Do the rest of GMM from Lect. 4.

$\theta = (\pi_1, \dots, \pi_C, \theta_1, \dots, \theta_C)$ where $\theta_c = (\mu_c, \sigma_c^2)$, θ' anc.

Expected complete log-lik. for GMM

$$\sum_n E_{p(z_n|x_n,\underset{\substack{\nearrow \\ \text{current} \\ \text{parameters}}}{\theta})} \left[\ell(\theta'; z_n, x_n) \right] =$$

$$= \sum_n E\left[\log \prod_c (\pi_c' \, p(x_n | z_n = c, \theta'))^{I(z_n=c)}\right]$$

$$= \sum_n E\left[\sum_c I(z_n=c) \log \pi_c' + \sum_c I(z_n=c) \log p(x_n | \theta_c')\right]$$

$$= \sum_n \sum_c \underbrace{E[I(z_n=c)]}_{=p(z_n=c|x_n,\theta)=r_{nc}} \log \pi_c' + \sum_n \sum_c \underbrace{E[I(z_n=c)]}_{=p(z_n=c|x_n,\theta)=r_{nc}} \log p(x_n | \theta_c')$$

$$= \underbrace{\sum_c \left[\sum_n r_{nc}\right] \log \pi_c'}_{(*)} + \underbrace{\sum_c \sum_n r_{nc} \log p(x_n | \theta_c')}_{(**)}$$

Cannot sum out n in this case

Obs r_{nc} dep. on θ, current para.

(*) and (**) can be maximized separately

Max by fractions gives

$$\pi'_c = \frac{\sum_n r_{nc}}{N} = \frac{r_c}{N} \quad \left(\text{so } r_c = \sum_n r_{nc}\right)$$

$\left(\text{Observe } \sum_{n,c} r_{n,c} = N\right)$

For (**), we want to maximize each

$$f_c(\mu'_c, \sigma'_c) = \sum_n r_{nc} \log\left[\frac{1}{\sqrt{2\pi\sigma'^2_c}} e^{-\frac{1}{2\sigma'^2_c}(x_n - \mu'_c)^2}\right]$$

$$\propto \sum_n r_{nc} \log \frac{1}{\sigma'^2_c} - \sum_n r_{nc} \frac{1}{2\sigma'^2_c}(x_n - \mu'_c)^2$$

$$\frac{\delta f_c}{\delta \mu'_c} = \sum_n r_{nc} \frac{1}{\sigma'^2_c} (x_n - \mu'_c) \quad \text{setting } \frac{\delta f_c}{\delta \mu_c} = 0$$

gives $\sum_n r_{nc} x_n = \mu'_c \sum r_{nc} = \mu'_c r_c$, i.e.,

$$\mu'_c = \frac{\sum_n r_{nc} x_n}{r_c}$$

Sim. for variance gives (but use $\alpha'_c = \frac{1}{\sigma'_c}$)

$$f_c(\sigma'_c, \alpha'_c) = \sum_n r_{nc} \log \alpha'_c - \sum_n r_{nc} \frac{\alpha'^2_c}{2} (x_n - \mu'_c)^2$$

$$\frac{\delta f}{\delta \alpha'_c} = \sum_n r_{nc} \frac{1}{\alpha'_c} - \sum_n r_{nc} \alpha'_c (x_n - \mu'_c)^2$$

$$\frac{\delta f}{\delta \alpha'_c} = 0 \Rightarrow \sum_n r_{nc} \frac{1}{\alpha'_c} = \sum_n r_{nc} \alpha'_c (x_n - \mu'_c)^2$$

So, $\sigma'^2_c = \frac{1}{\alpha'^2_c} = \sum_n r_{nc} (x_n - \mu'_c)^2 / N_c$.

Mixture of Bernoulli

Data $D = \{x_1, \dots, x_N\}$ where $x_n = x_{n1}, \dots, x_{nD}$

Complete log-likelihood:

(θ all param., $\mu_c = \mu_{c1}, \dots, \mu_{cD}$)

$$\ell(\theta'; z_n, x_n)$$

$$= \log p(x_n, z_n \mid \pi', \mu')$$

$$= \log \prod_c \left[\pi'_c \, p(x_n \mid \mu'_c)\right]^{I(z_n = c)}$$

$$= \sum_c I(z_n = c) \log \pi'_c \prod_d p(x_{nd} \mid \mu'_{cd})$$

$$= \sum_c I(z_n = c) \log \pi'_c + \sum_c \sum_d I(z_n = c) \log p(x_{nd} \mid \mu'_{cd})$$

$$\sum_n E_{p(z_n|x_n,\theta)}\left[\ell(\theta'; z_n, x_n)\right]$$

$$= \sum_n \sum_c \overbrace{E[I(z_n=c|x_n,\theta)]}^{r_{nc}} \log \pi'_c$$

$$+ \sum_n \sum_c \sum_d E[I(z_n=c|x_n,\theta)] \log p(x_{nd}|\mu'_{cd})$$

$$= \sum_c \left[\sum_n r_{nc}\right] \log \pi'_c + \sum_{c,d} \underbrace{\sum_n r_{nc} \log (\mu'_{cd})^{x_{nd}}(1-\mu'_{cd})^{1-x_{nd}}}_{\left[\sum_{n:x_{nd}=1} r_{nc}\right]\log \mu'_{cd} + \left[\sum_{n:x_{nd}=0} r_{nc}\right]\log(1-\mu'_{cd})}.$$

Max by

$$\pi'_c = \frac{\sum_n r_{nc}}{N} = \frac{r_c}{N}$$

$$\mu'_{cd} = \sum_{n:x_{nd}=1} r_{nc} \Big/ r_c$$

Baum-Welch

$$\sum_n E_{p(Z_{1:T+1}|x^n_{1:T},\theta)}\left[\ell(\theta'; Z_{1:T+1}, X^n_{1:T})\right]$$

$$= \sum_n E_{p(Z_{1:T+1}|x^n_{1:T},\theta)} \log \prod_t \left[\prod_{k,s} B'^{\,I(x^n_t=s,\, Z_t=k)}_{sk} \prod_{k,l} A'^{\,I(Z_t=k,\, Z_{t+1}=l)}_{lk}\right]$$

$\prod_{x^n}$

$$= \sum_n \sum_t \sum_{k,s} p(X_t=s, Z_t=k \mid X^n, \theta) \log B'_{sk}$$

$$\sum_n \sum_t \sum_{k,s} p(Z_t=k, Z_{t+1}=l \mid X^n, \theta) \log A'_{lk}$$

$$= \sum_k \sum_s \overbrace{\left[\sum_{n,t} p(X_t=s, Z_t=k \mid X^n, \theta)\right]}^{\bar{M}_{k,s}} \log B'_{sk}$$

same k comp. for ...

$$\sum_k \sum_l \overbrace{\left[\sum_{n,t} p(Z_t=k, Z_{t+1}=l \mid X^n, \theta)\right]}^{\bar{N}_{k,l}} \log A'_{lk}$$

prob. mass
but diff