Understanding Gaussian Mixture Models (GMM) with a Mathematical Example

March 11, 2025

Introduction

Gaussian Mixture Models (GMMs) are a powerful tool in unsupervised machine learning for clustering and density estimation. They assume that the data is generated from a mixture of several Gaussian (normal) distributions. Each Gaussian represents a cluster, and the model estimates the parameters of these Gaussians to best fit the data.

In this document, we will:

- Explain the basics of GMMs.
- Describe the Expectation-Maximization (EM) algorithm used to train GMMs.
- Walk through a step-by-step mathematical example to compute probabilities and update parameters.

Gaussian Mixture Model (GMM) Overview

A GMM is a probabilistic model that represents data as a combination of multiple Gaussian distributions. Each Gaussian component has three key parameters:

- **Mean (μ)**: The center of the Gaussian (cluster).
- **Covariance (Σ)**: The spread or shape of the Gaussian.
- **Mixing Coefficient (π)**: The weight of the Gaussian in the mixture. The sum of all mixing coefficients is 1.

The probability density function (PDF) of a GMM is given by:

$$p(x) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

Where:

- K: Number of Gaussian components (clusters).
- $\mathcal{N}(x \mid \mu_k, \Sigma_k)$: The PDF of the k^{th} Gaussian component.

Expectation-Maximization (EM) Algorithm

The EM algorithm is used to estimate the parameters (π_k, μ_k, Σ_k) of the GMM. It works in two steps:

1. Expectation Step (E-Step)

In this step, we calculate the **responsibilities** (r_{nk}), which represent the probability that a data point x_n belongs to cluster k. The formula is:

$$r_{nk} = \frac{\pi_k \cdot \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}$$

Here:

- π_k: Mixing coefficient of cluster k.
- $\mathcal{N}(x_n \mid \mu_k, \Sigma_k)$: Probability of x_n under the k^{th} Gaussian.

2. Maximization Step (M-Step)

In this step, we update the parameters of the GMM using the responsibilities calculated in the E-step:

- **Mixing Coefficients**:

$$\pi_k = \frac{N_k}{N}, \quad \text{where } N_k = \sum_{n=1}^{N} r_{nk}$$

- **Means**:

$$\mu_k = \frac{\sum_{n=1}^{N} r_{nk} x_n}{N_k}$$

- **Covariances**:

$$\Sigma_k = \frac{\sum_{n=1}^{N} r_{nk}(x_n - \mu_k)(x_n - \mu_k)^T}{N_k}$$

Mathematical Example: One-Dimensional GMM with Two Components

Let's work through a simple example to understand how GMMs work. We'll use a one-dimensional dataset with two Gaussian components.

Given Parameters

- **Component 1**:
  - Mean ($\mu_1 = 2$)
  - Variance ($\sigma_1^2 = 1$)
  - Mixing Coefficient ($\pi_1 = 0.6$)
- **Component 2**:
  - Mean ($\mu_2 = 5$)
  - Variance ($\sigma_2^2 = 2$)
  - Mixing Coefficient ($\pi_2 = 0.4$)

Step 1: Compute the Probability Density at $x = 3.5$

The Gaussian PDF for a single component is:

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

For $x = 3.5$:

- **Component 1**:

$$p_1(3.5) = \frac{1}{\sqrt{2\pi(1)}} e^{-\frac{(3.5-2)^2}{2(1)}} = 0.107$$

- **Component 2**:

$$p_2(3.5) = \frac{1}{\sqrt{2\pi(2)}} e^{-\frac{(3.5-5)^2}{2(2)}} = 0.120$$

The overall probability density at $x = 3.5$ is:

$$p(3.5) = \pi_1 \cdot p_1(3.5) + \pi_2 \cdot p_2(3.5) = (0.6)(0.107) + (0.4)(0.120) = 0.112$$

Step 2: Expectation Step (E-Step)

Calculate the responsibilities for $x = 3.5$:

- **Component 1**:

$$r_{n1} = \frac{\pi_1 \cdot p_1(3.5)}{\pi_1 \cdot p_1(3.5) + \pi_2 \cdot p_2(3.5)} = \frac{(0.6)(0.107)}{(0.6)(0.107) + (0.4)(0.120)} = 0.571$$

- **Component 2**:

$$r_{n2} = 1 - r_{n1} = 0.429$$

Step 3: Maximization Step (M-Step)

Update the parameters using the responsibilities:

- **Mixing Coefficients**:

$$\pi_1 = \frac{N_1}{N}, \quad \pi_2 = \frac{N_2}{N}$$

- **Means**:

$$\mu_1 = \frac{\sum_{n=1}^{N} r_{n1} x_n}{N_1}, \quad \mu_2 = \frac{\sum_{n=1}^{N} r_{n2} x_n}{N_2}$$

- **Covariances**:

$$\Sigma_1 = \frac{\sum_{n=1}^{N} r_{n1} (x_n - \mu_1)^2}{N_1}, \quad \Sigma_2 = \frac{\sum_{n=1}^{N} r_{n2} (x_n - \mu_2)^2}{N_2}$$

Gaussian Mixture Models are a flexible and powerful tool for clustering and density estimation. By using the EM algorithm, we can iteratively estimate the parameters of the Gaussians and assign data points to clusters. The example above demonstrates how to compute probabilities and update parameters step-by-step.