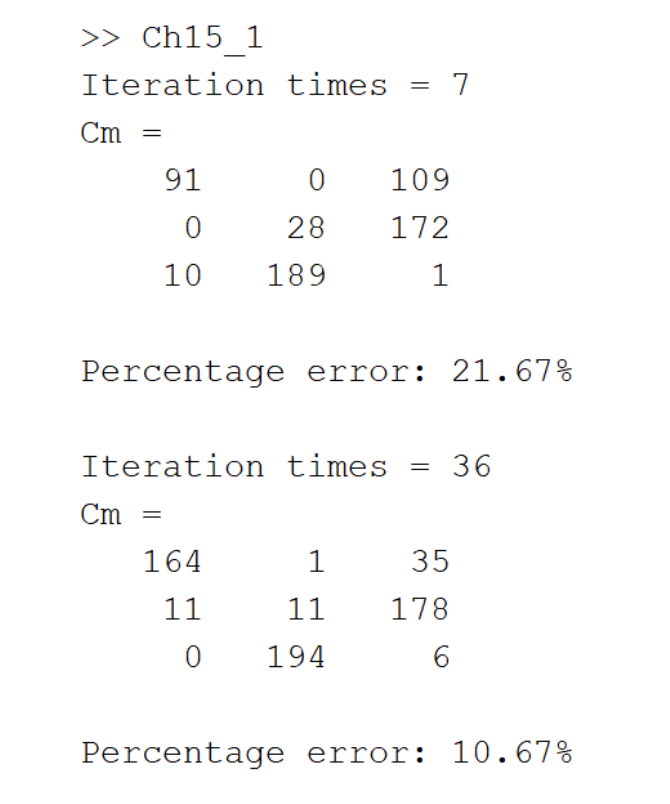
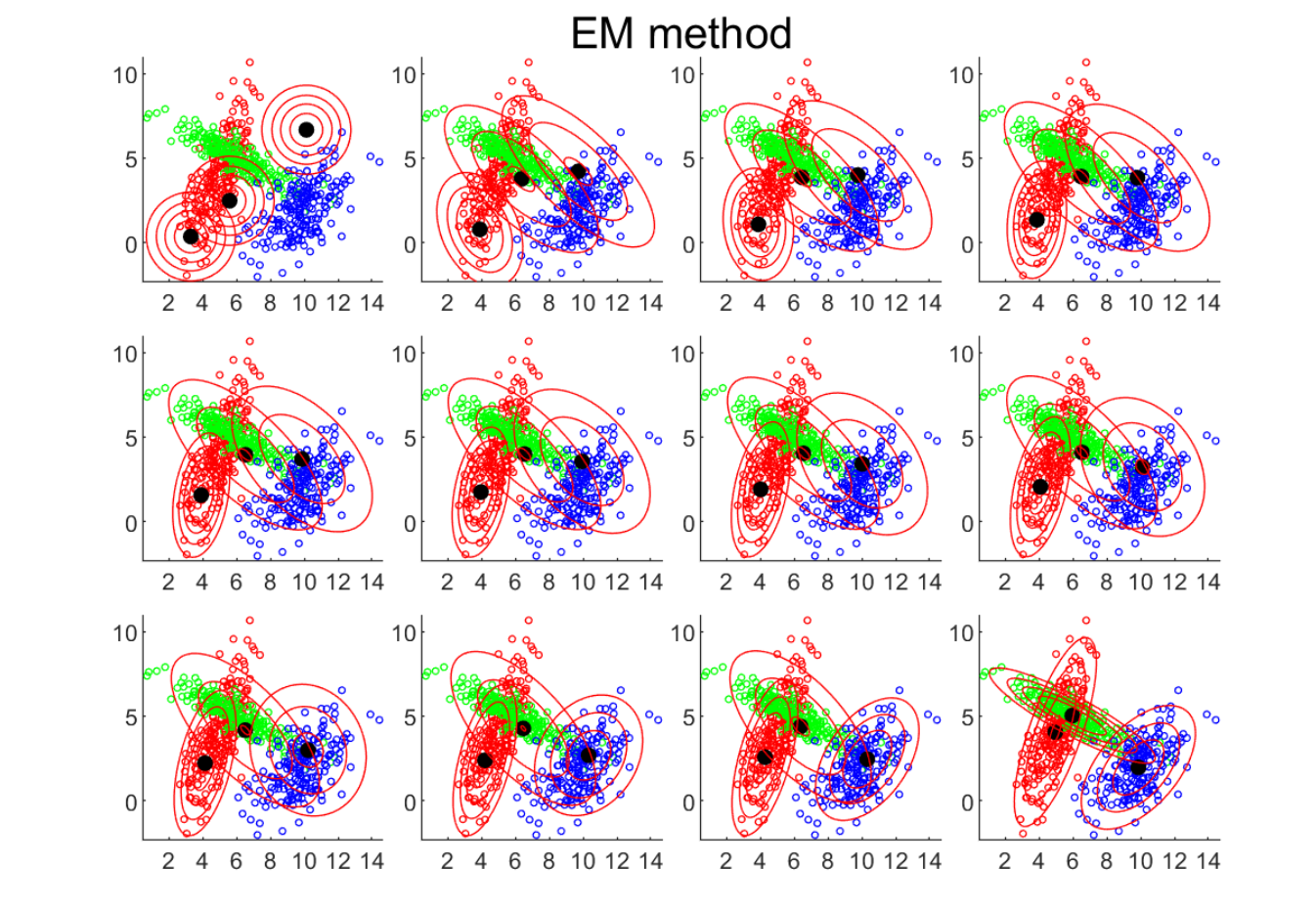
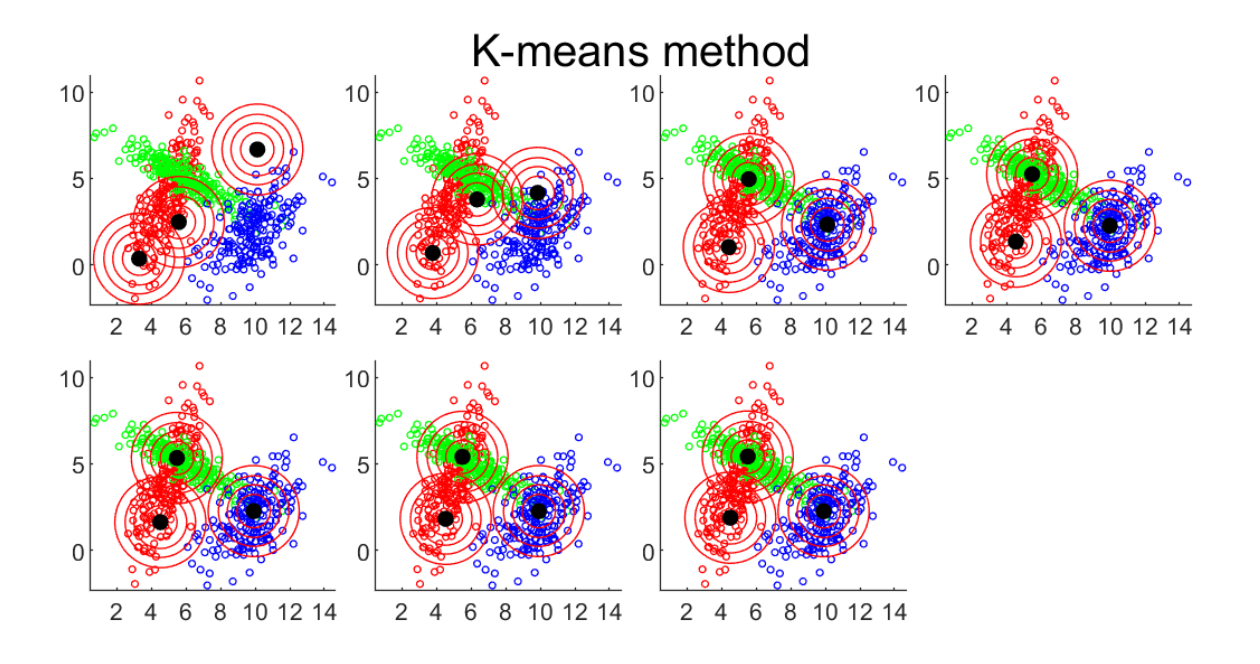
# Report —— Chapter 15

**Problem1**

This task is using GMM method and K means method to clustering data.

Output of the program:





By implementing and comparing K-means clustering and Gaussian Mixture Models on a synthetic dataset, we can conclude that GMM achieved superior performance in this case. The lower error rate confirms GMM's ability to more accurately separate nonlinearly separable clusters. Visualizations also demonstrate GMM produced compact, well-separated clusters capturing different covariances. In contrast, K-means struggles to correctly cluster non-globular distributions, leading to more overlap at boundaries.

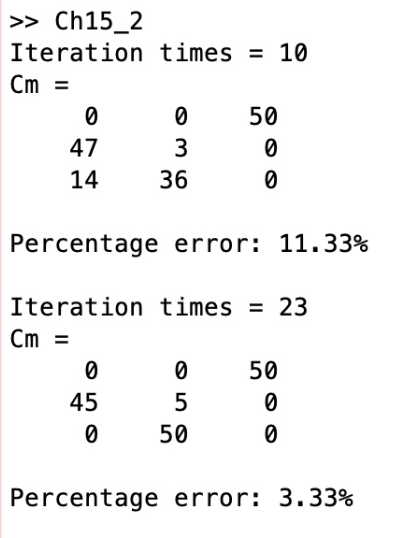
The visualization shows K-means and EM algorithms converged after 7 and 36 iterations respectively, indicating K-means converges faster than GMM. However, GMM achieved lower misclassification rate of 10.6% versus 21.67% for K-means.

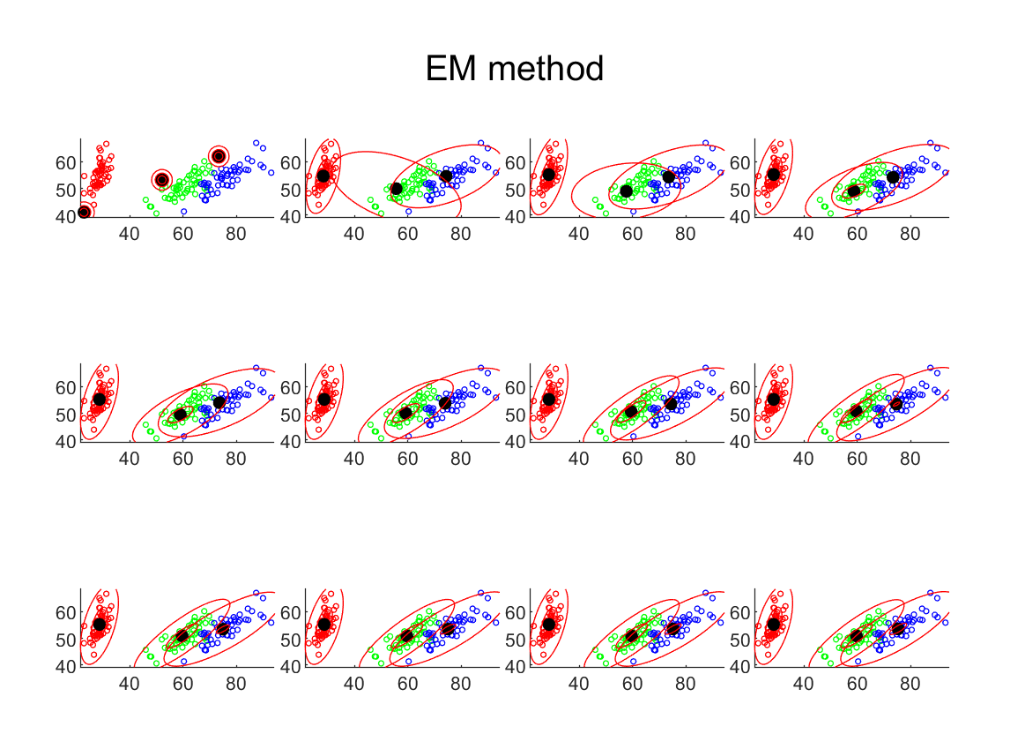
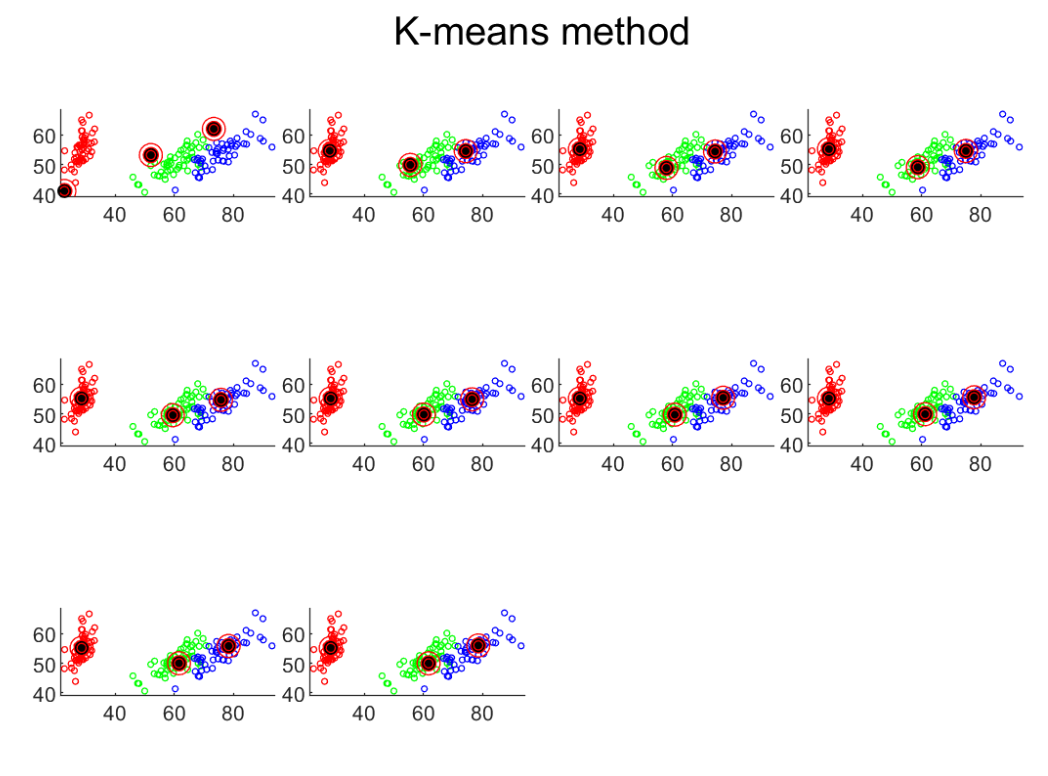
Overall, the flexibility of GMM's probabilistic model with Gaussian components enables it to effectively model complex clustering structures. By estimating means, covariances, and mixture coefficients, GMM can better fit multimodal distributions. On the other hand, K-means is limited by its spherical cluster assumption and sensitivity to initialization. Although simpler and faster, K-means performs poorly on non-convex cluster shapes and varying densities.

In summary, GMM outperforms K-means on this nonlinearly separable dataset, aside from convergence speed, due to its ability to model non-spherical clusters by fitting a probability model to the data. The results validate that GMM can capture complex clustering structures more accurately than the simpler K-means algorithm.

**Problem2**

Output:





Firstly, since we only change the dataset used, from the observations we are still able to arrive at the same conclusion made in the first part, which is that GMM achieved superior performance compared to K-means. This demonstrates that GMM's advantage is robus. Secondly, although the contour plots visualized have some flaws, by looking at the error rates reported in the command line output, we can very clearly see that GMM still has a lower error rate than K-means. This illustrates GMM's ability to adapt to different data distributions. Additionally, the flaws in the contour plots are likely due to ambiguous definition of the w parameter in this part. We tried different w values and the current contour plots are the best visualizations we could obtain. In summary, GMM's continued strong performance on the second dataset reaffirms its ability to model complex clustering shapes.