CONVERGENCE OF EM ALGORITHM FOR GAUSSIAN MIXTURE MODELS

AI61004 STAT AIML

STATEMENT OF OBJECTIVES

The speed of convergence of the **Expectation Maximization (EM)** algorithm for **Gaussian mixture model** fitting is known to be dependent on the amount of overlap among the mixture components. In this project, we study the impact of mixing coefficients on the convergence of EM. We show that when the mixture components exhibit some overlap, the convergence of EM becomes slower as the dynamic range among the mixing coefficients increases. We propose a **deterministic anti-annealing algorithm (DAAEM)**, that significantly improves the speed of convergence of EM for such mixtures with unbalanced mixing coefficients. However when the dataset has a relatively larger covariance (like that of the Iris dataset) , DAEM might get slow. We increase the speed by introducing a modification in DAEM algorithm without altering its convergence criteria.

The difference in Standard EM Algorithm and DAEM lies in the updated parameter ($h_j(t)$) in Expectation step. The Maximization step remains the same for all algorithms.

EM Algorithm:

$$h_j(t) = \frac{(\alpha_j P(x_t|\mu_j,\Sigma_j))}{\sum(\alpha_i P(x_t|\mu_j,\Sigma_j))}$$

DAEM Algorithm:

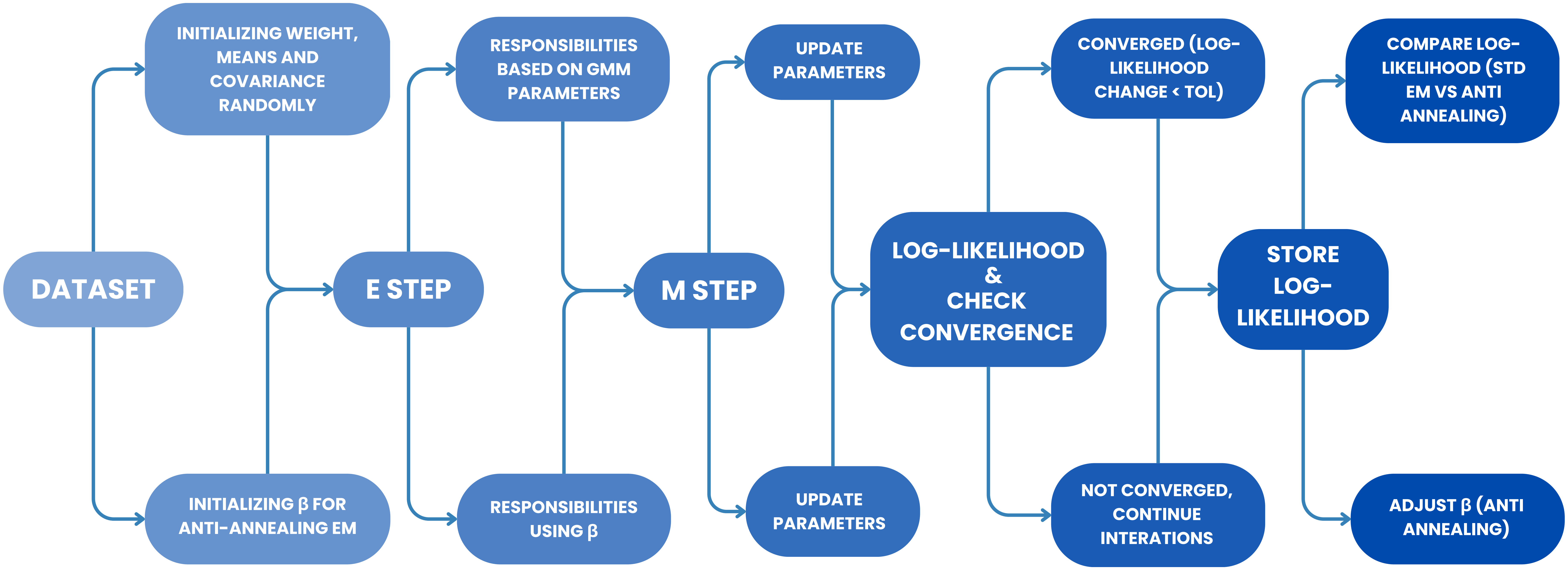
$$h_j(t) = \frac{(\alpha_j P(x_t|\mu_j,\Sigma_j))^{\beta}}{\sum(\alpha_i P(x_t|\mu_j,\Sigma_j))^{\beta}}$$

Modified DAEM Algorithm:

$$h_j(t) = \frac{(\alpha_j P(x_t|\mu_j,\Sigma_j))^{\beta+\alpha_j}}{\sum(\alpha_i P(x_t|\mu_j,\Sigma_j))^{\beta+\alpha_j}}$$

As the rate of convergence of the above-discussed algorithms increase, the number of computations as well as time complexity also increases. So for general purposes, the EM algorithm is preferred.

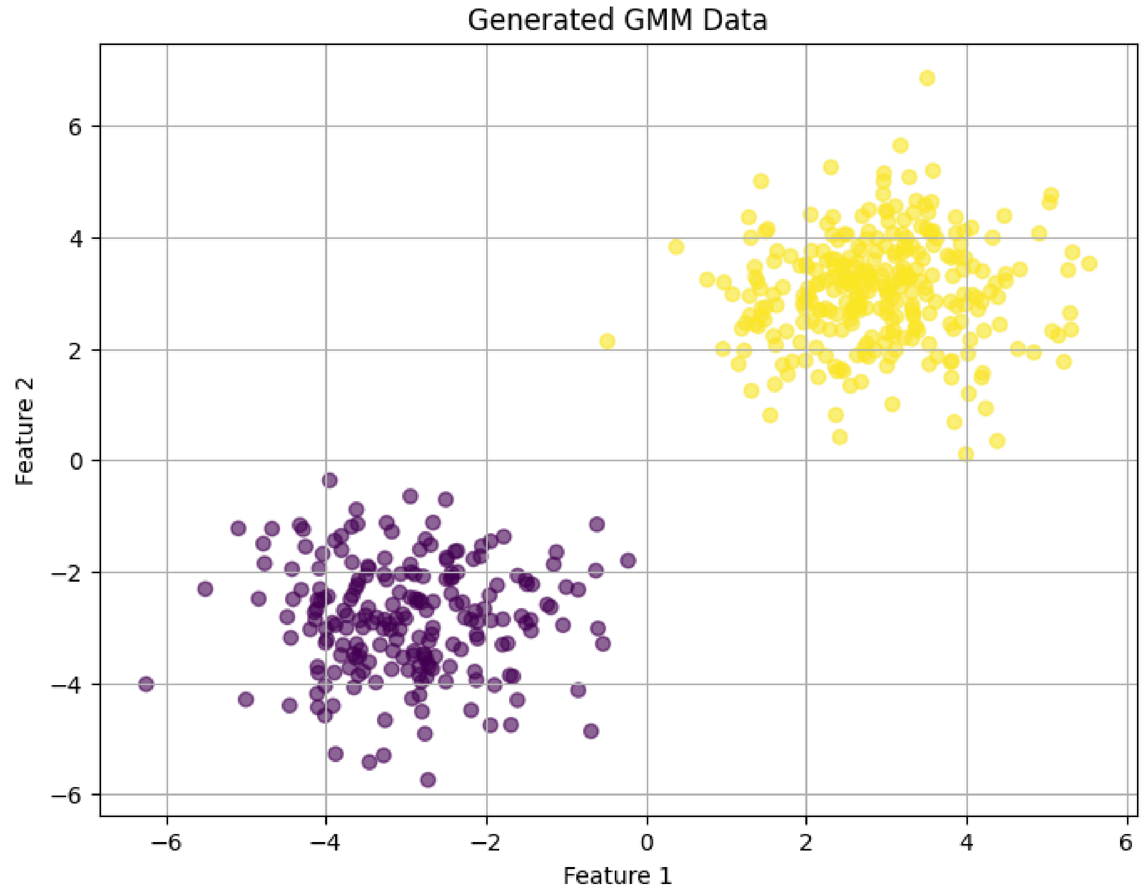
SOLUTION APPROACH FLOWCHART



MULTIVARIATE GAUSSIAN DISTRIBUTION SAMPLE

This dataset has two distinct clusters with very little or no overlap. We want to understand the performance of the algorithm when there is no major noise in the dataset

DATASET DESCRIPTION

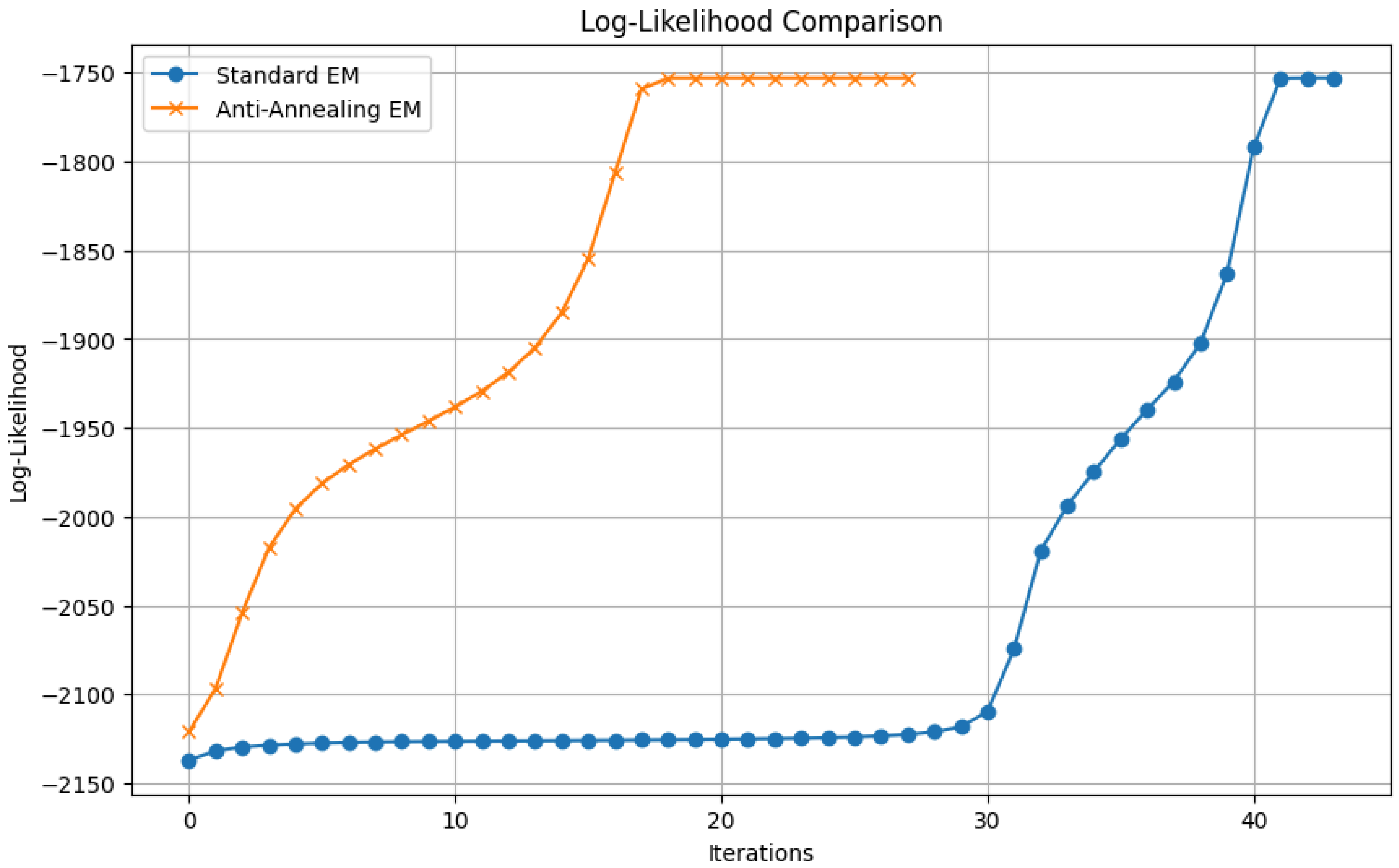


LEGEND

- Class 0
- Class 1

The dataset is a synthetic 2D Gaussian Mixture Model (GMM) with 500 samples generated from two components. The components have weights of 0.4 and 0.6, mean vectors [-3,-3][-3, -3] and [3,3][3, 3], and identity covariance matrices, representing independent features with unit variance. Each sample is labeled based on its generating component.

RESULTS

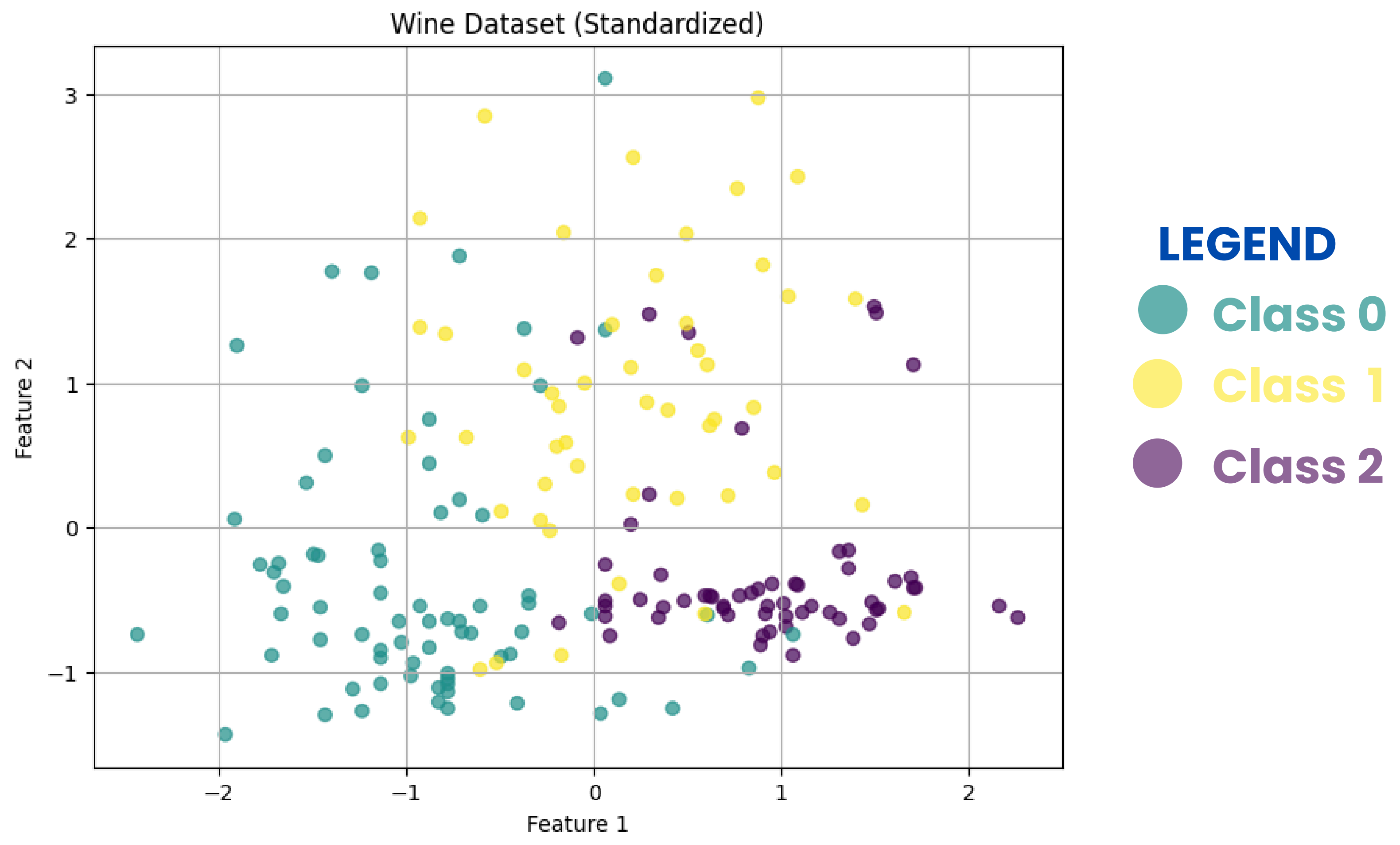


Therefore, we can say that the Deterministic Anti-Annealing EM (DAEM) algorithm gives faster convergence than the standard EM algorithm for multivariate Gaussian distribution. Next we will check the convergence rate for a noisy datasets like Wine aand Iris datasets and compare them.

WINE DATASET

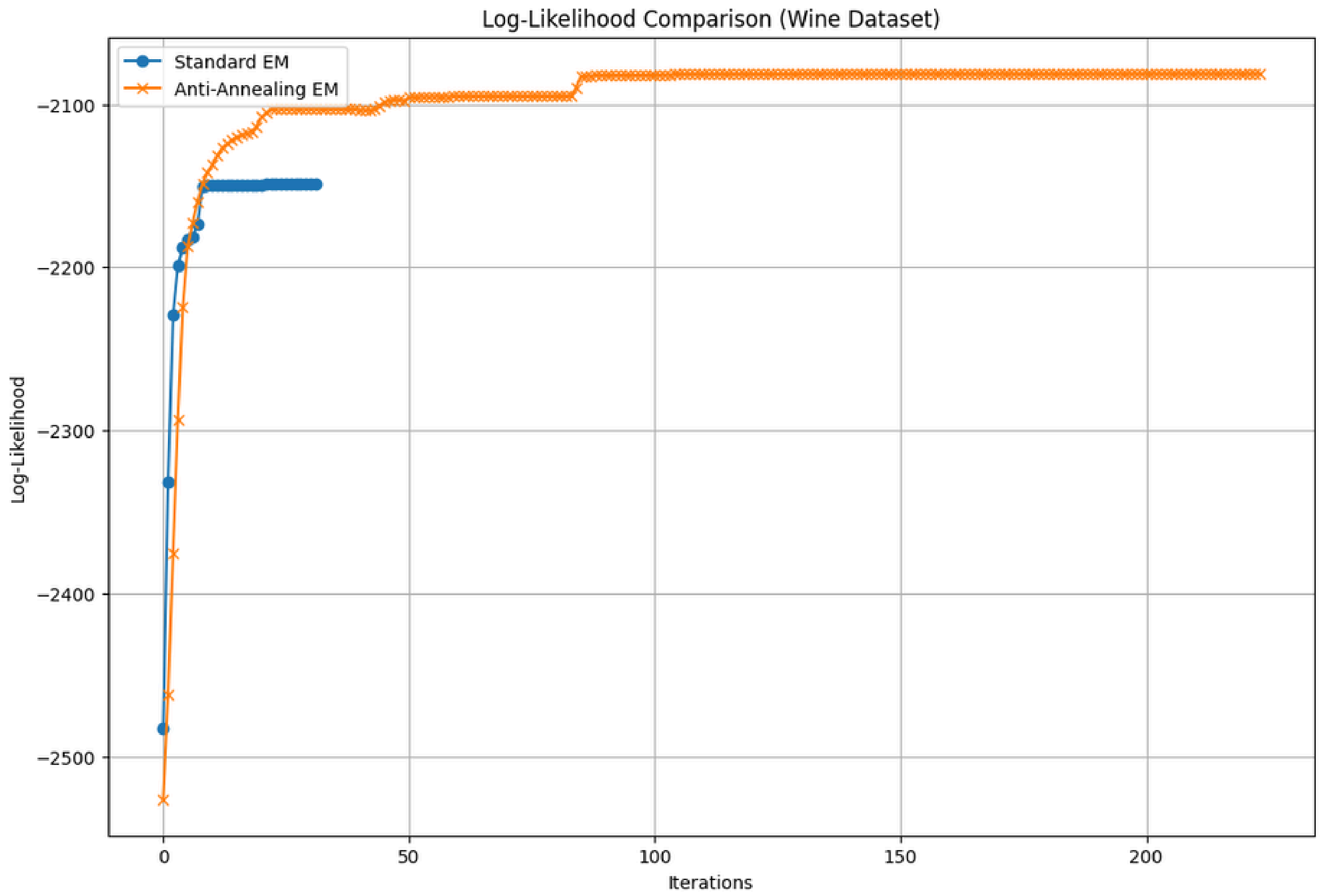
This dataset has significant overlap. We want to compare EM vs DAEM results for this dataset.

DATASET DESCRIPTION



Consists of 178 samples, each representing a wine characterized by 13 continuous features related to its chemical composition. These features are used to classify the wines into three distinct classes, Class 0 (Wine Type 1), Class 1 (Wine Type 2), and Class 2 (Wine Type 3).

RESULTS

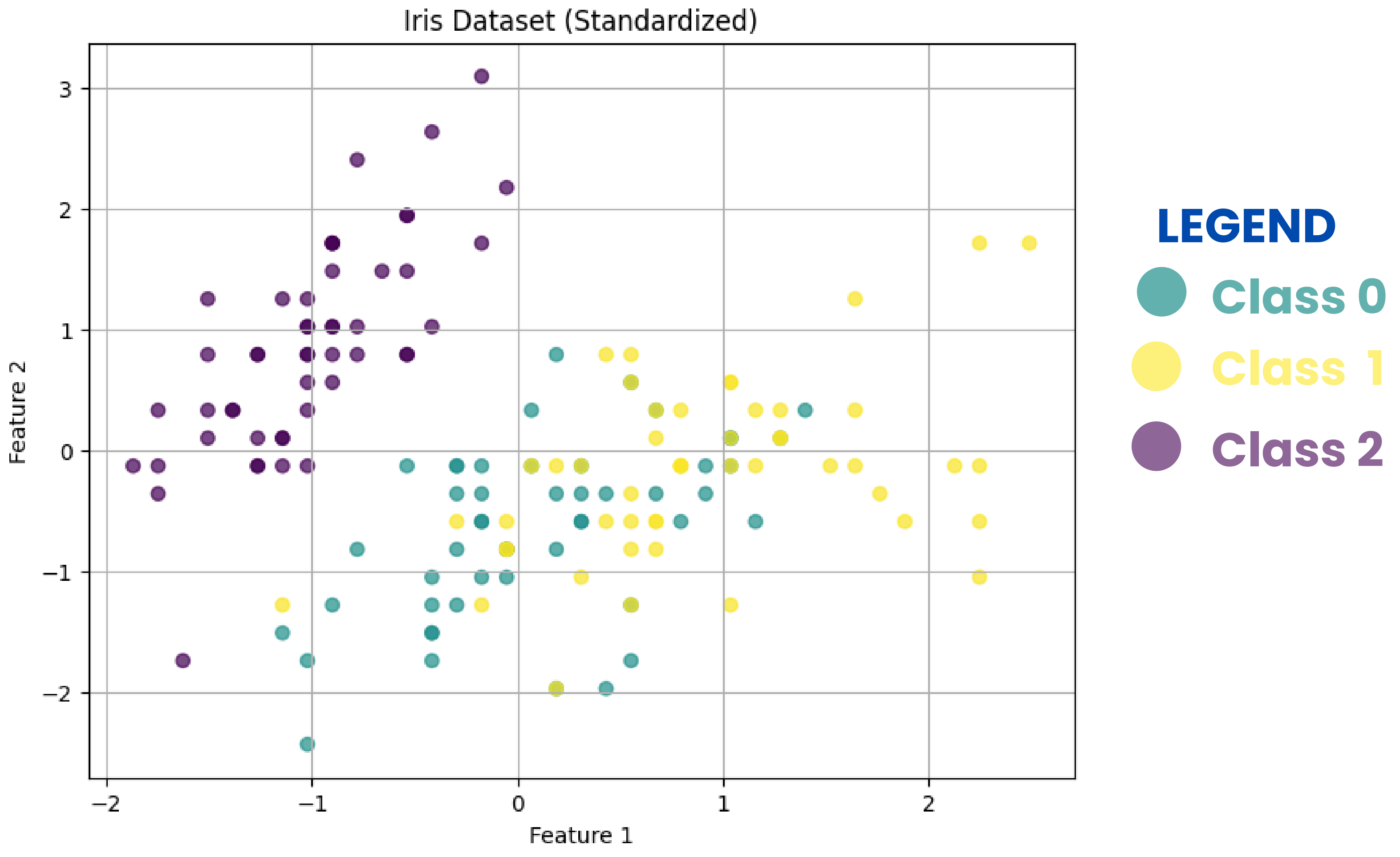


We can see that Standard EM gets stuck at a local maxima whereas the DAEM algorithm is able to overcome the local maxima and converge to global maxima (or higher local maxima).

IRIS DATASET

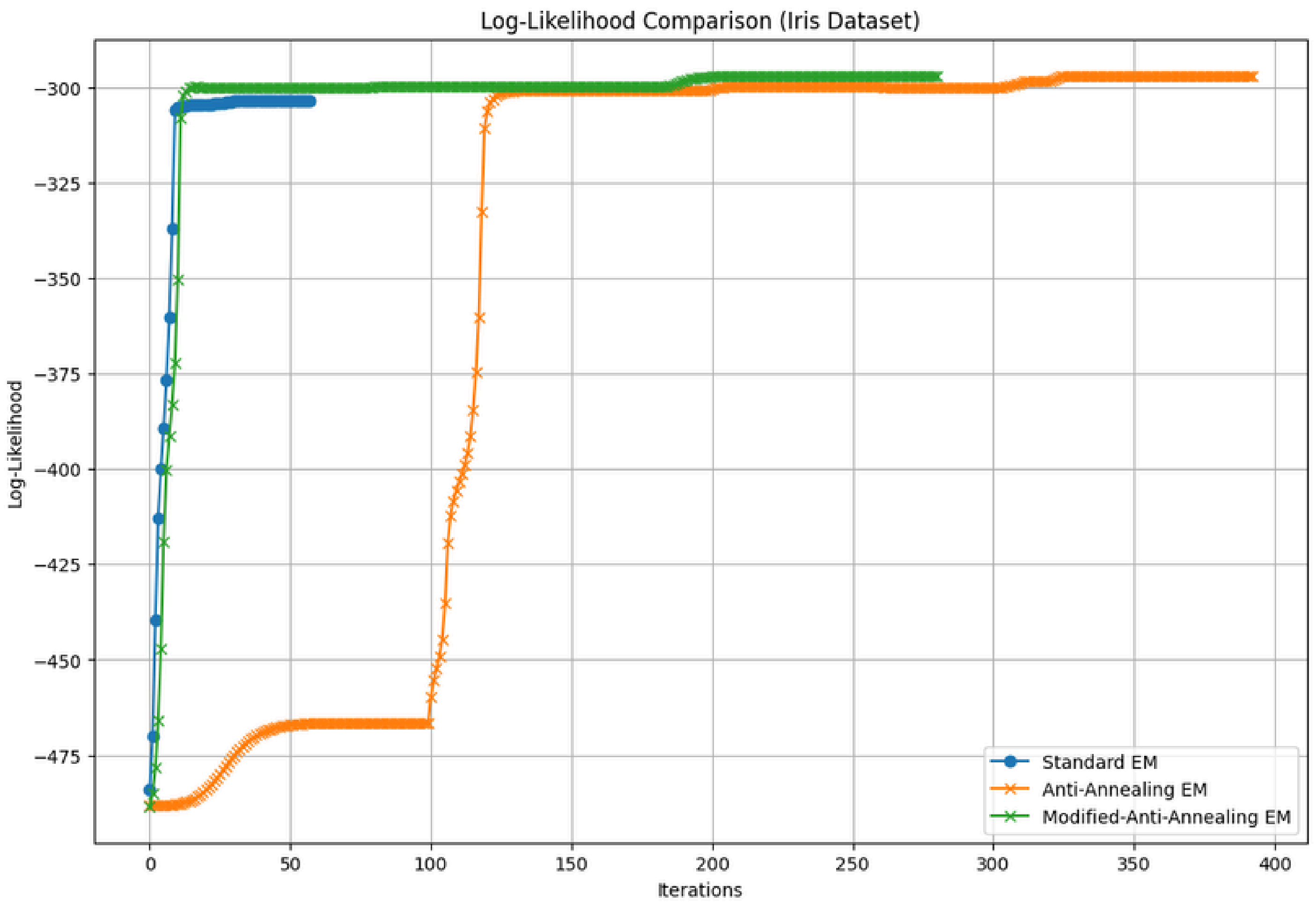
This dataset again has significant overlap. We want to compare EM vs a modified DAEM results for this dataset.

DATASET DESCRIPTION



The Iris dataset contains 150 samples from three species of Iris flowers, with 50 samples for each species. Each sample is characterized by four features: . These features are used for classification tasks, where the goal is to distinguish between the three species based on these measurements.

RESULTS



We see again that the standard EM algorithm gets stuck at a local maxima, which gets overcome by the DAEM algorithm. But since the convergence rate is slower, we implemented a modification of DAEM in which the scheduling parameter (β) is replaced with (β + α_j), where α_j is the weight of the j-th Gaussian cluster.

CONCLUSION

DATASET	STANDARD E-M		ANTI-ANNEALING E-M		MODIFIED ANTI-ANNEALING E-M	
	ARI	NMI	ARI	NMI	ARI	NMI
MULTIVARIATE GAUSSIAN DUSTRIBUTION	1.0000	1.0000	1.0000	1.0000	----	----
WINE	0.3259	0.3980	0.6511	0.7238	----	----
IRIS	0.5230	0.6154	0.7184	0.7629	0.7184	0.7629

REFERENCES

Iftekhar Naim, Daniel Gildea "Convergence of the EM Algorithm for Gaussian Mixtures with Unbalanced Mixing Coefficients"

Sarthak Chatterjee, Orlando Romero, and Sergio Pequito "Analysis of a Generalized Expectation-Maximization Algorithm for Gaussian Mixture Models: A Control Systems Perspective "

THANK YOU