

GMM Appendix C: Deriving the E-Step: From Indicators to Responsibilities

Starting Point: The Complete Data Log-Likelihood¹

[Appendix B](#) showed that if the cluster assignments $z_1, \dots, z_n$ were known, the log-likelihood of the data *and* the assignments together would be

$$\log p(X, Z \mid \theta) = \sum_{i=1}^n \sum_{k=1}^K \mathbb{1}(z_i = k) [\log \pi_k + \log \mathcal{N}(x_i \mid \mu_k, \Sigma_k)],$$

where the indicator $\mathbb{1}(z_i = k)$ is 1 if $z_i = k$ and 0 otherwise, so for each point exactly one term of the inner sum survives. No sum sits inside a log, which is what makes this form easy to maximize.

1. Courtesy of Claude.ai

The E-Step: Taking Expectations

The problem is that we don't observe Z , so we can't compute the complete data log-likelihood directly. The EM algorithm solves this by taking the **expected value** of the complete data log-likelihood with respect to the distribution of Z given X and the current parameters $\theta^{(t)}$.

The Q-Function

The lecture calls the result the *expected complete-data log-likelihood*; its standard name is the Q -function:

$$Q(\theta \mid \theta^{(t)}) = \mathbb{E}_{Z|X, \theta^{(t)}} [\log p(X, Z \mid \theta)]$$

Substituting our expression:

$$Q(\theta \mid \theta^{(t)}) = \mathbb{E}_{Z|X, \theta^{(t)}} \left[\sum_{i=1}^n \sum_{k=1}^K \mathbb{1}(z_i = k) [\log \pi_k + \log \mathcal{N}(x_i \mid \mu_k, \Sigma_k)] \right]$$

Computing the Expected Value of the Indicator

Now we need to compute:

$$\mathbb{E}_{Z|X, \theta^{(t)}} [\mathbb{1}(z_i = k)]$$

What is this expectation?

The expectation of an indicator variable is just the probability that the indicator equals 1:

$$\mathbb{E}[\mathbb{1}(z_i = k)] = 1 \cdot P(z_i = k) + 0 \cdot P(z_i \neq k) = P(z_i = k)$$

But we need the **conditional** expectation given X and $\theta^{(t)}$:

$$\mathbb{E}_{Z|X, \theta^{(t)}} [\mathbb{1}(z_i = k)] = P(z_i = k \mid X, \theta^{(t)})$$

Factorization of the conditional distribution

Defining the Responsibility

We define the **responsibility** of cluster k for point i as:

$$\gamma_{ik} = P(z_i = k \mid x_i, \theta^{(t)})$$

So we've shown that:

$$\mathbb{E}_{Z|X, \theta^{(t)}} [\mathbb{1}(z_i = k)] = \gamma_{ik}$$

Computing the Responsibility Using Bayes' Rule

Now we need to actually compute $\gamma_{ik} = P(z_i = k \mid x_i, \theta^{(t)})$.

Bayes' Rule

$$P(z_i = k \mid x_i, \theta^{(t)}) = \frac{p(x_i \mid z_i = k, \theta^{(t)}) \cdot P(z_i = k \mid \theta^{(t)})}{p(x_i \mid \theta^{(t)})}$$

Breaking down each term:

1. **Prior:** $P(z_i = k \mid \theta^{(t)}) = \pi_k^{(t)}$ (the mixing weight)
2. **Likelihood:** $p(x_i \mid z_i = k, \theta^{(t)}) = \mathcal{N}(x_i \mid \mu_k^{(t)}, \Sigma_k^{(t)})$
3. **Evidence** (marginal likelihood):

The Complete E-Step

Putting it all together, the Q-function becomes:

$$Q(\theta \mid \theta^{(t)}) = \sum_{i=1}^n \sum_{k=1}^K \gamma_{ik} [\log \pi_k + \log \mathcal{N}(x_i \mid \mu_k, \Sigma_k)]$$

where:

$$\gamma_{ik} = \frac{\pi_k^{(t)} \mathcal{N}(x_i \mid \mu_k^{(t)}, \Sigma_k^{(t)})}{\sum_{j=1}^K \pi_j^{(t)} \mathcal{N}(x_i \mid \mu_j^{(t)}, \Sigma_j^{(t)})}$$

Intuitive Interpretation

The Indicator Variable Perspective

- If we knew $z_i = k$, then $\mathbb{1}(z_i = k) = 1$ and this term contributes fully to the log-likelihood
- If we knew $z_i \neq k$, then $\mathbb{1}(z_i = k) = 0$ and this term contributes nothing

The Responsibility Perspective

- Since we don't know z_i , we use a **soft assignment** $\gamma_{ik} \in [0, 1]$
- γ_{ik} is the probability that point i came from cluster k
- If $\gamma_{ik} = 0.8$, then point i contributes 80% to cluster k 's statistics
- The responsibilities for each point sum to 1: $\sum_{k=1}^K \gamma_{ik} = 1$

Mathematical Equivalence

Numerical Example

► Code

```
Likelihoods  $p(x_i \mid z_i=k, \theta)$ :
```

```
Component 1: 0.002915
```

```
Component 2: 0.058550
```

```
Component 3: 0.000239
```

```
Numerators  $\pi_k * p(x_i \mid z_i=k, \theta)$ :
```

```
Component 1: 0.000875
```

```
Component 2: 0.029275
```

```
Component 3: 0.000048
```

```
Denominator (marginal): 0.030197
```

```
Responsibilities  $\gamma_{ik}$ :
```

```
Component 1: 0.028960 (2.9%)
```

```
Component 2: 0.969455 (96.9%)
```

```
Component 3: 0.001585 (0.2%)
```

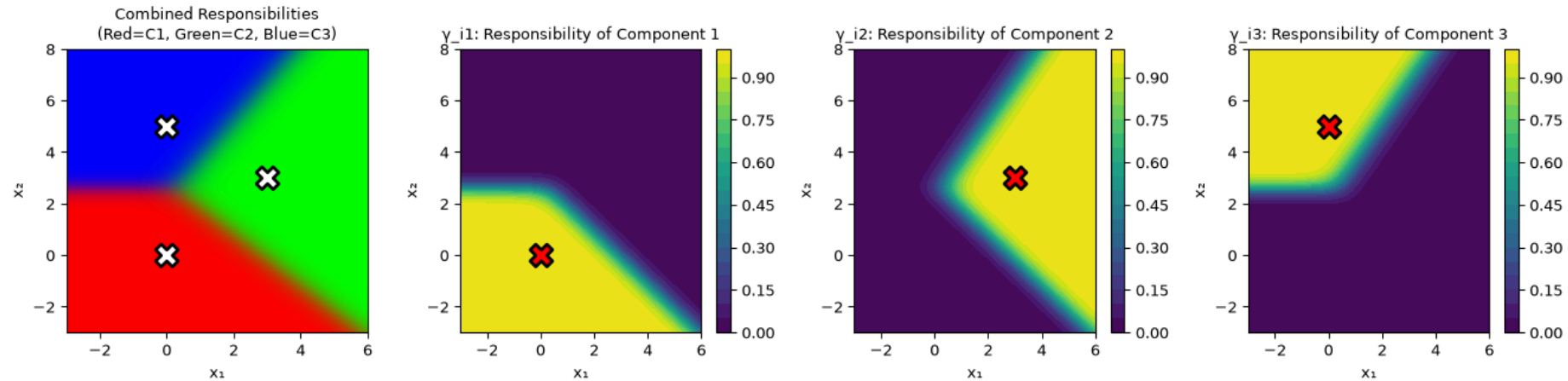
```
Sum of responsibilities: 1.000000
```

```
Interpretation:
```

```
- . . . . .
```

Visualization of Responsibilities

► Code



Key Properties of Responsibilities

1. Normalization

For each point i :

$$\sum_{k=1}^K \gamma_{ik} = 1$$

This follows from the law of total probability.

► Code

```
Verification that responsibilities sum to 1:  
Point 0: sum of  $\gamma$  = 1.0000000000  
Point 1: sum of  $\gamma$  = 1.0000000000  
Point 2: sum of  $\gamma$  = 1.0000000000  
Point 3: sum of  $\gamma$  = 1.0000000000  
Point 4: sum of  $\gamma$  = 1.0000000000
```

2. Soft Assignment

Summary

We've derived that:

$$\mathbb{E}_{Z|X, \theta^{(t)}} [\mathbb{1}(z_i = k)] = \gamma_{ik} = P(z_i = k \mid x_i, \theta^{(t)})$$

And computed it using Bayes' rule:

$$\gamma_{ik} = \frac{\pi_k^{(t)} \mathcal{N}(x_i \mid \mu_k^{(t)}, \Sigma_k^{(t)})}{\sum_{j=1}^K \pi_j^{(t)} \mathcal{N}(x_i \mid \mu_j^{(t)}, \Sigma_j^{(t)})}$$

Key insight: Since we don't know which cluster generated each point, we replace the binary indicator $\mathbb{1}(z_i = k)$ with its expected value under the posterior distribution, which is the probability (responsibility) γ_{ik} .