0.1 Gaussian Mixture Model (GMM)

Gaussian Mixture Model (GMM) 是一种基于概率模型的聚类算法，它假设数据由多个高斯分布的混合组成。与 K-Means 不同，GMM 通过概率来表示每个数据点属于不同簇的可能性，而不仅仅是将数据点硬性地分配给某个簇。

GMM 使用期望最大化 (Expectation-Maximization, EM) 算法来估计每个高斯分布的参数，包括均值、协方差矩阵和混合系数。EM 算法的主要步骤包括：

- 期望步骤 **(E-Step)**: 计算每个数据点属于每个高斯分布的后验概率（责任值）。
- 最大化步骤 **(M-Step)**: 使用这些概率更新每个高斯分布的参数，包括均值、协方差矩阵和混合系数。

GMM 的目标是通过以下概率密度函数描述数据：

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

其中：

- K 是高斯分布的数量（即簇数）。
- π_k 是第 k 个高斯分布的权重，满足 $\sum_{k=1}^{K} \pi_k = 1$。
- $\mathcal{N}(x|\mu_k, \Sigma_k)$ 是第 k 个高斯分布的概率密度函数，μ_k 为均值，Σ_k 为协方差矩阵。

GMM 的优点在于其灵活性，可以处理不同形状和大小的簇，并能进行软聚类，即一个数据点可以部分属于多个簇。然而，GMM 的计算复杂度较高，尤其是在高维数据集上。此外，GMM 对初始参数的选择较为敏感，可能会陷入局部最优。