=1}^{5} \sum_{k=1}^{5} x_{t-1j}^l x_{tk}^l a_{jk} \tag{16}$$

$$\tag{17}$$

Now we take

$$N_k^l = \sum_{l=1}^{N} x_{1k}^l \tag{18}$$

$$N_{jk} = \sum_{l=1}^{N} \sum_{T=2}^{T} x_{t-1j}^l x_{tk}^l \tag{19}$$

$$\tag{20}$$

so

$$\ln(p(D_1 ... D_N|\theta)) \tag{21}$$

$$= \sum_{l=1}^{N} N_k^l \pi_k + \sum_{j=1}^{5} \sum_{k=1}^{5} N_{jk} a_{jk} \tag{22}$$

$$\tag{23}$$

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

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

$$\sum_{k=1}^{K} a_{jk} = 1 \tag{25}$$

$$\tag{26}$$

So

$$\boldsymbol{L}(\boldsymbol{\pi}, \boldsymbol{A}) = \sum_{l=1}^{N} N_k^l \pi_k + \sum_{j=1}^{5} \sum_{k=1}^{5} N_{jk} a_{jk} - \lambda(\sum_{k=1}^{K} \pi_k - 1) \tag{27}$$

1
1
1

(iii) dwdw