Data: $\{x_1, x_2, \ldots, x_n\}$ $\quad x_i \in \mathbb{R}^d$

$(\pi_1, \pi_2, \ldots, \pi_K)$ $\quad \sum_{i=1} \pi_i = 1$ $\quad Pr[z_n = k] = \pi_k$

$x_i | z_i = k \sim \text{Normal}(\mu_k, \Sigma_k)$

Learn: $(\pi_1, \pi_2, \ldots, \pi_K)$ & $(\mu_1, \Sigma_1), (\mu_2, \Sigma_2) \ldots$
(μ_K, Σ_K)

① Expectation - Step : Posterior $Pr[z_n = k)$
$= \gamma_{nk}$

Poster $\propto$ Prior $\times$ Likelihood

$$\gamma_{nk} = \pi_k \cdot N(x_n | \mu_k, \Sigma_k)$$

↳ PDF of Normal

$$\sum_{k=1}^{K} \gamma_{nk} = 1$$

↳ normalization

② Maximization-Step.

$$\mathcal{L}_k^{\mu,\Sigma} = \sum_{n=1}^{N} \gamma_{nk} \log \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$$

$$\frac{\partial \mathcal{L}_k^{\mu,\Sigma}}{\partial \mu_k} = 0 \qquad \frac{\partial \mathcal{L}_k^{\mu,\Sigma}}{\partial \Sigma_k} = 0$$

$$\Rightarrow \mu_k = \frac{\sum_{n=1}^{N} \gamma_{nk} x_n}{N_k}, \quad N_k = \sum_{n=1}^{N} \gamma_{nk}$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} (x_n - \mu_k)(x_n - \mu_k)^T$$

$$\pi_k = \frac{N_k}{N}, \quad \sum_{k=1}^{K} \pi_k = 1$$

Repeat E-step & M-step until convergence $\square$.

$$\left.\begin{array}{l} \text{Doc } 1: \; w_1 \; w_2 \cdots w_n \\ \text{Doc } 2: \; w_1, w_2 \cdots w_n \\ \vdots \\ \text{Doc } N_d: \; w_1 \; w_2 \cdots w_n \end{array}\right\} \text{Data}$$

$\alpha \longrightarrow \theta$: Distribution of topics for each Doc, $\longrightarrow z_n$:

$\alpha \in [0,1]^K$

$\theta \in$ Simplex$(K-1)$

$\sum_{i=1}^{K} \theta_i = 1$

$0 \le \theta_i \le 1$

$z_n \in \{1, 2, \cdots, K\}$ Topic of each word

$\theta \sim \text{Dirichlet}(\alpha)$

$z_n \sim \text{Multinomial}(\theta)$

$\beta \in \mathbb{R}^{|V| \times K}$

$\beta, z_n \longrightarrow w_n$

$$\Pr[w_n = i \mid z_n = j] = \beta_{ij}$$

$$\text{Data} \xrightarrow{\text{learn}} (\alpha, \beta)$$

We want to infer $P(\theta, z \mid w, \alpha, \beta)$

intractable

$$q(\theta \mid \gamma) \prod_{i=1}^{n} q(z_i \mid \phi_i)$$

close

Dirichlet(γ) Multinomial

$$KL[q \| p] = -\mathbb{E}_q\left[\log\left(\frac{p}{q}\right)\right] \geq 0$$

KL-Divergence

"=" $\Leftrightarrow$ $p = q$

$$KL[q \| p] \neq KL[p \| q]$$

$$KL[q\|p] = -\int \Big\{ \log[p(\theta, z \mid w, \alpha, \beta)]$$

$$- \log q(\theta \mid \gamma)$$

$$- \sum_{n=1}^{N} \log q(z_n \mid \phi_n) \Big\}$$

$$q(\theta \mid \gamma) \prod_{n=1}^{N} q(z_n \mid \phi_n)$$

$$d\theta \, dz_1 \, dz_2 \ldots dz_n$$

Variational

Inference