we have input labeled data (training) $\{x_i, t_i\}_{i=1}^{N}$

$\uparrow$ Categorical

To estimate the data likelihood $P(x \mid C_k)$:

STEP 1: Observed data likelihood $\mathcal{L}^0$

$$\mathcal{L}^0 = P(x_1, x_2, \dots, x_N \mid C_k)$$

$$= \prod_{n=1}^{N} P(x_n \mid C_k)$$

$$= \prod_{n=1}^{N} \frac{1}{(2\pi)^{1/2} \cdot |\Sigma_k|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x_n - \mu_k)^T \cdot \Sigma_k^{-1} (x_n - \mu_k)\right\}$$

i.i.d.

assume multivariate Gaussian likelihood

For simplification, consider

$$\Sigma_k = \sigma_k^2 \cdot I$$

$$\mathcal{L}^{o} = \prod_{n=1}^{N} \frac{1}{(2\pi)^{1/2} \left| \sigma_k^2 \cdot I \right|^{1/2}} \cdot \mathrm{Exp}\left\{ -\frac{1}{2\sigma_k^2} (x_n - \mu_k)^T (x_n - \mu_k) \right\}$$

$$= \prod_{n=1}^{N} \frac{1}{(2\pi)^{1/2} \cdot (\sigma_k^2)^{1/2}} \cdot \mathrm{Exp}\left\{ -\frac{1}{2\sigma_k^2} (x_n - \mu_k)^T (x_n - \mu_k) \right\}$$

STEP 2 : Take the log-likelihood, $\mathcal{L}$.

$$\mathcal{L} = \ln \mathcal{L}^{o}$$

$$= \sum_{n=1}^{N} \ln\left(\frac{1}{(2\pi)^{1/2} (\sigma_k^2)^{1/2}} \right) - \frac{1}{2\sigma_k^2} (x_n - \mu_k)^T (x_n - \mu_k)$$

$$= \sum_{n=1}^{N} -\underline{\frac{1}{2} \ln(2\pi)} - \underline{\frac{1}{2} \ln(\sigma_k^2)} - \frac{1}{2\sigma_k^2} (x_n - \mu_k)^T (x_n - \mu_k)$$

STEP 3 : optimize log-likelihood to find the parameters μ_k, $\Sigma_k' = \sigma_k^2 I$. $\forall$ k (classes).

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = 0 \iff \boxed{\mu_k = \frac{1}{N_k} \sum_{n \in C_k}' x_n}$$

↳ sample mean of samples in class k.

$$\frac{\partial \mathcal{L}}{\partial \sigma_k^2} = 0$$

$$\Leftrightarrow \sum_{n=1}^{N} -\frac{1}{\cancel{2}} \cdot \frac{1}{\cancel{\sigma_k^2}} + \frac{2(x_n-\mu_k)^T(x_n-\mu_k)}{(2\sigma_k^2)^{\cancel{2}}} = 0$$

$$\Leftrightarrow \sum_{n=1}^{N} \left(-\frac{1}{} + \frac{(x_n-\mu_k)^T \cdot (x_n-\mu_k)}{\sigma_k^2} \right) = 0$$

$$\boxed{\Leftrightarrow \sum_{n=1}^{N} \frac{(x_n-\mu_k)^T(x_n-\mu_k)}{N} = \sigma_k^2}$$

↳ sample variance of data

$$\Sigma_k = \sigma_k^2 \cdot I$$

Mixture Model

↳ to fit data likelihood

some probabilistic model, e.g. Gaussian, Exponential, Bernoulli

$$\underline{\underline{P(x|\theta)}} = \sum_{k=1}^{K} \pi_k \cdot \underline{\underline{P}}(x|\theta_k)$$

$\pi_k \equiv$ weight / prob. of each mixture dist.

$$0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1$$

Gaussian Mixture Model (GMM)

$$P(x \mid \underline{\underline{\theta}}) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

$$\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}, \quad 0 \leq \pi_k \leq 1, \quad \sum_k \pi_k = 1$$

Data Likelihood w/ GMM:

1 $$\mathcal{L}^{\phi} = \prod_{n=1}^{N} p(x_n | \theta)$$

$$p(x_n | \theta) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n | \mu_k, \Sigma_k)$$

$$\mathcal{L}^{o} = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n | \mu_k, \Sigma_k)$$

② Log-likelihood

$$\mathcal{L} = \ln\left(\prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right)$$

$$= \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right)$$

③ $\frac{\partial \mathcal{L}}{\partial \mu_k} = 0, \quad \frac{\partial \mathcal{L}}{\partial \Sigma_k} = 0, \quad \frac{\partial \mathcal{L}}{\partial \pi_k} = 0$

Expectation - Maximization (EM) algorithm to optimize $\mathcal{L}$.

Example: $\{x_i\}_{i=1}^{N}$ are collected with some sensor

Data $x_1, x_2, \ldots, x_n$ is censored.

$$f(x) = \begin{cases} x, & \text{if } x \leq a \\ a, & \text{if } x > a \end{cases}$$

samples from $x_1 - x_m$ less than a, $x_{m+1} - x_n$ larger than a

$$\mathcal{L}^{0} = P(x_1, x_2, \ldots, x_n \mid \theta)$$

$$= P(x_1 \mid \theta) \cdot P(x_2 \mid \theta) \ldots\ldots P(x_n \mid \theta)$$

$$= \prod_{i=1}^{m} P(x_i \mid \theta) \cdot \prod_{j=m+1}^{N} \int_{a}^{+\infty} p(x_j \mid \theta) \cdot dx_j$$

EM algorithm introduces hidden latent variables

Hidden variable

$z_j \equiv$ true value for x_j, $j = m+1, \dots, N$

$$\mathcal{L}^{o} = \prod_{i=1}^{m} p(x_i|\theta) \cdot \prod_{j=m+1}^{N} p(z_j|\theta)$$

$$\boxed{\underset{\theta^{o}}{\arg\max}\; E_{z}[\ln \mathcal{L}^{o}]}$$