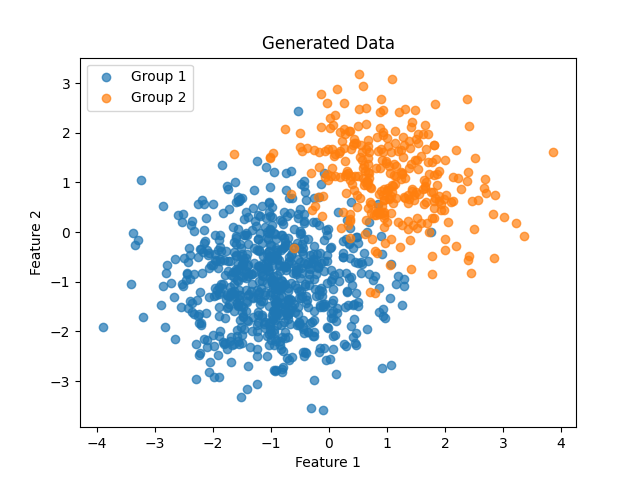
**Clustering Analysis using K-means and Gaussian Mixture Models (GMM)**

Yahya Saad 322944869 –Layla Mahameed 322799313

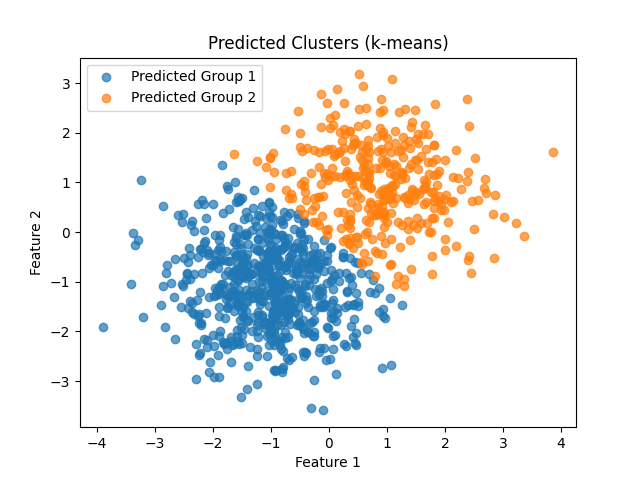
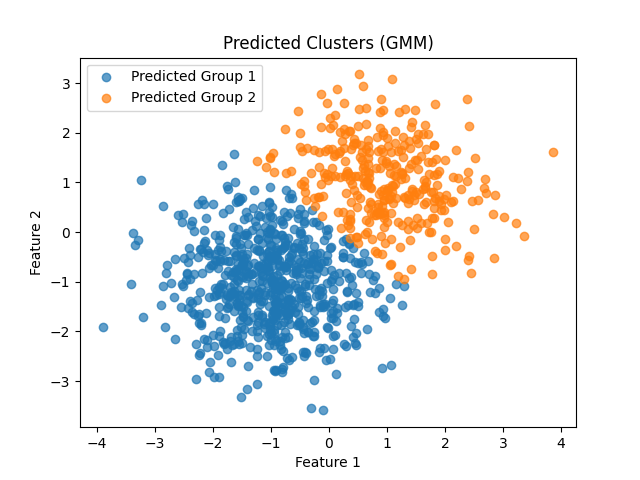
This code demonstrates clustering analysis using the K-means and Gaussian Mixture Models (GMM) algorithms on a generated dataset. The purpose of this analysis is to separate the data points into distinct groups based on their features.

The code begins by importing the necessary libraries, including numpy for numerical operations, matplotlib.pyplot for data visualization, sklearn.cluster.KMeans for the K-means algorithm, sklearn.mixture.GaussianMixture for the GMM algorithm, and matplotlib.cm for colormap functionality.

A synthetic dataset is created with two groups, each defined by its mean and covariance. Group 1 and Group 2 data are generated using multivariate normal distributions. The data from both groups are then combined into a single dataset.



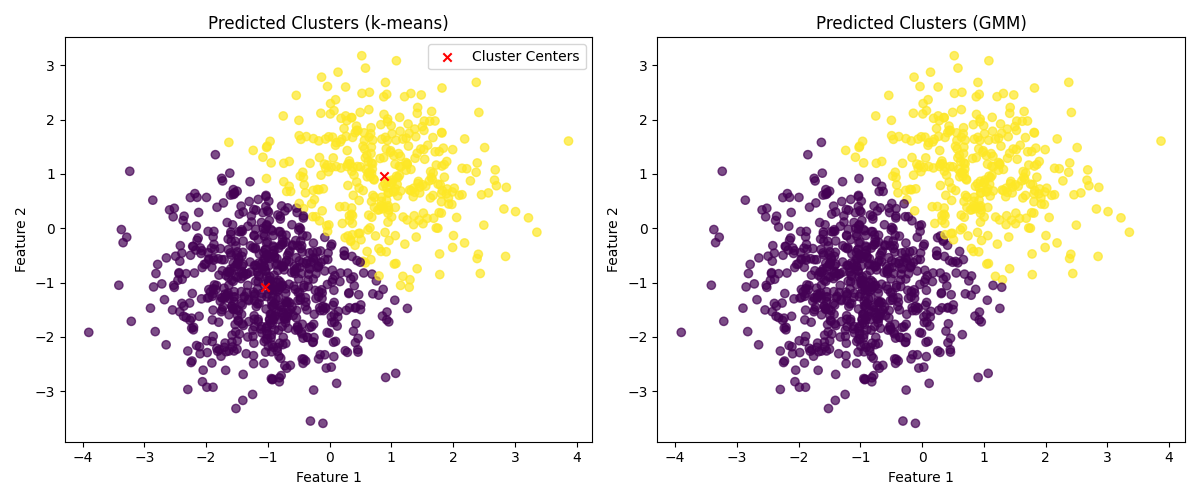
True labels are assigned to the data points to evaluate the performance of the clustering algorithms. These labels are created by concatenating arrays of zeros and ones, corresponding to the two groups.



The code proceeds to perform clustering using K-means and GMM. The K-means model is initialized with the desired number of clusters (k = 2), and the model is fitted to the data. The predicted labels for each data point are obtained. Similarly, the GMM model is initialized with the same number of components, and the model is fitted to the data to obtain the predicted labels.

To visualize the results, scatter plots are created to display the data points. True labels are represented by different markers and colors, while K-means and GMM predicted labels are also displayed with distinct markers and colors. Colormaps are used to differentiate between the labels.

The plot includes labels for the x and y axes, a title, and a legend indicating the true labels, K-means predicted labels, and GMM predicted labels.



Furthermore, the code identifies the misclassified data points by comparing the true labels with the predicted labels. These misclassified points are plotted separately, highlighting the differences between the true and predicted labels. Different markers and colors are used to distinguish misclassified points from correctly classified points.

A misclassification legend is added to the plot to indicate the meaning of the markers and colors.

In conclusion, this code provides an example of how to perform clustering analysis using K-means and GMM algorithms. It demonstrates how to generate data, fit the models, and visualize the results. The inclusion of misclassified points allows for a deeper understanding of the performance of the clustering algorithms. This code can be used as a starting point for further exploration and experimentation with clustering techniques.

