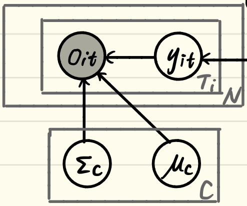
ubm + i-vector note.

Views from probabilistic graphical Model.

$\pi = C$ ‖ π is a $C \times 1$ vector

ubm is an ordinary GMM.



$y_{it} \sim Multi(\pi)$.
$O_{it} | y_{it} \sim \mathcal{N}(\mu_c, \Sigma_c)$.

For utterance i

For frame t

- Select a component indicator $y_{it} = C$ with probability π_c $(c = 1 \cdots C)$.
- Select a speech frame O_{it} from $\mathcal{N}(\mu_c, \Sigma_c)$.

The joint distribution $p(O_{i\cdot}, y_{i\cdot} | \pi, \Sigma_c, \mu_c) = p(y_{i\cdot} | \pi) \cdot p(O_{i\cdot} | y_{i\cdot}, \Sigma_c, \mu_c)$

$$= \prod_{t=1}^{T_i} \prod_{c=1}^{C} \pi_c^{1_c(y_{it})} \exp\{1_c(y_{it}) [-\tfrac{1}{2}(O_{it} - \mu_c)^T \Sigma_c^{-1} (O_{it} - \mu_c) - \tfrac{D}{2} \ln \pi - \tfrac{1}{2} \ln |\Sigma_c|]\}$$

$$L_i = \ln p(O_{i\cdot}, y_{i\cdot} | \pi, \Sigma_c, \mu_c) = \sum_{t=1}^{T_i} \sum_{c=1}^{C} 1_c(y_{it}) \cdot (\ln(\pi_c) + O_{it}^T \Sigma_c^{-1} \mu_c - \tfrac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \tfrac{1}{2} \mu_c^T \Sigma_c^{-1} \mu_c - \tfrac{D}{2} \ln \pi - \tfrac{1}{2} \ln |\Sigma_c|)$$

E-step:

$$E_p[1_c(y_{it})] = p(y_{it} = c | O_{it}, \Sigma_c, \mu_c, \pi) = \frac{\pi_c N(O_{it} | \mu_c, \Sigma_c)}{\sum_c \pi_c N(O_{it} | \mu_c, \Sigma_c)}$$ for each O_{it}.

Denote $\gamma_c(O_{it}) = E_p[1_c(y_{it})]$

M-step:

Maximize $E_p[\sum_i L_i]$ w.r.t. π_c, Σ_c, μ_c.

$$\mathcal{L} = E_p[\sum_i L_i] = \sum_{i,t} \sum_{c=1}^{C} E_p[1_c(y_{it})] \cdot (\ln \pi_c + O_{it}^T \Sigma_c^{-1} \mu_c - \tfrac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \tfrac{1}{2} \mu_c^T \Sigma_c^{-1} \mu_c - \tfrac{D}{2} \ln \pi - \tfrac{1}{2} \ln |\Sigma_c|)$$

$$= \sum_{i,t} \sum_{c=1}^{C} \gamma_c(O_{it}) (\ln \pi_c + O_{it}^T \Sigma_c^{-1} \mu_c - \tfrac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \tfrac{1}{2} \mu_c^T \Sigma_c^{-1} \mu_c - \tfrac{1}{2} \ln |\Sigma_c| + const)$$

① $\mathcal{L}(\pi_c) = \sum_{i,t} \gamma_c(O_{it}) \cdot \ln \pi_c$, s.t. $\sum_c \pi_c = 1$

$$0 = \frac{\partial \mathcal{L}}{\partial \pi_c} = \frac{1}{\pi_c} (\sum_{i,t} \gamma_c(O_{it})) - \lambda_c \Rightarrow \pi_c = \frac{\sum_{i,t} \gamma_c(O_{it})}{\sum_c \sum_{i,t} \gamma_c(O_{it})}$$

② $\mathcal{L}(\mu_c) = \sum_{i,t} \gamma_c(O_{it}) \cdot (O_{it}^T \Sigma_c^{-1} \mu_c - \tfrac{1}{2} \mu_c^T \Sigma_c^{-1} \mu_c)$

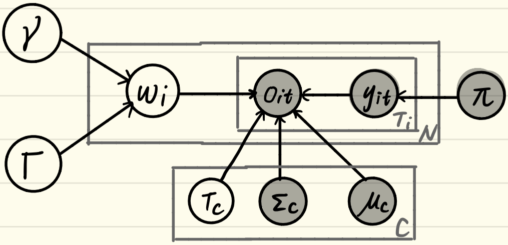
$$0 = \frac{\partial \mathcal{L}}{\partial \mu_c} = \sum_{i,t} \gamma_c(O_{it}) (\Sigma_c^{-1} O_{it} - \Sigma_c^{-1} \mu_c)$$

$$\Rightarrow \mu_c = \frac{\sum_{i,t} \gamma_c(O_{it}) O_{it}}{\sum_{i,t} \gamma_c(O_{it})}$$

③ $\mathcal{L}(\Sigma_c^{-1}) = \sum_{i,t} \gamma_c(O_{it}) (O_{it}^T \Sigma_c^{-1} \mu_c - \tfrac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \tfrac{1}{2} \mu_c^T \Sigma_c^{-1} \mu_c + \tfrac{1}{2} \ln |\Sigma_c^{-1}|)$

$$\Rightarrow \Sigma_c^{-1} = \frac{\sum_{i,t} \gamma_c(O_{it}) (O_{it} - \mu_c)(O_{it} - \mu_c)^T}{\sum_{i,t} \gamma_c(O_{it})}$$

i-vector is factor Analysis.



For utterance i

- Select a total variability variable w_i from $\mathcal{N}(\gamma, \Gamma)$. where $\gamma = 0$. $\Gamma = I$

For frame t

Select O_{it} from $\mathcal{N}(O_{it} | c = y_{it}, T_c w_c + \mu_c, \Sigma_c)$

The joint distribution: $p(O_{i\cdot}, w_i | \pi, y_{it}, \Sigma_c, \mu_c, T_c, \gamma, \Gamma)$

$$= p(w_i | \gamma, \Gamma)\, p(O_{i\cdot} | w_i, T_c, \mu_c, \Sigma_c, y_{it})$$

$$= \mathcal{N}(w_i | 0, I) \prod_{t=1}^{T_i} p(O_{it} | w_i, T_c, \mu_c, \Sigma_c, y_{it}).$$

$$\propto \exp\left(-\tfrac{1}{2} w_i^T w_i + \sum_{t=1}^{T_i}\sum_{c=1}^{C}\left[-\tfrac{1}{2}(O_{it} - T_c w_i - \mu_c)^T \Sigma_c^{-1}(O_{it} - T_c w_i - \mu_c) - \tfrac{1}{2}\ln|\Sigma_c|\right] \cdot 1_c(y_{it})\right).$$

$$\ln p(O_{i\cdot}, w_i | \cdots) = -\tfrac{1}{2}[w_i^T w_i] + \sum_{t=1}^{T_i}\sum_{c=1}^{C}\left[-\tfrac{1}{2}(O_{it} - T_c w_i - \mu_c)^T \Sigma_c^{-1}(O_{it} - T_c w_i - \mu_c) - \tfrac{1}{2}\ln|\Sigma_c|\right] 1_c(y_{it})$$

E-step: Compute statistics of w_i.

$$p(w_i | O_{i\cdot}) = \exp\left[-\tfrac{1}{2} w_i^T w_i + \sum_{t=1}^{T_i}\sum_{c=1}^{C}\left(O_{it}^T \Sigma_c^{-1}(T_c w_i + \mu_c) - \tfrac{1}{2} w_i^T T_c^T \Sigma_c^{-1} T_c w_i\right) \gamma_c(y_{it})\right)$$

which is a gaussian with mean γ_i. and covariance Γ_i^{-1}.

$$\gamma_i = \Gamma_i^{-1} \sum_{t=1}^{T_i}\sum_{c=1}^{C} \gamma_c(y_{it}) T_c^T \Sigma_c^{-1}(O_{it} - \mu_c$$

$$\Gamma_i = I + \sum_{t=1}^{T_i}\sum_{c=1}^{C} T_c^T \Sigma_c^{-1} T_c \cdot \gamma_c(y_{it})$$

So $E[w_i] = \gamma_i \quad E[w_i w_i^T] = \gamma_i \gamma_i^T + \Gamma_i^{-1}$

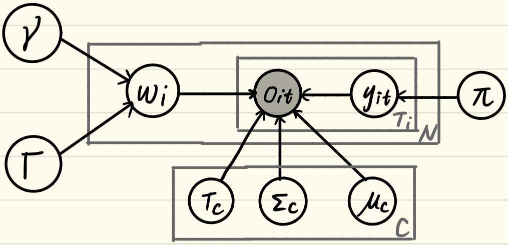
M-step: Compute T_c.

$$\mathcal{L}(T_c) = \sum_i E_p[\ln p(O_{i\cdot}, w_i | \cdots)] = \text{const} + \sum_i \sum_t \gamma_i^T T_c^T \Sigma_c^{-1}(O_{it} - \mu_c) - \tfrac{1}{2} tr(T_c^T \Sigma_c^{-1} T_c (\Gamma_i + \gamma_i \gamma_i^T)) \cdot \gamma_c(O_{it})$$

$$\frac{\partial \mathcal{L}}{\partial T_c} = 0 \Rightarrow \sum_i \sum_t \left[\Sigma_c^{-1}(O_{it} - \mu_c)\gamma_i^T - \Sigma_c^{-1} T_c (\Gamma_i + \gamma_i \gamma_i^T)\right] \cdot \gamma_c(O_{it}) = 0$$

$$T_c = \left[\sum_i \sum_t \gamma_c(O_{it}) \cdot \gamma_i (O_{it} - \mu_c)^T\right]\left[\sum_i \sum_t \gamma_c(O_{it})(\Gamma_i + \gamma_i \gamma_i^T)\right]^{-1}$$

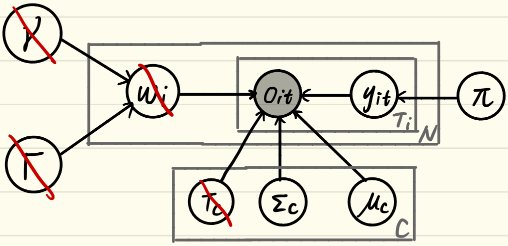
ubm + i-vector:



Problem behind this idea:

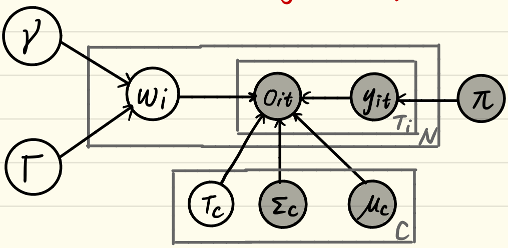
① May computational intractable.

① When traing ubm: Assume $T_c \equiv 0$



② When traing i-vector extractor.

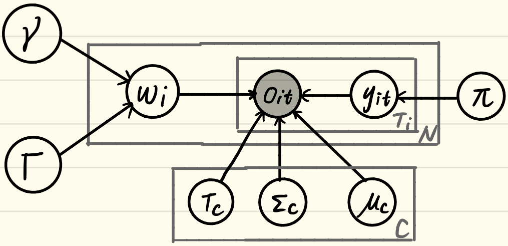
Assume y_{it}, Σ_c, μ_c π_c is known,



③ Extract i-vector $E[w_i]$ for each utterance.

Inspiration: ① Can we iterate over these two procedure?

② What about Variational Inference?



Now, we derive the update formulations.

For an utterance i. the joint distribution is:

$$p(w_i, y_{i\cdot}, O_{i\cdot} \mid \pi, T_c, \Sigma_c, \mu_c) = p(w_i \mid \gamma, \Gamma) \prod_{t=1}^{T_i} p(y_{it} \mid \pi) \cdot p(O_{it} \mid w_i, y_{it})$$

where $p(w_i \mid \gamma, \Gamma) = \mathcal{N}(w_i \mid 0, I) \propto \exp(-\frac{1}{2} w_i^T w_i)$

$$p(y_{it} \mid \pi) = \exp\left(\sum_{c=1}^{C} 1_c(y_{it}) \cdot \log \pi_c\right).$$

$$p(O_{it} \mid w_i, y_{it}) \propto \exp\Big[\sum_{c=1}^{C} 1_c(y_{it}) \big(O_{it}^T \Sigma_c^{-1} (T_c w_i + \mu_c) - \mu_c^T \Sigma_c^{-1} T_c w_i - \frac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \frac{1}{2} \mu_c^T \Sigma_c^{-1} \mu_c - \frac{1}{2} w_i^T T_c^T \Sigma_c^{-1} T_c w_i - \frac{1}{2} \ln|\Sigma_c|\big)\Big]$$

As, w_i and y_{it} and conditionally dependent given. O_{it}, we can not use E-M. directly. In this case, variational inference is suitable.

Now, we consider the partition of the posterior:

$$p(w_i, y_{i\cdot}) \approx q(y_{i\cdot}, w_i) = q(w_i) \cdot \prod_{t=1}^{T_i} q(y_{it})$$ (Mean-field method.).

The optimum for $q^*(w_i)$:

$$\begin{aligned}
\ln q^*(w_i) &= E_{y_{i\cdot}}[\ln p(w_i, y_{i\cdot}, O_{i\cdot})] + \text{const} \\
&= E_{y_{i\cdot}}[\ln p(O_{i\cdot} \mid y_{i\cdot}, w_i)] + \ln p(w_i) + \text{const} \\
&= \sum_{t=1}^{T_i} \sum_{c=1}^{C} E_{y_{it}}[1_c(y_{it})]\left[-\frac{1}{2}(O_{it} - T_c w_i - \mu_c)^T \Sigma_c^{-1} (O_{it} - T_c w_i - \mu_c)\right] - \frac{1}{2} w_i^T w_i + \text{const} \\
&= \sum_{t=1}^{T_i} \sum_{c=1}^{C} \gamma_c(O_{it}) \cdot \left[(O_{it} - \mu_c)^T \Sigma_c^{-1} T_c w_i - \frac{1}{2} w_i^T T_c^T \Sigma_c^{-1} T_c w_i\right] - \frac{1}{2} w_i^T w_i + \text{const} \\
&= \left[\sum_{t=1}^{T_i} \sum_{c=1}^{C} \gamma_c(O_{it})(O_{it} - \mu_c)^T \Sigma_c^{-1} T_c\right] w_i - \frac{1}{2} w_i^T \left[I + \sum_{t=1}^{T_i} \sum_{c=1}^{C} \gamma_c(O_{it}) T_c^T \Sigma_c^{-1} T_c\right] w_i + \text{const}
\end{aligned}$$

So $q^*(w_i)$ follows gaussian distribution. with mean γ_i. and covariance Γ_i^{-1}.

$$\gamma_i = \Gamma_i^{-1} \sum_{t=1}^{T_i} \sum_{c=1}^{C} T_c^T \Sigma_c^{-1} \gamma_c(O_{it})(O_{it} - \mu_c)$$

$$\Gamma_i = I + \sum_{t=1}^{T_i} \sum_{c=1}^{C} \gamma_c(O_{it}) T_c^T \Sigma_c^{-1} T_c$$

The optimum for $q^*(y_{i\cdot})$:

$$\ln q^*(y_{i\cdot}) = E_{w_i}[\ln p(w_i, y_{i\cdot}, O_{i\cdot})] + const$$
$$= E_{w_i}[\ln p(O_{i\cdot} \mid y_{i\cdot}, w_i)] + \ln p(y_{i\cdot}) + const$$
$$= \sum_{c=1}^{C} 1_c(y_{it})\Big(O_{it}^T \Sigma_c^{-1}(T_c \gamma_i + \mu_c) - \mu_c^T \Sigma_c^{-1} T_c \gamma_i - \tfrac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \tfrac{1}{2}\mu_c^T \Sigma_c^{-1}\mu_c$$
$$- \tfrac{1}{2}\underbrace{E_{w_i}[w_i^T T_c^T \Sigma_c^{-1} T_c w_i]}_{\Downarrow\ tr(T_c^T \Sigma_c^{-1} T_c(\Gamma_i + \gamma_i \gamma_i^T))} - \tfrac{1}{2}\ln|\Sigma_c| - \tfrac{D}{2}\Big)$$
$$+ \sum_{c=1}^{C} 1_c(y_{it}) \log \pi_c + const$$

So $q^*(y_{it})$ follows categorical distribution.

Let $q^*(y_{it}) = \mathrm{Multi}(\phi_{it}^c)$ where $\phi_{it}^c \propto \pi_c \exp(-\tfrac{1}{2}tr(\Sigma_c^{-1} T_c \Gamma_i T_c^T))\, N(O_{it} \mid T_c \gamma_i + \mu_c, \Sigma_c)$

The variational lower bound:

$$\mathcal{L} = E_q[\log p(w_{\cdot})] + E_q[\log p(y_{\cdot\cdot} \mid \pi)] + E_q[\log p(O_{\cdot\cdot} \mid w_{\cdot\cdot}, y_{\cdot\cdot})]$$
$$- E_q[\log q(w_{\cdot})] - E_q[\log q(y_{\cdot\cdot})]$$

$$E_q[\log p(w_{\cdot})] = \sum_i E_q[-\tfrac{1}{2} w_i^T w_i] = -\tfrac{1}{2}\sum_i (\gamma_i^T \gamma_i + tr(\Gamma_i^{-1})) + const$$

$$E_q[\log p(y_{\cdot\cdot} \mid \pi)] = \sum_i E_q\Big[\sum_{t=1}^{T_i}\sum_{c=1}^{C} 1_c(y_{it}) \log \pi_c\Big] = \sum_{i=1}\sum_{t=1}^{T_i}\sum_{c=1}^{C} \phi_{it}^c \cdot \log \pi_c$$

$$E_q[\log p(O_{\cdot\cdot} \mid w_{\cdot}, y_{\cdot\cdot})] = \sum_i \sum_{t=1}^{T_i}\sum_{c=1}^{C} \phi_{it}^c \Big\{ O_{it}^T \Sigma_c^{-1}(T_c \gamma_i + \mu_c) - \tfrac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \mu_c^T \Sigma_c^{-1} T_c \gamma_i$$
$$- \tfrac{1}{2} tr(\Sigma_c^{-1} T_c \Gamma_i^{-1} T_c^T + \Sigma_c^{-1} T_c \gamma_i \gamma_i^T T_c^T) - \tfrac{1}{2}\mu_c^T \Sigma_c^{-1}\mu_c$$
$$- \tfrac{D}{2}\ln 2\pi - \tfrac{1}{2}\ln|\Sigma_c| \Big\}$$

$$E_q[\log q(w_{\cdot})] = \sum_i \Big\{ E_q[\gamma_i^T \Gamma_i w_i] - \tfrac{1}{2} E_q[w_i^T \Gamma_i^{-1} w_i] - \tfrac{1}{2}\gamma_i^T \Gamma_i^{-1} \gamma_i - \tfrac{1}{2}\ln|\Gamma_i^{-1}| \Big\}$$
$$= \sum_i \Big\{ \tfrac{1}{2}\gamma_i^T \Gamma_i \gamma_i - \tfrac{1}{2} tr(\gamma_i^T \Gamma_i \gamma_i + I) - \tfrac{1}{2}\ln|\Gamma_i^{-1}| \Big\}$$
$$= \sum_i \tfrac{1}{2}\ln|\Gamma_i| + const$$

$$E_q[\log q(y_{\cdot\cdot})] = \sum_i \sum_t \sum_{c=1}^{C} \phi_{it}^c \cdot \log \phi_{it}^c$$

So the lower bound is

$$
\begin{aligned}
\mathcal{L} = & \sum_i -\frac{1}{2}(\gamma_i^T\gamma_i + tr(\Gamma_i^{-1})) \\
& + \sum_{i=1}\sum_{t=1}\sum_{c=1}^{C} \phi_{it}^c \cdot \log \pi_c \\
& + \sum_{i=1}\sum_{t=1}\sum_{c=1}^{C} \phi_{it}^c \{ O_{it}^T \Sigma_c^{-1}(T_c\gamma_i + \mu_c) - \frac{1}{2} O_{it}^T \Sigma_c^{-1} O_{it} - \mu_c^T \Sigma_c^{-1} T_c \gamma_i - \underline{\frac{1}{2} tr(\Sigma_c^{-1} T_c \Gamma_i^{-1} T_c^T)} \\
& \qquad - \frac{1}{2}\gamma_i^T T_c^T \Sigma_c^{-1} T_c \gamma_i - \frac{1}{2}\mu_c^T \Sigma_c^{-1}\mu_c - \frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_c| \}. \\
& + \sum_i \frac{1}{2}\ln|\Gamma_i^{-1}| - \sum_i\sum_t\sum_{c=1} \phi_{it}^c \log \phi_{it}^c \quad + const.
\end{aligned}
$$

Update parameters: $\pi_c, \Sigma_c^{-1}, T_c, \mu_c$, given. $\gamma_i, \Gamma_i^{-1}, O_{it}, \phi_{it}^c$

① $\mathcal{L}(\pi_c) = \sum_{i=1}\sum_{t=1} \phi_{it}^c \log \pi_c \qquad s.t. \sum_c \pi_c = 1$

$$\pi_c = \frac{\sum_{i=1}\sum_{t=1}\phi_{it}^c}{\sum_c\sum_{i,t}\phi_{it}^c}$$

② $\mathcal{L}(\mu_c) = \sum_{i=1}\sum_{t=1} \phi_{it}^c \{ O_{it}^T \Sigma_c^{-1}\mu_c - \gamma_i^T T_c^T \Sigma_c^{-1}\mu_c - \frac{1}{2}\mu_c^T\Sigma_c^{-1}\mu_c \}$

$$\frac{dL}{d\mu_c} = 0 \Rightarrow \mu_c = \frac{\sum_{i,t}\phi_{it}^c (O_{it} - T_c\gamma_i)}{\sum_{i,t}\phi_{it}^c}$$

③ $\mathcal{L}(\Sigma_c^{-1}) = \sum_{i=1}\sum_{t=1} \phi_{it}^c \{ O_{it}^T\Sigma_c^{-1}(T_c\gamma_i + \mu_c) - \frac{1}{2}O_{it}^T\Sigma_c^{-1}O_{it} - \mu_c^T\Sigma_c^{-1}T_c\gamma_i - \frac{1}{2}tr(\Sigma_c^{-1}T_c\Gamma_i^{-1}T_c^T)$
$\qquad - \frac{1}{2}\gamma_i^T T_c^T\Sigma_c^{-1}T_c\gamma_i - \frac{1}{2}\mu_c^T\Sigma_c^{-1}\mu_c - \frac{1}{2}\ln|\Sigma_c| \}$

$$
\begin{aligned}
\frac{dL}{d\Sigma_c^{-1}} = 0 \Rightarrow & \sum_{i,t}\phi_{it}^c \{ O_{it}(T_c\gamma_i + \mu_c)^T - \frac{1}{2}O_{it}O_{it}^T - \mu_c\gamma_i^T T_c^T - \frac{1}{2}T_c\Gamma_i^{-1}T_c^T \\
& - \frac{1}{2}T_c\gamma_i\gamma_i^T T_c^T - \frac{1}{2}\mu_c\mu_c^T + \frac{1}{2}\Sigma_c \} = 0
\end{aligned}
$$

$$
\begin{aligned}
\Sigma_c = & \frac{1}{\sum_{i,t}\phi_{it}^c} \{ \sum_{i,t}\phi_{it}^c (T_c\Gamma_i^{-1}T_c^T + T_c\gamma_i\gamma_i^T T_c^T + \mu_c\mu_c^T + O_{it}O_{it}^T + 2\mu_c\gamma_i^T T_c^T \\
& \qquad - 2O_{it}(T_c\gamma_i + \mu_c)^T) \} \\
= & \frac{1}{\sum_{i,t}\phi_{it}^c} \{ \sum_{i,t}\phi_{it}^c (O_{it} - T_c\gamma_i - \mu_c)(O_{it} - T_c\gamma_i - \mu_c)^T \} \\
& + T_c \left[\frac{1}{\sum_{i,t}\phi_{it}^c} \sum_{i,t}\phi_{it}^c \Gamma_i^{-1} \right] T_c^T
\end{aligned}
$$

④ $\mathcal{L}(T_c) = \sum_{i=1}\sum_{t=1} \phi_{it}^c \{ O_{it}^T\Sigma_c^{-1}T_c\gamma_i - \mu_c^T\Sigma_c^{-1}T_c\gamma_i - \frac{1}{2}tr(\Sigma_c^{-1}T_c\Gamma_i^{-1}T_c^T)$
$\qquad - \frac{1}{2}\gamma_i^T T_c^T \Sigma_c^{-1} T_c \gamma_i \}$

$$\frac{dL}{dT_c} = 0 \Rightarrow \sum_{i=1}\sum_{t=1}\phi_{it}^c \Sigma_c^{-1}(O_{it}\gamma_i^T - \mu_c\gamma_i^T - T_c\Gamma_i^{-1} - T_c\gamma_i\gamma_i^T) = 0$$

$$\therefore T_c = \left[\sum_{i,t}\phi_{it}^c (O_{it} - \mu_c)\gamma_i^T\right]\left[\sum_i\sum_t \phi_{it}^c (\Gamma_i^{-1} + \gamma_i\gamma_i^T)\right]$$

$$\phi_{it}^c = \gamma_c(O_{it}).$$

The Result:

$$\begin{cases} y_i = \Gamma_i^{-1} \sum_{t=1}^{T_i} \sum_{c=1}^{C} T_c^T \Sigma_c^{-1} \gamma_c(O_{it})(O_{it} - \mu_c). \\ \Gamma_i = I + \sum_{t=1}^{T_i} \sum_{c=1}^{C} \phi_{it}^c T_c^T \Sigma_c^{-1} T_c \end{cases}$$

$$\phi_{it}^c \propto \pi_c \exp\left(-\tfrac{1}{2} tr(\Sigma_c^{-1} T_c \Gamma_i T_c^T)\right) \mathcal{N}(O_{it} \mid T_c y_i + \mu_c, \Sigma_c).$$

$$\pi_c = \frac{\sum_{i,t} \phi_{it}^c}{\sum_c \sum_{i,t} \phi_{it}^c}$$

: Indicates the difference.

$$\mu_c = \frac{\sum_{i,t} \phi_{it}^c (O_{it} - T_c y_i)}{\sum_{i,t} \phi_{it}^c}$$

$$\Sigma_c = \frac{1}{\sum_{i,t} \phi_{it}^c} \left\{ \sum_{i,t} \phi_{it}^c \left[(O_{it} - T_c y_i - \mu_c)(O_{it} - T_c y_i - \mu_c)^T \right] \right\}$$

$$+ T_c \left[\frac{1}{\sum_{i,t} \phi_{it}^c} \sum_{i,t} \phi_{it}^c \Gamma_i^{-1} \right] T_c^T$$

$$T_c = \left[\sum_{i,t} \phi_{it}^c (O_{it} - \mu_c) y_i^T \right] \left[\sum_{i,t} \phi_{it}^c (\Gamma_i^{-1} + y_i y_i^T) \right]^{-1}$$

Initialize. $T_c = 0$ $y_i = 0$. $\Gamma_i = I$. $\pi_c = \frac{1}{N_c}$. $\mu_c = \text{randn}$, $\Sigma_c = \text{randn}$.

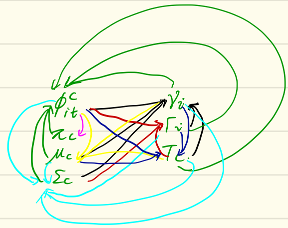
Now, it's a GMM-UBM.

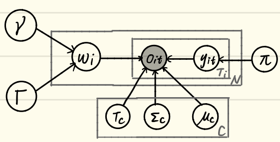
Iteratively update. ϕ_{it}^c and π_c μ_c, Σ_c.

Then Regard. ϕ_{it}^c, π_c, μ_c, Σ_c as given.

Update. T_c y_i Γ_i

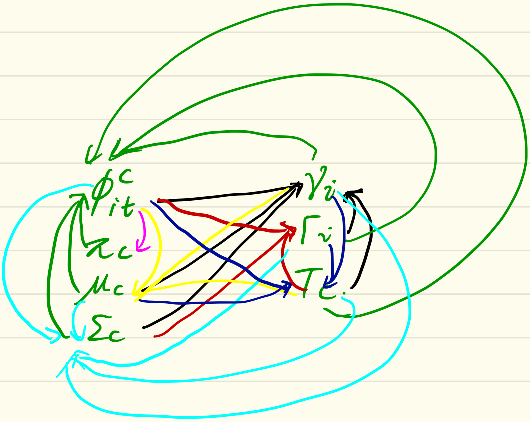
(Cannot decide the update order)





From / To

To \ From	ϕ_{it}^c	π_c	μ_c	Σ_c	y_i	Γ_i	T_c	
ϕ_{it}^c		/	/	/	/	/	/	6
π_c	/							1
μ_c	/				/		/	3
Σ_c	/		/		/	/	/	5
y_i	/		/	/		/	/	5
Γ_i	/			/			/	3
T_c	/		/		/	/		4
	6	1	4	3	4	4	5	



From

To	ϕ_{it}^c	π_c	μ_c	Σ_c	ν_i	Γ_i	$T_{c.}$	
ϕ_{it}^c		/	/	/	/	/	/	6
π_c	/							1
μ_c	/				/		/	3
Σ_c	/		/		/	/	/	5
ν_i	/		/	/		/	/	5
Γ_i	/			/			/	3
$T_{c.}$	/		/		/	/		4
	6	1	4	3	4	4	5	

Check the code of i-vector to ensure the initialization and update order. and the computational. expense.

① sid/train_diag_ubm.sh.
② sid/train_full_ubm.sh
③ sid/train_ivector_extractor.sh
④ sid/extract_ivectors.sh

① num_gauss_init 1024
num_gauss 2048
num_iters_init 20
gmm-global-init-from-feats.