



## Hierarchical Model

$$(1) \mu_k \sim N(0, \sigma^2) \quad k = 1, \dots, K$$

$$p(\mu_k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu_k^2}{2\sigma^2}}$$

$$(2) c_i \sim \text{categorical}(\frac{1}{K}, \dots, \frac{1}{K}) \quad i = 1, \dots, n$$

$$(3) x_i | c_i, \mu \sim N(c_i^T \mu, 1) \quad i = 1, \dots, n$$

$$p(x_i | c_i, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \mu_{c_i})^2}{2}}$$

$\mu = (\mu_1, \dots, \mu_K)$

$c_i = (0, \dots, \underset{k}{1}, \dots, 0)$

$\underbrace{\qquad\qquad\qquad}_{K\text{-length vector}}$   $\underset{k}{\text{k}^{\text{th}} \text{ location}}$

if data point  $x_i$   
belongs to cluster  $k$ .

$\Rightarrow c_i^T \mu = \mu_k$

Data =  $x_1, \dots, x_n$

Sampled in the following way:

(1) Assume  $\mu$  known

(we've already sampled  $\mu_k \sim N(0, \sigma^2)$ )

(2) Step 1: randomly select a cluster  
 $k=1, \dots, K$  all equally likely

Assume our selection is cluster  $k$ .

(3) Step 2: Sample  $x_i \sim N(\mu_k, 1)$ .

Joint Density:

$$\begin{aligned} p(\mu, c, x) &= p(x | c, \mu) \underbrace{p(c, \mu)}_{\text{indep.}} \\ &= \underbrace{p(x | c, \mu)}_{\text{ind.}} \underbrace{p(c)}_{\text{ind.}} p(\mu) \\ &= \prod_{i=1}^n p(x_i | c_i, \mu) p(c_i) p(\mu) \end{aligned}$$

Marginal is found by integrating a joint.

$$p(x) = \int p(x, y) dy$$

---

$$(8) \quad p(x) = \iint_{\mathcal{C} \times \mathcal{M}} q(x, \mu, \sigma) d\sigma d\mu$$

---

$$(10) \quad q^*(z) = \arg \min_{q(z) \in \mathcal{Q}} KL(q(z) || p(z|x))$$

Using  
(12)

$$= \arg \min_{q(z) \in \mathcal{Q}} \left[ \mathbb{E}(\log q(z)) - \mathbb{E}(\log p(z|x)) + \log p(x) \right]$$

$$= \arg \min_{q(z) \in \mathcal{Q}} \left[ \mathbb{E}(\log q(z)) - \mathbb{E}(\log p(z|x)) \right]$$

—————  
— ELBO(q)

$$= \arg \max_{q(z) \in \mathcal{Q}} \mathbb{E} \text{ELBO}(q)$$

Model:

$$\underbrace{p(z, x)}_{\text{joint}} = \underbrace{p(z)}_{\substack{\text{prior} \\ \uparrow \\ \text{latent variable}}} p(x|z) \quad \text{on pg 2.}$$

(13)  $\text{ELBO}(q) = \mathbb{E}[\log p(z, x)] - \mathbb{E}[\log q(z)]$

$$= \mathbb{E}[\log p(z) p(x|z)] - \mathbb{E}[\log q(z)]$$
$$= \mathbb{E}[\log p(z)] + \underbrace{\mathbb{E}[\log p(x|z)]}_{\substack{\text{expected log} \\ \text{likelihood}}} - \mathbb{E}[\log q(z)]$$

$$\mathbb{E}[\log q(\mu_k); m_k, s_k^2]$$

$$\mu_k \sim N(m_k, s_k^2)$$

$$= \int q(\mu_k; m_k, s_k^2) \log p(\mu_k) d\mu_k$$
$$= \int \frac{1}{\sqrt{2\pi s_k^2}} e^{-\frac{(\mu_k - m_k)^2}{2s_k^2}} \log \left[ \frac{1}{\sqrt{2\pi s_k^2}} e^{-\frac{\mu_k^2}{2s_k^2}} \right] d\mu_k$$

$$y_{ik} \propto e^{\mathbb{E}(\mu_k) x_i - \mathbb{E}(\mu_k^2)/2}$$

$$\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iK})$$

to represent a prob. dist.

need  $\sum_{k=1}^K y_{ik} = 1$

discrete dist:

$$\mathbf{p} = (p_1, p_2, p_3)$$

$$\frac{4}{8}, \frac{3}{8}, \frac{1}{8}$$

$$\tilde{\mathbf{p}} = (4, 3, 1)$$

$$\mathbf{P} = \frac{\tilde{\mathbf{p}}}{\text{sum}(\tilde{\mathbf{p}})}$$

$$\mathbf{p} \propto \tilde{\mathbf{p}}$$

$$\mathbb{E} \left[ \log_p(x_i | c_i, \mu) ; \gamma_i, m, s^2 \right]$$

$$\rightarrow x_i | c_i, \mu \sim \mathcal{N}(c_i^\top \mu, 1)$$

$$p(x_i | c_i, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - c_i^\top \mu)^2}$$

$$= \prod_{k=1}^K \left[ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu_k)^2} \right]^{c_{ik}}$$

$$c_i = (0, 0, \dots, \underset{k^{\text{th}} \text{ location}}{1}, \dots, 0)$$

$$c_i^\top \mu = \mu_k$$

$$q(c_i; \gamma_i) = \text{the } k^{\text{th}} \text{ position} = 1 \text{ w.p. } \gamma_{ik}$$

$$q(\mu; m, s^2) = \prod_{k=1}^K q(\mu_k; m_k, s_k^2)$$

Gaussian

$$\int \int q(c_i; \gamma_i) q(\mu; m, s^2) \sum_{k=1}^K c_{ik} \left( \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2}(x_i - \mu_k)^2 \right)$$

$c_i, \mu$

$$= \quad \cdot$$

$$\mathbb{E} \left[ \log p(x_i | c_i, \mu) \right] = \mathbb{E}_{c_i} \mathbb{E}_\mu \left[ \log p(x_i | c_i, \mu) \right]$$

$$= -\mathbb{E}_\mu \mathbb{E}_{c_i} \left[ \sum_k c_{ik} \left( \frac{1}{2} (x_i - \mu_k)^2 \right) \right] + \text{const}$$

$$= -\mathbb{E}_\mu \left[ \sum_k g_{ik} \left( \frac{1}{2} (x_i - \mu_k)^2 \right) \right] + \text{const}$$

$$\rightarrow \frac{1}{2} \mathbb{E}_{c_i} \left[ \sum_k c_{ik} (x_i - \mu_k)^2 \right] =$$

$$= \frac{1}{2} \sum_{k'} g_{ik'} \sum_k \underbrace{\mathbf{1}\{c_{ik'} = 1 \cap c_{ik} = 0\}}_{\text{when } k \neq k'} \{c_{ik}(x_i - \mu_k)^2\}$$

$$= \frac{1}{2} \sum_{k'} g_{ik'} (x_i - \mu_{k'})^2$$

$$\sum c_{ik} \mathbf{1}\{c_{ik'} = 1 \text{ and } c_{ik} = 0 \text{ for } k \neq k'\} (x_i - \mu_k)^2$$

$$\mathbf{1} = \begin{cases} 0 & \text{if } k \neq k' \\ 1 & \text{if } k = k' \end{cases}$$

$$X = \begin{cases} 1 & p_1 \\ 2 & p_2 \\ 3 & p_3 \end{cases} \quad \mathbb{E}(X) = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3$$

$$\mathbf{1}\{A\} = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false} \end{cases}$$

$$c_i = \begin{cases} (1, 0, 0) & \text{w.p. } \varphi_{i1} \\ (0, 1, 0) & \text{w.p. } \varphi_{i2} \\ (0, 0, 1) & \text{w.p. } \varphi_{i3} \end{cases} = q(c_i; \varphi_i)$$

$$\mathbb{E}_{c_i} \left[ \sum_{k=1}^3 c_{ik} \Delta_k \right] = \mathbb{E}_{c_i} [c_{i1}\Delta_1 + c_{i2}\Delta_2 + c_{i3}\Delta_3]$$

const. wrt  $c_i$

$$= \varphi_{i1} \left[ \underbrace{c_{i1}\Delta_1}_{=1} \right] + \varphi_{i2} \left[ \underbrace{c_{i2}\Delta_2}_{=1} \right] + \varphi_{i3} \left[ \underbrace{c_{i3}\Delta_3}_{=1} \right]$$

$$= \sum_{k'=1}^3 \varphi_{ik'} \underbrace{\Delta_{k'}}_{\Delta_k = (x_i - \mu_k)^2}$$

$$\frac{1}{2} \mathbb{E}_{c_i} \left[ \sum_{k=1}^K c_{ik} (x_i - \mu_k)^2 \right] = \frac{1}{2} \sum_{k=1}^K \varphi_{ik} (x_i - \mu_k)^2$$

---


$$\frac{1}{2} \mathbb{E}_{c_i} \left[ (x_i - c_i^\top \mu)^2 \right]$$

$$= \frac{1}{2} \left[ \varphi_{i1} (x_i - \mu_1)^2 + \varphi_{i2} (x_i - \mu_2)^2 \right.$$

$$+ \dots + \varphi_{iK} (x_i - \mu_K)^2 \left. \right] q(c_i; \varphi_i) = \begin{cases} (1, 0, \dots, 0) & \varphi_{i1} \\ (0, 1, \dots, 0) & \varphi_{i2} \\ \vdots & \vdots \\ (0, 0, \dots, 1) & \varphi_{iK} \end{cases}$$

$$= \frac{1}{2} \sum_k \varphi_{ik} (x_i - \mu_k)^2$$

$$\mu = (\mu_1, \dots, \mu_K)$$

$$c_i = (c_{i1}, \dots, c_{iK})$$

$$\begin{aligned}
\mathbb{E} \left[ \log p(x_i | c_i, \mu) \right] &= \mathbb{E}_\mu \mathbb{E}_{c_i} \left[ -\frac{1}{2} \log(2\pi) - \frac{1}{2} (x_i - c_i^\top \mu)^2 \right] \\
&= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \mathbb{E}_\mu \mathbb{E}_{c_i} \left[ (x_i - c_i^\top \mu)^2 \right] \\
&= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \mathbb{E}_\mu \left[ \sum_{k=1}^K \varphi_{ik} (x_i - \mu_k)^2 \right] \\
&= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \sum_{k=1}^K \varphi_{ik} \mathbb{E}_{\mu_k} [(x_i - \mu_k)^2] \\
&= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \sum_{k=1}^K \varphi_{ik} \int_{\mu_k} q(\mu_k; m_k, s_k^2) (x_i - \mu_k)^2 d\mu_k \\
&= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \sum_{k=1}^K \varphi_{ik} \left[ x_i^2 - 2x_i m_k + s_k^2 + m_k^2 \right] \\
&= -\frac{1}{2} \log(2\pi) - \frac{1}{2} x_i^2 + x_i \sum_k \varphi_{ik} m_k - \frac{1}{2} \sum_k \varphi_{ik} (s_k^2 + m_k^2)
\end{aligned}$$


---


$$\mathbb{E}_{c_i} \left[ (x_i - c_i^\top \mu)^2 \right] = \varphi_{i1} (x_i - (1, \dots, 0)^\top \mu)^2 + \varphi_{i2} (x_i - (0, 1, \dots, 0)^\top \mu)^2 + \dots + \varphi_{iK} (x_i - (0, \dots, 0, 1)^\top \mu)^2 = \varphi_{i1} (x_i - \mu_1)^2 + \varphi_{i2} (x_i - \mu_2)^2 + \dots + \varphi_{iK} (x_i - \mu_K)^2$$

$$c_i^\top \mu = \sum_k c_{ik} \mu_k$$

$$= (c_{i1}, c_{i2}, \dots, c_{iK}) \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix}$$

$$\mathbb{E}_{\mu_K}[(x_i - \mu_K)^2] = \mathbb{E}_{\mu_K}[x_i^2 - 2x_i\mu_K + \mu_K^2]$$

$$= x_i^2 - 2x_i \mathbb{E}[\mu_K] + \mathbb{E}[\mu_K^2]$$

$$q(\mu_K; \mu_K, s_K^2) = x_i^2 - 2x_i \mu_K + s_K^2 + \mu_K^2$$

$$\sim \mathcal{N}(\mu_K, s_K^2)$$

$$\mu_K \sim \mathcal{N}(\mu_K, s_K^2)$$

↑      ↓  
mean      variance

$$\mathbb{E}[x] = \mu \quad x \sim \mathcal{N}(\mu, \sigma^2)$$

mean      variance

$$\mathbb{E}[x^2] = \sigma^2 + \mu^2$$

$$\text{Var}(x) = \mathbb{E}x^2 - (\mathbb{E}x)^2$$

definition of variance.

$$\sum_{k=1}^K \varphi_{ik} x_i^2 = x_i^2 \sum_{k=1}^K \varphi_{ik} = 'x_i^2$$

$$(y_{i1} x_i^2 + y_{i2} x_i^2 + \dots + y_{ik} x_i^2)$$

$$= x_i^2 (y_{i1} + y_{i2} + \dots + y_{ik})$$

$$\begin{aligned}
 & -\frac{1}{2} \sum_{k=1}^K y_{ik} (x_i^2 - 2x_i m_k + s_k^2 + m_k^2) \\
 & = -\frac{1}{2} \sum_k y_{ik} x_i^2 + \underbrace{\sum_k x_i y_{ik} m_k}_{\text{circled}} - \frac{1}{2} \sum_k y_{ik} (s_k^2 + m_k^2) \\
 & = -\frac{1}{2} x_i^2 + x_i \sum_k y_{ik} m_k - \frac{1}{2} \sum_k y_{ik} (s_k^2 + m_k^2)
 \end{aligned}$$