Maximum Likelihood Estimate

$$L(\mu, \underline{\sigma}, \underline{\pi}) = -\ln P(X)$$

$$= -\ln \prod_n \sum_k P(\underline{X}_n | k) \cdot P(k)$$

$$= -\sum_n \left(\ln \sum_k P(\underline{X}_n | k) \cdot P(k) \right)$$

$$= -\sum_n \left(\ln \sum_k e^{\ln P(X_n | k) \cdot P(k)} \right)$$

$$= -\sum_n \left(\ln \sum_k e^{\ln P(X_n | k) + \ln P(k)} \right)$$

$$= -\sum_{cols} \left(\underbrace{\ln \sum_{rows} e^{(P + \ln \underline{\pi})}}_{\text{log-sum-exp}} \right)$$

K-Means

$$X = \begin{bmatrix} - & \underline{x}_1^T & - \\ & \vdots & \\ - & \underline{x}_N^T & - \end{bmatrix} \qquad \mu = \begin{bmatrix} - & \underline{\mu}_1 & - \\ & \vdots & \\ - & \underline{\mu}_K & - \end{bmatrix}$$

Let the distance matrix be

$$D = \left[\|\underline{x}_i - \underline{\mu}_j\|^2 \right] = \begin{bmatrix} \overline{\underline{x}_1\underline{\mu}_1} & \cdots & \overline{\underline{x}_1\underline{\mu}_K} \\ \vdots & & \\ \overline{\underline{x}_N\underline{\mu}_1} & \cdots & \overline{\underline{x}_N\underline{\mu}_n} \end{bmatrix}$$

where $\overline{\underline{x}_i\underline{\mu}_j} = \|\underline{x}_i - \underline{\mu}_j\|^2$

$$\Rightarrow d_{ij} = \|\underline{x}_i - \underline{\mu}_j\|^2$$

$$= \sum_d (x_{id} - \mu_{jd})^2 = \sum_d (x_{id}^2 - 2x_{id}\mu_{jd} + \mu_{jd}^2)$$

$$= \sum_d x_{id}^2 \cdot 1_{dj} - 2x_{id}\mu_{dj}^T + 1_{id}(\mu_{dj}^T)^2$$

$$\Rightarrow \boxed{D = X^2 \cdot 1 - 2X\mu^T + 1(\mu^T)^2}$$

GMM

$$X = \begin{bmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_N^T \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1^T \\ \vdots \\ \mu_K^T \end{bmatrix} \qquad \underline{\sigma} = [\sigma_1 \cdots \sigma_K] \qquad \underline{\pi} = [P(1) \cdots P(K)]$$

$$P = [\ln P(\underline{x}_i \mid j)] = \ln \begin{bmatrix} P(\underline{x}_1 \mid 1) & \cdots & P(\underline{x}_1 \mid K) \\ \vdots & & \vdots \\ P(\underline{x}_N \mid 1) & \cdots & P(\underline{x}_N \mid K) \end{bmatrix}$$

where $P(\underline{x}_i \mid j) = \mathcal{N}(\underline{x}_i; \mu_j, \sigma_j^2)$

Let P be the log gaussian matrix

$$\Rightarrow \quad P_{ij} = \ln \mathcal{N}(\underline{x}_i; \mu_j, \sigma_j^2)$$

where $\begin{cases} \Sigma_j = \sigma_j^2 I \\ |\Sigma_j| = \sigma_j^{2d} \\ \Sigma_j^{-1} = \frac{1}{\sigma_j^2} I \end{cases}$ (due to independence of RVs and same standard deviation)

$$= \ln \left[\frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} \cdot e^{-\frac{1}{2}(\underline{x}_i - \mu_j)^T \Sigma_j^{-1} (\underline{x}_i - \mu_j)} \right]$$

$$= \ln \left[\frac{1}{(\sigma_j \sqrt{2\pi})^d} \cdot e^{-\frac{1}{2\sigma_j^2}(\underline{x}_i - \mu_j)^2} \right]$$

$$= -d \ln(\sigma_j \sqrt{2\pi}) - \frac{1}{2\sigma_j^2} (\underline{x}_i - \mu_j)^2 = -d \ln(\sigma_j \sqrt{2\pi}) - \frac{1}{2\sigma_j^2} \sum_d (x_{id} - \mu_{jd})^2$$

$$= -d \ln(\sigma_j \sqrt{2\pi}) - \frac{1}{2\sigma_j^2} \sum_d \left[x_{id}^2 \cdot 1_{dj} - 2 x_{id} \mu_{dj}^T + 1_{id} (\mu_{dj}^T)^2 \right]$$

$$\Rightarrow \boxed{P = -d \ln(\dot{\underline{\sigma}} \sqrt{2\pi}) - \frac{1}{2\dot{\underline{\sigma}}^2} \otimes \left[X^2 \cdot 1 - 2X\mu^T + 1(\mu^T)^2 \right]}$$

Similarly let Q be the log posterior matrix

$$Q = [\ln P(j \mid \underline{x}_i)]$$

$$\Rightarrow q_{ij} = \ln \left[\frac{P(\underline{x}_i \mid j) \cdot P(j)}{\sum_1^K P(\underline{x}_i \mid k) \cdot P(k)} \right] = \ln P(\underline{x}_i \mid j) + \ln P(j) - \ln \left[\sum_1^K P(\underline{x}_i \mid k) \cdot P(k) \right]$$

$$= \ln P(\underline{x}_i \mid j) + \ln P(j) - \ln \left[\sum_{k=1}^K e^{\ln P(\underline{x}_i \mid k) \cdot P(k)} \right]$$

$$= \ln P(\underline{x}_i \mid j) + \ln P(j) - \ln \left[\sum e^{\ln P(\underline{x}_i \mid k) + \ln P(k)} \right]$$

$$\Rightarrow \boxed{Q = P + \ln \dot{\underline{\pi}} - \underbrace{\ln \sum_{\text{rows}} e^{(P + \ln \dot{\underline{\pi}})}}_{\text{log-sum-exp } \circledast}}$$