Assignment 1

Chetan Goel
CS-GY 6613: Artificial Intelligence, NYU Tandon

September 30, 2025

1 Setup

Likelihood for a point x_i:

$$p(x_i \mid \pi, \mu, \sigma^2) = \pi_1 \mathcal{N}(x_i \mid \mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(x_i \mid \mu_2, \sigma_2^2).$$

Log-likelihood for m data points:

$$\mathcal{L} = \sum_{i=1}^{m} \log \Big(\sum_{k=1}^{2} \pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2) \Big).$$

Responsibility of component k for data point i:

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

2 Derivation: Mean gradient

For any single data point x_i let

$$S_i \;=\; \sum_{j=1}^{2} \pi_j \, \mathcal{N}(x_i \mid \mu_j, \sigma_j^2).$$

Then,

$$\mathcal{L} = \sum_{i=1}^{m} \log \Big(\sum_{k=1}^{2} \pi_k \, \mathcal{N}(x_i \mid \mu_k, \sigma_k^2) \Big) = \sum_{i=1}^{m} \log S_i.$$

Differentiating $\mathcal{L}$ w.r.t. μ_k:

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_{i=1}^{m} \frac{1}{S_i} \cdot \frac{\partial}{\partial \mu_k} \big(\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2) \big) = \sum_{i=1}^{m} \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{S_i} \cdot \frac{\partial \log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\partial \mu_k}.$$

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_{i=1}^{m} \frac{\pi_k \, \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\sum_{j=1}^{2} \pi_j \, \mathcal{N}(x_i \mid \mu_j, \sigma_j^2)} \cdot \frac{\partial \log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\partial \mu_k}$$

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_{i=1}^{m} \gamma_{ik} \cdot \frac{\partial \log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\partial \mu_k}.$$

For a univariate Gaussian,

$$\log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2) = -\tfrac{1}{2} \log(2\pi\sigma_k^2) - \frac{(x_i - \mu_k)^2}{2\sigma_k^2},$$

so

$$\frac{\partial \log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\partial \mu_k} = \frac{x_i - \mu_k}{\sigma_k^2}.$$

Finally,

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_{i=1}^{m} \gamma_{ik} \, \frac{(x_i - \mu_k)}{\sigma_k^2}$$

3 Derivation: Variance gradient

Differentiating the log-likelihood w.r.t. σ_k^2:

$$\frac{\partial \mathcal{L}}{\partial \sigma_k^2} = \sum_{i=1}^{m} \frac{1}{S_i} \cdot \frac{\partial}{\partial \sigma_k^2} \big(\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)\big) = \sum_{i=1}^{m} \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{S_i} \cdot \frac{\partial \log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\partial \sigma_k^2}.$$

For the univariate Gaussian,

$$\frac{\partial \log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\partial \sigma_k^2} = -\frac{1}{2\sigma_k^2} + \frac{(x_i - \mu_k)^2}{2(\sigma_k^2)^2}.$$

Following the same approach as the mean gradient derivation, we apply the chain rule and observe that $\frac{\pi_k \mathcal{N}(x_i|\mu_k, \sigma_k^2)}{S_i}$ is exactly γ_{ik},

$$\frac{\partial \mathcal{L}}{\partial \sigma_k^2} = \sum_{i=1}^{m} \gamma_{ik} \left(-\frac{1}{2\sigma_k^2} + \frac{(x_i - \mu_k)^2}{2(\sigma_k^2)^2} \right).$$

Rearranging,

$$\frac{\partial \mathcal{L}}{\partial \sigma_k^2} = \frac{1}{2} \sum_{i=1}^{m} \gamma_{ik} \left[\frac{(x_i - \mu_k)^2}{\sigma_k^4} - \frac{1}{\sigma_k^2} \right]$$

4 Derivation: Mixture-weights gradient

Differentiate $\mathcal{L}$ with respect to π_k:

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = \frac{\partial}{\partial \pi_k} \sum_{i=1}^{m} \log S_i = \sum_{i=1}^{m} \frac{\partial \log S_i}{\partial \pi_k}.$$

For each x_i,

$$\frac{\partial}{\partial \pi_k} \log S_i = \frac{1}{S_i} \cdot \frac{\partial}{\partial \pi_k} (\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)) = \frac{\mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{S_i}.$$

Thus

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = \sum_{i=1}^{m} \frac{\mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{S_i}.$$

According to the definition of Responsibility,

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{S_i}.$$

Rearranging,

$$\frac{\mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{S_i} = \frac{\gamma_{ik}}{\pi_k},$$

Finally,

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = \sum_{i=1}^{m} \frac{\gamma_{ik}}{\pi_k}$$