Backward variable

$$b_t(k) = p(x_{t+1,T} \mid z_t = k)$$

$$= \sum_{\ell} p(x_{t+1,T}, z_{t+1} = \ell \mid z_t = k)$$

$$= \sum_{\ell} p(z_{t+1} = \ell \mid z_t = k)\, p(x_{t+1,T} \mid z_{t+1} = \ell)$$

$$= \sum_{\ell} \underbrace{p(z_{t+1} = \ell \mid z_t = k)}_{\text{transition}}\, \underbrace{p(x_{t+2,T} \mid z_{t+1} = \ell)}_{b_{t+1}(\ell)}\, \underbrace{p(x_{t+1} \mid z_{t+1} = \ell)}_{\text{emission}}$$

Smoothing

$$p(z_t = k \mid x_{1:T}) = \frac{p(z_t = k, x_{1:T})}{p(x_{1:T})}$$

$$\propto p(x_{1:t-1}, z_t = k)\, p(x_{t:T} \mid z_t = k)$$

$$= \underbrace{f_t(k)}_{\text{forw.}}\, \underbrace{p(x_t \mid z_t = k)}_{\text{emiss.}}\, \underbrace{b_t(k)}_{\text{back.}}$$

Sampling

$$p(z_{1:T+1}|x_{1:T}) = p(z_{T+1}|x_{1:T})\, p(z_{1:T}|z_{T+1}, x_{1:T})$$

$$= \cancel{p(z_{T+1}|x_{1:T})}\, p(z_T|z_{T+1}, x_{1:T})\, p(z_{1:T-1}|z_T, x_{1:T-1})$$

$$= \prod_{t=1}^{T} p(z_t|z_{t+1}, x_{1:t-1})$$

So sam $z_{T+1} \xrightarrow{p(z_t|z_{t+1}, x_{1:t})} z_T \rightarrow z_{T-1} \dashrightarrow z_1$

$$p(z_t|z_{t+1}, x_{1:t}) \underbrace{\propto}_{\text{up to a constant (same } \forall z_t)} p(z_t, z_{t+1}, x_{1:t})$$

$$= \underbrace{p(z_t, x_{1:t-1})}_{\text{forward}}\, \underbrace{p(z_{t+1}, x_t|z_t)}_{\text{trans. \& emiss.}}$$

And the special case

$$p(z_{T+1} \mid x_{1:T}) \propto \underbrace{p(z_{T+1}, x_{1:T})}_{\text{forward}}$$

Viterbi

$$v_t(k) = \max_{z_{1:t-1}} p(z_{1:t-1}, z_t = k, x_{1:t})$$

$$= \max_{l} \max_{z_{1:t-2}} p(z_{1:t-2}, z_{t-1} = l, z_t = k, x_{1:t})$$

$$= \max_{l} \max_{z_{t-2}} (z_{t-2}, z_{t-1} = l, x_{1:t-1})\, p(z_t = k, x_t \mid z_{t-1} = l)$$

$$= \max_{l} \underbrace{v_{t-1}(l)}_{\text{smaller}}\, \underbrace{p(z_t = k \mid z_{t-1} = l)}_{\text{transition}}\, \underbrace{p(x_t \mid z_t = k)}_{\text{emission}}$$

GMM

Complete log-like. (complete data)

all para.

$$\ell(\theta'; D) = \log \prod_n p(z_n, x_n \mid \theta')$$

1 for one
0 for rest

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

$$= \sum_n \sum_c I(z_n = c) \log \left[\pi'_c \, p(x_n \mid z_n = c)\right]$$

class c's
para.

$$= \sum_{n,c} I(z_n = c) \log \pi'_c + \sum_{n,c} I(z_n = c) \log p(x_n \mid \theta'_c)$$

$$= \sum_c \sum_{n: z_n = c} \log \pi'_c + \sum_c \sum_{n: z_n = c} \log p(x_n \mid \theta'_c)$$

$$= \sum_c N_c \log \pi'_c + \underbrace{\sum_c \sum_{n: z_n = c} \log p(x_n \mid \theta'_c)}$$

(*)

$$N_c = |\{n : z_n = c\}| = \overbrace{\underbrace{\sum_n}_{\text{ML style}} I(z_n = c)}^{(**)}$$

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

Let $$f(\sigma'_c, \mu'_c) = \sum_{n: z_n = c} \log \frac{1}{\sqrt{2\pi\sigma_c'^2}} - \sum_{n: z_n = c} \frac{1}{2\sigma_c'^2}(x_n - \mu'_c)^2$$

Derivation

$$\frac{\partial f}{\partial \mu'_c} = \sum_{n: z_n = c} \frac{1}{\sigma_c'^2}(x_n - \mu'_c)$$

Solving $\frac{\partial f}{\partial \mu'_c} = 0 \Rightarrow \sum_{n: z_n = c} x_n = \sum_{n: z_n = c} \mu'_c = \underbrace{N_c}_{|\{n : z_n = c\}|} \mu'_c$

So,

$$\mu'_c = \frac{\sum_{n: z_n = c} x_n}{N_c}$$

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

$$f(\sigma'_c, \alpha'_c) = \sum_{n: z_n = c} \log \frac{\alpha'_c}{\sqrt{2\pi}} - \sum_{n: z_n = c} \frac{\alpha'^2_c}{2} (x_n - \mu'_c)^2$$

$$\frac{\delta f}{\delta \alpha'_c} = \sum_{n: z_n = c} \frac{1}{\alpha'_c} - \sum_{n: z_n = c} \alpha'_c (x_n - \mu'_c)^2$$

$$\frac{\delta f}{\delta \alpha'_c} = 0 \Rightarrow \sum_{n: z_n = c} \frac{1}{\alpha'_c} = \sum_{n: z_n = c} \alpha'_c (x_n - \mu'_c)^2$$

So, $\sigma'^2_c = \frac{1}{\alpha'^2_c} = \sum_{n: z_n = c} (x_n - \mu'_c)^2 / N_c$.