Project Report

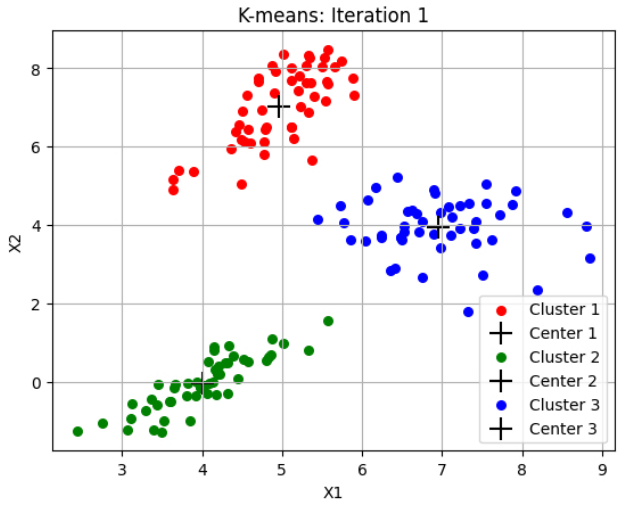
Psarros Filippos

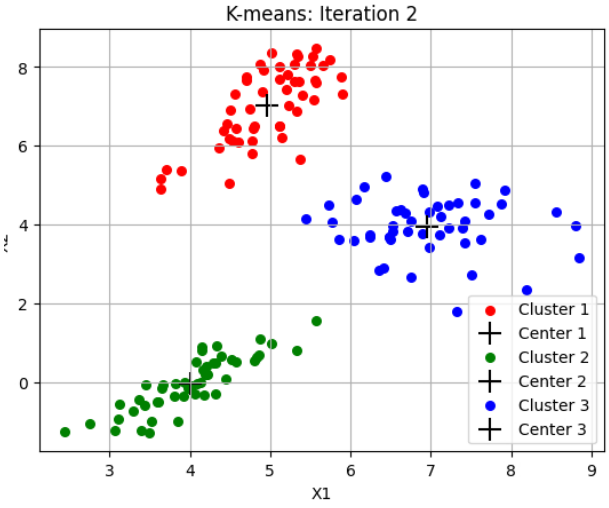
A full analysis of the project is provided in the K_Means_Clustering_Implementation_and_Visualization_from_Scratch_report.pdf file.

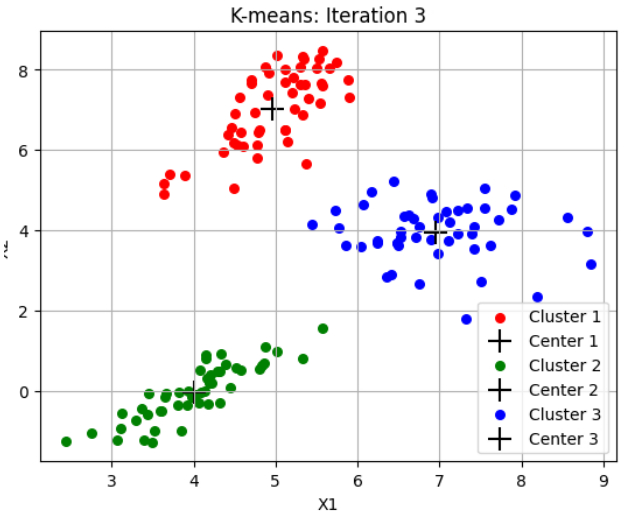
Here are the results generated by the code below.

```
Iteration 1: SSE = 167.77165
Iteration 2: SSE = 158.68924
Iteration 3: SSE = 158.68924
Convergence reached at iteration 3
```

We observe that in the 3rd iteration, the SSE remains stable, indicating that convergence has been achieved. The significant decrease in SSE during the initial iterations shows that the clusters improve rapidly at the beginning.
Below are the plots of the K-Means algorithm.

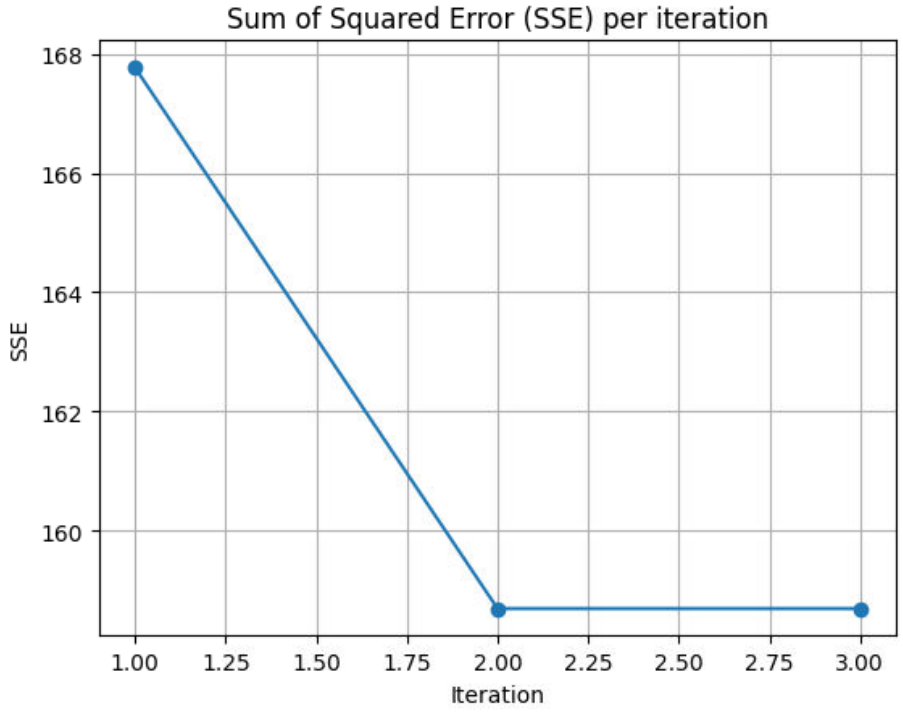






We observe that in the 1st iteration, the algorithm performed the initial classification based on the minimum distance of the points from the initial cluster centers, and the Sum of Squared Error (SSE) at this iteration is 167.77165, indicating a large deviation between the points and the centers.
In the 2nd iteration, the point classification has significantly improved, and the SSE has decreased to 158.68924, showing a notable improvement in overall accuracy.
In the 3rd iteration, all points have been correctly classified into their respective clusters. The movement of the centers in this iteration is zero, which signals the convergence of the algorithm, while the SSE remains constant at 158.68924, confirming that the algorithm has converged.

Below is the plot of the Sum of Squared Error.



The graph shows the decrease in the Sum of Squared Error (SSE) with each iteration. A sharp drop in SSE is observed between the 1st and 2nd iteration. This significant drop in the first iteration indicates that the initial centers were quite far from their optimal positions. The SSE remains stable from the 3rd iteration onward, confirming the convergence of the algorithm.

Conclusions:

1. The K-Means algorithm successfully classified the data into three clusters, as shown by the final state in the plot from iteration 3.
2. The SSE plot demonstrates that convergence was achieved in just 3 iterations, indicating fast and efficient operation.
3. The initial classification in Iteration 1 was less accurate due to the large distance between the initial centers and the data points, but the algorithm corrected these errors in the following iterations.

The final cluster centers are:

```
Final cluster centers:
[[ 4.96719651  7.01490467]
 [ 4.00922758 -0.02733973]
 [ 6.95717588  3.94556642]]
```