GMM Numerical Problems

Prof. Pranay Kumar Saha
Academic Year: 2025-2026
Institution: Indian Institute of Technology (Indian School of Mines) Dhanbad,Jharkhand.

Problem 1: One-Dimensional GMM (Single EM Iteration)

A dataset has four one-dimensional observations:

$$x = [1.0,\ 2.0,\ 5.0,\ 6.0].$$

We assume a Gaussian mixture model with two components ($K = 2$) and initial parameters:

$$\pi_1 = \pi_2 = 0.5, \quad \mu_1 = 2.0, \quad \mu_2 = 5.0, \quad \sigma_1^2 = \sigma_2^2 = 1.0.$$

Tasks:

1. **E-step:** Compute the responsibilities

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\sum_{j=1}^{2} \pi_j \mathcal{N}(x_i \mid \mu_j, \sigma_j^2)},$$

   for each data point x_i and component $k = 1, 2$.

2. **M-step:** Update parameters using:

$$\pi_k^{(new)} = \frac{1}{n}\sum_{i=1}^{n} \gamma_{ik}, \qquad \mu_k^{(new)} = \frac{\sum_{i=1}^{n} \gamma_{ik} x_i}{\sum_{i=1}^{n} \gamma_{ik}}, \qquad \sigma_k^{2,(new)} = \frac{\sum_{i=1}^{n} \gamma_{ik}(x_i - \mu_k^{(new)})^2}{\sum_{i=1}^{n} \gamma_{ik}}.$$

3. **Interpretation:** Identify which cluster each data point most likely belongs to based on the computed responsibilities.

Expected Outcome: The first component ($k = 1$) should primarily explain points near $x = 1, 2$, while the second component ($k = 2$) should explain points near $x = 5, 6$.

Problem 2: The E-Step (Calculating Responsibilities)

You have a GMM with two 1D components (Cluster 1 and Cluster 2) and the following *current* parameters:

- **Cluster 1:** Mean (μ_1) = 15, Variance (σ_1^2) = 4, Mixture Weight (π_1) = 0.5
- **Cluster 2:** Mean (μ_2) = 25, Variance (σ_2^2) = 9, Mixture Weight (π_2) = 0.5

You observe a new data point, $x = 17$.

Recall the 1D Gaussian probability formula:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Useful values (you can use these to simplify):

- $\sqrt{2\pi} \approx 2.507$
- $e^{-0.5} \approx 0.607$
- $e^{-4.44} \approx 0.012$

Your Task: Calculate the **responsibility** (the soft assignment) of Cluster 1 and Cluster 2 for the data point $x = 17$.

1. Calculate the likelihood of $x = 17$ under Cluster 1: $\mathcal{N}(17|15, 4)$.
2. Calculate the likelihood of $x = 17$ under Cluster 2: $\mathcal{N}(17|25, 9)$.
3. Calculate the total probability $p(x)$ by adding the *weighted* likelihoods:
$$p(x) = (\pi_1 \times \mathcal{N}_1) + (\pi_2 \times \mathcal{N}_2)$$
4. Calculate the responsibility for each cluster:
   - Responsibility (Cluster 1) = $\frac{\pi_1 \times \mathcal{N}_1}{p(x)}$
   - Responsibility (Cluster 2) = $\frac{\pi_2 \times \mathcal{N}_2}{p(x)}$

Problem 3: The M-Step (Updating Parameters)

After running the E-step on a small dataset of $N = 4$ points, you get the following responsibilities for **Cluster A**:

Data Point (x_i)	Responsibility for Cluster A (γ_{iA})
5	0.9
8	0.8
10	0.7
20	0.2

Your Task: Use these responsibilities to calculate the *new* parameters for Cluster A.

1. **Calculate the new mean (μ_A^{new}):**
   Hint: This is the weighted average of the data points.
$$\mu_A^{\text{new}} = \frac{\sum(\gamma_{iA} \times x_i)}{\sum \gamma_{iA}}$$

2. **Calculate the new mixture weight (π_A^{new}):**
   Hint: This is the average responsibility for this cluster across all N data points.

$$\pi_A^{\text{new}} = \frac{\sum \gamma_{iA}}{N}$$