FS Questions Ex9

Montag, 26. Juni 2023 10:09

Q_1: Show decomposition of Q

Q_2: Explain why $\sum_{z} = \sum_{z^1} \cdots \sum_{z^N}$

Q_3: Explain 9.56 // what is $E[z_{nk}]$

Q_1:

Let us revisit the EM-algorithm and also answer the question:

"Why do we call steps in EM
E-step = "expectation"-step +
M-step = "maximization"-step?"

See: Bishop 9.4 annotated!

Q_2:

Suppose $z = \{a, b\}$, $a \in \{a_1, a_2\}$, $b \in \{b_1, b_2\}$

Now, what does $\sum_{z} f(a, b)$ mean?

Note that $\sum_{z}$ usually refers to:
"Sum over all (possible) values of z".

These are: $\{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

Thus: $\sum_{z} f(a, b) = f(a_1, b_1) + f(a_1, b_2)$
$+ f(a_2, b_1) + f(a_2, b_2)$
$= \sum f(a, b_1) + f(a, b_2)$

$$= \sum_a f(a, b_1) + f(a, b_2)$$

$$= \sum_a \sum_b f(a, b)$$

$$= \sum_{\{a, b\}} f(a, b)$$

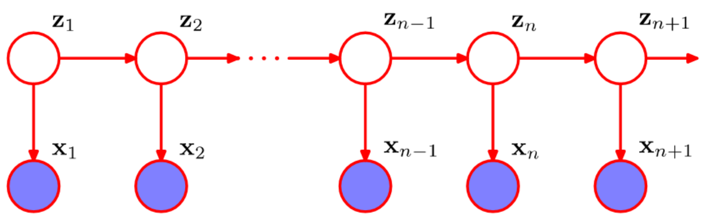
Q_3:

We look at this for the HMM case.
(Single sequence with notation following Bishop, i.e. n, N denote time index.)

Let us have a look at the form of Q:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}). \qquad (13.12)$$

Where for $p(X, Z|\theta)$, we have:



$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = p(\mathbf{z}_1|\boldsymbol{\pi}) \left[\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) \right] \prod_{m=1}^{N} p(\mathbf{x}_m|\mathbf{z}_m, \boldsymbol{\phi}) \qquad (13.10)$$

With 1-of-K encoding for z_n,
e.g. $z_n = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ for "z_n is in state 2" we have:

$$p(\mathbf{z}_1|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \qquad (13.8)$$

$$p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{n-1,j} z_{nk}}. \qquad (13.7)$$

$$p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\phi}) = \prod_{k=1}^{K} p(\mathbf{x}_n|\boldsymbol{\phi}_k)^{z_{nk}}. \qquad (13.9)$$

(13.12) then takes the form

$$\mathbb{E}_{z \sim p(z|x,\theta^{old})}\left[\ln p(X, Z|\theta)\right]$$

$$= \mathbb{E}_{z \sim p(z|x,\theta^{old})}\left[\sum_{k=1}^{K} z_{1k} \ln \pi_k\right]$$

$$+ \mathbb{E}_{z \sim p(z|x,\theta^{old})}\left[\sum_{n=2}^{N}\sum_{k=1}^{K}\sum_{j=1}^{K} z_{n-1,j} \cdot z_{nk} \ln A_{jk}\right]$$

$$+ \mathbb{E}_{z \sim p(z|x,\theta^{old})}\left[\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk} \ln p(x_n|\phi_k)\right]$$

$$= \sum_{k=1}^{K} \mathbb{E}_z\left[z_{1k}\right] \ln \pi$$

$$+ \sum_{n=2}^{N}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{E}_z\left[z_{n-1,j} \, z_{nk}\right] \ln A_{jk}$$

$$+ \sum_{n=1}^{N}\sum_{k=1}^{K} \mathbb{E}_z\left[z_{nk}\right] \ln p(x_n|\phi_k) = ELL$$

Where we use $\mathbb{E}_z$ as shorthand for $\mathbb{E}_{z \sim p(z|x,\theta^{old})}$ ...

We now define

$$\gamma(z_{nk}) = \mathbb{E}_z[z_{nk}] \overset{*}{=} \sum_{z_n} p(z_n|x,\theta^{old}) \, z_{nk}$$

and

$$\xi(z_{n-1,j}, z_{nk}) = \mathbb{E}_z[z_{n-1,j} \cdot z_{nk}] \overset{**}{=} \sum_{z_{n-1}, z_n} p(z_{n-1}, z_n | x, \theta^{old}) \, z_{n-1,j} \cdot z_{nk}$$

* noting that $\mathbb{E}_z[z_{nk}] = \sum_{z_1} \cdots \sum_{z_N} \prod_{m=1}^{N} p(z_m|x,\theta^{old}) \, z_{nk}$

$$= \underbrace{\sum_{z_1} p(z_1|X,\theta^{old}) \cdots}_{= \prod_{m=1}^{n-1} 1 = 1} \sum_{z_n} p(z_n|X,\theta^{old})\, z_{nk} \cdots \underbrace{\sum_{z_N} p(z_N|X,\theta^{old})}_{= \prod_{m=n+1}^{N} 1 = 1}$$

** analogous

With these definitions, we have:

$$ELL = \sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi + \sum_{n=2}^{N} \sum_{k=1}^{K} \sum_{j=1}^{K} \xi(z_{n-1,j}, z_{nk}) \ln A_{jk}$$
$$+ \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \ln p(x_n|\phi_k)$$

We thus showed, that

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \ln A_{jk}$$
$$+ \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \ln p(\mathbf{x}_n|\boldsymbol{\phi}_k). \qquad (13.17)$$