

Applied machine learning: Gaussian Mixtures

Problem sheet 5

In this exercise, you will implement and apply the Expectation-Maximization (EM) algorithm to fit a Gaussian Mixture Model (GMM) to a dataset.

Tasks

1. Data Generation:

- Generate 1,000 samples from a mixture of three Gaussian distributions with the following parameters (you can try other later):
 - Weights: $[0.3, 0.2, 0.5]$
 - Means: $[-2, 1, 5]$
 - Standard deviations: $[0.5, 0.3, 1.5]$
- Combine the samples into a single dataset.
- Plot a histogram of the dataset to visualize the distribution.

2. Expectation Step:

- Implement the *Expectation step* of the EM algorithm to compute the responsibilities of each data point for each Gaussian component.
- Use the Gaussian probability density function (PDF) to calculate the likelihoods. You can use `norm.pdf` from `scipy.stats`.

3. Maximization Step:

- Implement the *Maximization step* to update the weights, means, and variances of the Gaussian components based on the responsibilities.

4. EM Algorithm:

- Initialize the parameters randomly.
- Combine the *Expectation* and *Maximization* steps into a loop to iteratively update the parameters.
- Run the algorithm for 100 iterations or until convergence.

5. Results and Visualization:

- Compare the true weights, means, and standard deviations to the estimated ones.
- Plot the estimated Gaussian components overlaid on the histogram of the data.