Algorithm Implement

Traditional K-means optimizes SSE (Sum of Squared Errors), while OKM aims to minimize the "degree of overlap between clusters." The specific approach is as follows:

1. Define a method to measure the degree of overlap between clusters, such as minimum distance, overlapping regions, or distance ratios.
2. During the iterative process, strive to make points within the same cluster as close as possible (intra-cluster compactness) while keeping different clusters as separate as possible (inter-cluster separation).
3. In the paper, the authors propose a metric called Overlap and integrate it into the K-means iteration process.

Key Features of the Implementation

1. Two Variants:
    - Standard Overlap K-means (OK) using centroids (Algorithm 1 in the paper)
    - Localized variant (OK-local) using K-nearest neighbors (Algorithm 2 in the paper)
2. Overlap Calculation:
    - For OK: Uses distance to own centroid divided by distance to nearest neighbor in different cluster
    - For OK-local: Uses mean shift distance (average distance to KNN in same cluster) divided by distance to nearest neighbor in different cluster
3. Weighted Updates:
    - Points are weighted inversely proportional to their overlap values when updating centroids (for OK) or labels (for OK-local)
    - Uses exponential weighting with bandwidth parameter γ
4. Efficiency Considerations:
    - Uses sklearn's NearestNeighbors for efficient KNN calculations
    - Uses pairwise_distances for distance computations
5. Stopping Criterion:
    - Stops when labels stop changing or max iterations reached

Cluster Result: