Example solution

Minimize $L(\gamma, \theta)$ by setting the gradient of L to $\vec{0}$.

$$
\begin{aligned}
\frac{\partial L}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i}\left(\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{kn} \| x^{(n)} - \theta_k \|_2^2 \right) \\
&= \frac{\partial}{\partial \theta_i}\left(\sum_{k=1}^{K} \sum_{n=1}^{N} \gamma_{kn} (x^{(n)} - \theta_k)^T (x^{(n)} - \theta_k) \right) \\
&= \frac{\partial}{\partial \theta_i}\left(\sum_{n=1}^{N} \gamma_{in} (x^{(n)} - \theta_i)^T (x^{(n)} - \theta_i) \right) \\
&= \sum_{n=1}^{N} \gamma_{in} \quad 2 (x^{(n)} - \theta_i)(-1) \\
&= -2 \sum_{n=1}^{N} \gamma_{in} x^{(n)} - 2 \sum_{n=1}^{N} \gamma_{in} \theta_i \\
&= -2 \sum_{n=1}^{N} \gamma_{in} x^{(n)} - 2 \theta_i \sum_{n=1}^{N} \gamma_{in} \overset{!}{=} 0
\end{aligned}
$$

$$
\Rightarrow \boxed{\hat{\theta}_i = \frac{\sum_{n=1}^{N} \gamma_{in} x^{(n)}}{\sum_{n=1}^{N} \gamma_{in}}}
$$

- $\sum_{n=1}^{N} \gamma_{in} x^{n}$:= the sum of the data point assigned to cluster i
- $\sum_{n=1}^{N} \gamma_{in}$:= the number 〃

$\Rightarrow \hat{\theta}_i$:= the mean 〃

This is exactly the cluster update step in K-means.

Check if $(\hat{\theta}_1, \dots, \hat{\theta}_K)$ is indeed a minimum. The Hessian of L is

$$
\mathcal{H}(L) = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_1^2} & \cdots & \frac{\partial^2 L}{\partial \theta_1 \partial \theta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \theta_K \partial \theta_1} & \cdots & \frac{\partial^2 L}{\partial \theta_K^2} \end{bmatrix} = \begin{bmatrix} 2 \sum_{n=1}^{N} \gamma_{1n} & & 0 \\ & \ddots & \\ 0 & & 2 \sum_{n=1}^{N} \gamma_{Kn} \end{bmatrix}
$$

Since

$$
\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \frac{\partial}{\partial \theta_j}\left(-2 \sum_{n=1}^{N} \gamma_{in} x^{(n)} + 2 \theta_i \sum_{n=1}^{N} \gamma_{in} \right) = 0
$$

$$
\frac{\partial^2 L}{\partial \theta_i^2} = \frac{\partial}{\partial \theta_i}\left(\quad〃\quad \right) = 2 \sum_{n=1}^{N} \gamma_{in} .
$$

$\sum_{n=1}^{N} \gamma_{in} > 0$ for all $i = 1, \dots, K$

$\Rightarrow$ all the eigenvalues of $\mathcal{H}(L) > 0$

$\Rightarrow$ $\mathcal{H}(L)$ is (symmetric) positive definite.

$\Rightarrow (\hat{\theta}_1, \dots, \hat{\theta}_K)$ is a global minimum.