② EM Algorithm and Factor Analysis

ⓐ $\hat{q}_i(h_i) = Pr(h_i \mid x_i, \theta)$.

<u>Apply Bayes Rule:</u>

$$Pr(h_i \mid x_i, \theta) = \frac{Pr(x_i \mid h_i, \theta) \cdot Pr(h_i)}{Pr(x_i \mid \theta)}$$

$$= \frac{N_{x_i}(\mu + \phi h, \Sigma) \cdot N_{h_i}(0, I)}{Pr(x_i \mid \theta)}$$

$Pr(x_i \mid \theta)$ is independent of h_i, hence can be treated as constant.

Hence, $\hat{q}_i(h_i) = k\, N_{x_i}(\mu + \phi h, \Sigma)\, N_{h_i}(0, I)$ — ①

Change of variable in Normal Distribution:

$$\text{Norm}_x[Ay + B, \Sigma] = k\, \text{Norm}_y[A'x + B', \Sigma']$$

where

$$A' = (A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1}$$

$$B' = -(A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1} B.$$

$$\Sigma' = -(A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1} B.$$

Applying this in equation ①, we get:

$$\hat{q}_i(h_i) = k_1 \cdot N_{h_i}\left((\phi^T \Sigma^{-1} \phi)^{-1} \phi^T \Sigma^{-1} (x_i - \mu), (\phi^T \Sigma^{-1} \phi)^{-1}\right) N_{h_i}(0, I).$$

constant term.

— ②

Product of 2 Normal distributions is proportional to a 3rd Normal Distribution.

$$\text{Norm}_x[a, A] \cdot \text{Norm}_x[b, B] = k\, \text{Norm}_x\left[(A^{-1} + B^{-1})^{-1}(A^{-1}a + B^{-1}b), (A^{-1} + B^{-1})^{-1}\right]$$

Putting this in equation ②, we get :

$$\hat{q}_i^*(h_i) = k_2 \mathcal{N}_{h_i}\left((\phi^T \Sigma^{-1} \phi + I)^{-1} \phi^T \Sigma^{-1}(x_i - \mu), \; (\phi^T \Sigma^{-1} \phi + I)^{-1} \right)$$

constant k_2 can be ignored.

(b) step 1

Take the log-likelihood function :

$$L(\tilde{\mu}, \tilde{\phi}, \tilde{\Sigma}) = \sum_{i=1}^{I} E[-\log|\tilde{\Sigma}| - (x_i - \tilde{\mu} - \tilde{\phi} h_i)^T \Sigma^{-1} (x_i - \tilde{\mu} - \tilde{\phi} h_i)]$$

$$= \sum_{i=1}^{I} [-\log|\tilde{\Sigma}| - x_i^T \tilde{\Sigma}^{-1} x_i - \tilde{\mu}_i^T \tilde{\Sigma}^{-1} \tilde{\mu}_i - E(h_i^T \tilde{\phi}^T \tilde{\Sigma}^{-1} \tilde{\phi} h_i) + 2\tilde{\mu}^T \tilde{\Sigma}^{-1} x_i - 2\tilde{\mu}^T \tilde{\Sigma}^{-1} \tilde{\phi} E(h_i) + 2 x_i \tilde{\Sigma}^{-1} \tilde{\phi} E(h_i)]$$

step 2

~~Dat~~ Differentiating $L(\tilde{\mu}, \tilde{\phi}, \tilde{\Sigma})$, partially wrt ~~the~~ parameters : (and equating it to 0)

$$\frac{\partial L}{\partial \tilde{\mu}} = \sum_{i=1}^{I} [-2\tilde{\Sigma}^{-1}\tilde{\mu} + 2\tilde{\Sigma}^{-1} x_i - 2\tilde{\Sigma}^{-1}\tilde{\phi} E(h_i)] = 0$$

or, $\sum_{i=1}^{I} (-2\tilde{\mu} + 2x_i - 2\tilde{\phi} E(h_i)) = 0$ [multiply by $\tilde{\Sigma}$, on both sides]

or, $\tilde{\mu} = \frac{1}{I} \sum_{i=1}^{I} (x_i - \tilde{\phi} E(h_i))$. —— ①

$$\frac{\partial L}{\partial \tilde{\phi}} = \sum_{i=1}^{I} \left(-2\tilde{\Sigma}^{-1} \tilde{\phi}\, E(h_i h_i^T) - 2\tilde{\Sigma}^{-1} \tilde{\mu}\, E(h_i)^T + 2\tilde{\Sigma}^{-1} x_i\, E(h_i)^T \right) = 0.$$

or $$\tilde{\phi} \sum_{i=1}^{I} E(h_i h_i^T) = \sum_{i=1}^{I} (x_i - \tilde{\mu})\, E(h_i)^T$$

or $$\tilde{\phi} = \left[\sum_{i=1}^{I} (x_i - \tilde{\mu})\, E(h_i)^T \right] \left[\sum_{i=1}^{I} E(h_i h_i^T) \right]^{-1} \quad \text{— (2)}$$

$$\frac{\partial L}{\partial \tilde{\Sigma}^{-1}} = \sum_{i=1}^{I} \mathrm{diag}\left[\tilde{\Sigma} - x_i x_i^T - \tilde{\mu}\tilde{\mu}^T - \phi\, E(h_i h_i^T)\, \tilde{\phi}^T + 2 x_i \tilde{\mu}^T - 2\tilde{\mu}\, E(h_i)^T \tilde{\phi}^T + 2 x_i\, E(h_i)^T \tilde{\phi}^T \right]$$

$$= 0$$

~~Perform ...~~

or $$\sum_{i=1}^{I} \mathrm{diag}\left[(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T - x_i x_i^T - \tilde{\mu}\tilde{\mu}^T - \phi\, E(h_i h_i^T)\, \tilde{\phi}^T + 2 x_i \tilde{\mu}^T - 2\tilde{\mu}\, E(h_i)^T \tilde{\phi}^T + 2 x_i\, E(h_i)^T \tilde{\phi}^T \right] = 0$$

or $$I.\tilde{\Sigma} = \sum_{i=1}^{I} \mathrm{diag}\left[(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T + \tilde{\phi}\, E(h_i h_i^T)\, \tilde{\phi}^T - 2(x_i - \tilde{\mu})\, E(h_i)^T \tilde{\phi}^T \right]$$

$$= \mathrm{diag}\left[\sum_{i=1}^{I} (x_i - \tilde{\mu})(x_i - \tilde{\mu})^T + \tilde{\phi} \sum_{i=1}^{I} E(h_i h_i^T)\, \tilde{\phi}^T - 2 \sum_{i=1}^{I} (x_i - \tilde{\mu})\, E(h_i)^T \tilde{\phi}^T \right].$$

— (3)

Put the results of ② in ③.

We get :

$$\text{I. } \tilde{\Sigma} = \text{diag}\Big(\sum_{i=1}^{I} (x_i - \tilde{\mu})(x_i - \tilde{\mu})^T + \Big[\sum_{i=1}^{I} (x_i - \tilde{\mu}) E(h_i)^T\Big]\Big[\sum_{i=1}^{I} E(h_i h_i^T)\Big]^{-1} \Big[\sum_{i=1}^{I} E(h_i h_i^T)\Big] \tilde{\phi}^T - 2 \sum_{i=1}^{I} (x_i - \tilde{\mu}) E(h_i)^T \tilde{\phi}^T \Big).$$

$$\boxed{\tilde{\Sigma} = \frac{1}{I} \sum_{i=1}^{I} \text{diag}\left[(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T - \tilde{\phi} E(h_i)(x_i - \tilde{\mu})^T \right]} \quad \text{—— ④}$$

ⓒ If μ is initialized with the emprical mean, $\boxed{\mu^{(0)} = \frac{1}{I} \sum_{i=1}^{I} x_i}$

$$\sum_{i=1}^{I} E(h_i) = \sum_{i=1}^{I} (\phi^T \Sigma^{-1} \phi + I)^{-1} \phi^T \Sigma^{-1} \Big(x_i - \frac{1}{I} \sum_{j=1}^{I} x_j \Big)$$

$$= (\phi^T \Sigma^{-1} \phi + I)^{-1} \phi^T \Sigma^{-1} \sum_{i=1}^{I} \Big(x_i - \frac{1}{I} \sum_{j=1}^{I} x_j \Big)$$

$$= (\phi^T \Sigma^{-1} \phi + I)^{-1} \phi^T \Sigma^{-1} \Big(\sum_{i=1}^{I} x_i - \frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{I} x_j \Big)$$

$$= (\phi^T \Sigma^{-1} \phi + I)^{-1} \phi^T \Sigma^{-1} \Big(\sum_{i=1}^{I} x_i - \sum_{j=1}^{I} x_j \Big)$$

$$= 0$$

So, for $\tilde{\mu} = \frac{1}{I} \sum_{i=1}^{I} (x_i - \tilde{\Phi} E(h_i))$.

$$= \frac{1}{I} \sum_{i=1}^{I} x_i$$

It means that mean always remains its initial value.

(d)

If we initialize the model with diagonal co-variance, then $\tilde{\mu}$, $\tilde{\Phi}$ and $\tilde{\Sigma}$ never change, and they remain in their initial values. Hence this is not beneficial.

$$\boxed{\begin{array}{l} \text{If } \phi^{(0)} = 0 \\ E(h_i) = 0 \end{array}}$$

Manzil Roy.

Dhananjay Bhandiwad.