Lecture 15 EM Algorithm & Factor Analysis

< Recap >:

E-step: $Q_i(z^{(i)}) = P(z^{(i)} | x^{(i)}; \theta)$

M-step: $\theta := \arg\max_\theta \underbrace{\sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}}_{\text{lower bound}}$

1. From EM to Mixture of Gaussian model.

< Recap Mixture of Gaussian model >:

$P(x^{(i)}, z^{(i)}) = P(x^{(i)} | z^{(i)}) \cdot P(z^{(i)})$

$z^{(i)} \sim \text{Multinomial}(\phi) \Rightarrow P(z^{(i)} = j) = \phi_j$

$x^{(i)} | z^{(i)} = j \sim N(\mu_j, \Sigma_j)$

E-step: $Q_i(z^{(i)}) = P(z^{(i)} | x^{(i)}; \theta) = P(z^{(i)} = j | x^{(i)}; \phi, \mu, \Sigma)$

$\Rightarrow w_j^{(i)} = Q_i(z^{(i)}) = P(z^{(i)} = j | x^{(i)}, \phi, \mu, \Sigma)$

M-step: $\max_{\phi,\mu,\Sigma} \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{\boxed{p(x^{(i)}, z^{(i)}; \phi, \mu, \Sigma)}}{Q_i(z^{(i)})}$

$p(x^{(i)} | z^{(i)}; \cdots)$ $\qquad$ $p(z^{(i)})$

$$= \sum_i \sum_j w_j^{(i)} \log \frac{\boxed{\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_j|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_j)^T \Sigma_j^{-1} (x^{(i)}-\mu_j)\right)} \cdot \boxed{\phi_j}}{w_j^{(i)}}$$

$\Rightarrow \nabla_{\mu_j}(\cdots) \overset{\text{set}}{=} 0 \Rightarrow \mu_j = \frac{\sum_i w_j^{(i)} x^{(i)}}{\sum_i w_j^{(i)}}$ $\leftarrow$ strength with which $x^{(i)}$ is assigned to Gaussian j

"$p(z^{(i)} = j | x^{(i)}; \cdots)$"

$\nabla_{\phi_j}(\cdots) \overset{\text{set}}{=} 0 \Rightarrow \phi_j = \frac{\sum_i w_j^{(i)}}{\sum_i \sum_j w_j^{(i)}}$, $\nabla_{\Sigma_j}(\cdots) \overset{\text{set}}{=} 0 \Rightarrow \Sigma_j =$

2. Why EM converge?

$J(\theta, Q) = \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$

We know $l(\theta) \geq J(\theta, Q)$, for any θ, Q

E-step: Maximize J wrt. Q_i $\Rightarrow$ coordinate ascend

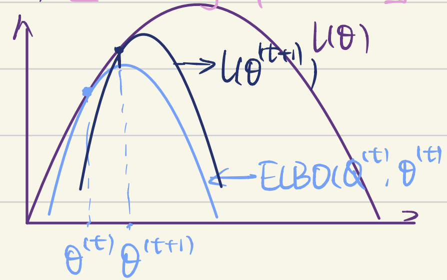
M-step: Maximize J wrt. θ

$\Rightarrow$ plot $J(\theta, Q) / l(\theta)$ to see if convergence

$l(\theta^{(t+1)}) \geq \mathrm{ELBO}(x; Q^{(t)}, \theta^{(t+1)})$ [Jensen]

$\geq \mathrm{ELBO}(x; Q^{(t)}, \theta^{(t)})$ [argmax M-step]

$= l(\theta^{(t)})$ [Corollary of Jensen]



3. when $m \approx n$ or $m \ll n$.

If model as a single Gaussian: $x \sim N(\mu, \Sigma)$

MLE: $\mu = \frac{1}{m} \sum_i^m x^{(i)}$

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \mu)(x^{(i)} - \mu)^T$$ $\Rightarrow$ singular / non-invertible

① Option 1: constrain Σ to be diagonal.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & \ddots & \sigma_n^2 \end{bmatrix}$$

$\Rightarrow$ MLE: $\sigma_j^2 = \frac{1}{m} \sum_i (x_j^{(i)} - \mu_j)^2$

But this assumes that the n features are completely independent.

② Option 2: construct Σ to be $\Sigma = \sigma^2 I$

i.e. $\Sigma = \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & \ddots & \sigma^2 \end{bmatrix}$ $\Rightarrow$ MLE: $\sigma^2 = \frac{1}{m} \frac{1}{n} \sum_i \sum_j (x_j^{(i)} - \mu_j)^2$

Only 1 parameter, even worse.

③ Option 3: Factor Analysis

4. Factor Analysis

$P(x, z) = P(x|z) \cdot P(z)$, z is hidden

$z \sim N(0, I)$, $z \in \mathbb{R}^d$ $(d < n)$

$x = \mu + \Lambda z + \varepsilon$, where $\varepsilon \sim N(0, \Psi)$ → read sai

Equivalently: $x|z \sim N(\mu + \Lambda z, \Psi)$

Parameters:

$\mu \in \mathbb{R}^n$, $\Lambda \in \mathbb{R}^{n \times d}$, $\Psi \in \mathbb{R}^{n \times n}$ diagonal

- Multivariate Gaussian

$$x = \begin{bmatrix} x_1 \\ \hline x_2 \end{bmatrix} \begin{matrix} \updownarrow r \\ \updownarrow s \end{matrix} \qquad x_1 \in \mathbb{R}^r, \; x_2 \in \mathbb{R}^s, \; x \in \mathbb{R}^{r+s}$$

$$x \sim N(\mu, \Sigma), \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{array}{cc} \overset{r}{\longleftrightarrow} & \overset{s}{\longleftrightarrow} \\ \end{array} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Marginal $P(x_1) = ?$

$P(x) = P(x_1, x_2)$

$\int_{x_2} P(x_1, x_2)\, dx_2 = P(x_1)$

$\Rightarrow \frac{1}{(2\pi)^{\frac{5}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}^T \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right) \Rightarrow x_1 \sim N(\mu_1, \Sigma_{11})$

Conditional $P(x_1 | x_2) = ?$ (equals to $\frac{P(x_1, x_2)}{P(x_2)}$)

$\Rightarrow x_1 | x_2 \sim N(\mu_{1|2}, \Sigma_{1|2})$

$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$

$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$

- Derive EM algorithm for Factor Analysis

① Derive $P(x, z)$

$\begin{pmatrix} z \\ x \end{pmatrix} \sim N(\mu_{x,z}, \Sigma)$

$z \sim N(0, I) \Rightarrow E(z) = 0$

$x = \mu + \Lambda z + \varepsilon \Rightarrow E(x) = E[\mu + \Lambda \overset{0}{z} + \overset{0}{\varepsilon}] = \mu$

$$\Rightarrow \mu_{xz} = \begin{array}{cc} \begin{bmatrix} 0 \\ \mu \end{bmatrix} & \begin{matrix} \updownarrow d \\ \updownarrow n \end{matrix} \end{array}, \quad \boxed{\Sigma = \begin{array}{c} \begin{bmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \\ \begin{matrix} \leftrightarrow & \leftrightarrow \\ d & n \end{matrix} \end{array}} = \begin{bmatrix} E(z-Ez)(z-Ez)^T & E(z-Ez)(x-Ex) \\ E(x-Ex)(z-Ez)^T & E(x-Ex)(x-Ex)^T \end{bmatrix}$$

<proof Σ_{22}>

$$\begin{aligned} \Sigma_{22} &= E(x-Ex)(x-Ex)^T = E(\Lambda z+\cancel{\mu}+\varepsilon-\cancel{\mu})(\Lambda z+\cancel{\mu}+\varepsilon-\cancel{\mu})^T \\ &= E[\Lambda z z^T\Lambda^T + \underline{\Lambda z\varepsilon^T + \varepsilon z^T\Lambda^T} + \varepsilon\varepsilon^T] \\ &= E[\Lambda z z^T \Lambda^T] + {}^0\, E[\varepsilon\varepsilon^T]^0 \\ &= \Lambda E[zz^T]\Lambda^T + \Psi = \Lambda\Lambda^T + \Psi \end{aligned}$$

$$\Sigma = \begin{bmatrix} I & \Lambda^T \\ \Lambda & \Lambda\Lambda^T+\Psi \end{bmatrix}$$

$$\Rightarrow \begin{pmatrix} z \\ x \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} I & \Lambda^T \\ \Lambda & \Lambda\Lambda^T+\Psi \end{bmatrix}\right)$$

② E-step:

$$Q_i(z^{(i)}) = P(z^{(i)} \mid x^{(i)}; \theta)$$

$$z^{(i)} \mid x^{(i)} \sim N(\mu_{z^{(i)}|x^{(i)}}, \Sigma_{z^{(i)}|x^{(i)}}),$$

where $\mu_{z^{(i)}|x^{(i)}} = \vec{0} + \Lambda^T(\Lambda\Lambda^T+\Psi)^{-1}(x^{(i)}-\mu)$

$$\Sigma_{z^{(i)}|x^{(i)}} = I - \Lambda^T(\Lambda\Lambda^T+\Psi)^{-1}\Lambda$$

③ M-step:

$$Q_i(z^{(i)}) = \frac{1}{(2\pi)^{d/2}|\cdots|} \exp\left(-\frac{1}{2}(\cdots)\right)$$

$$\int_{z^{(i)}} Q_i(z^{(i)})\, z^{(i)}\, dz^{(i)} = \int_{z^{(i)}} \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\cdots)\right)\cdot z^{(i)}\, dz^{(i)}$$

$\because E z = \int P(z)\cdot z\, dz$ $\quad E[z^{(i)}] = \mu_{z^{(i)}|x^{(i)}},\ z^{(i)} \sim Q_i$

$$\theta := \arg\max_{\theta} \sum_i \int_{z^{(i)}} Q_i(z^{(i)}) \log \frac{P(x^{(i)}, z^{(i)})}{Q_i(z^{(i)})}\, d(z^{(i)})$$

$$= \sum_i E_{z^{(i)}\sim Q_i}\left[\log \frac{P(x^{(i)}, z^{(i)})}{Q_i(z^{(i)})}\right]$$

plug in Gaussian density.

$\because \log \cdots \exp(\cdots) \Rightarrow$ quadratic function

$\nabla \mu(\cdots) \overset{\text{set}}{=} 0 \quad - \cdots$