Analysis and Insights from Sessa Empirical Estimator (SEE) Implementation

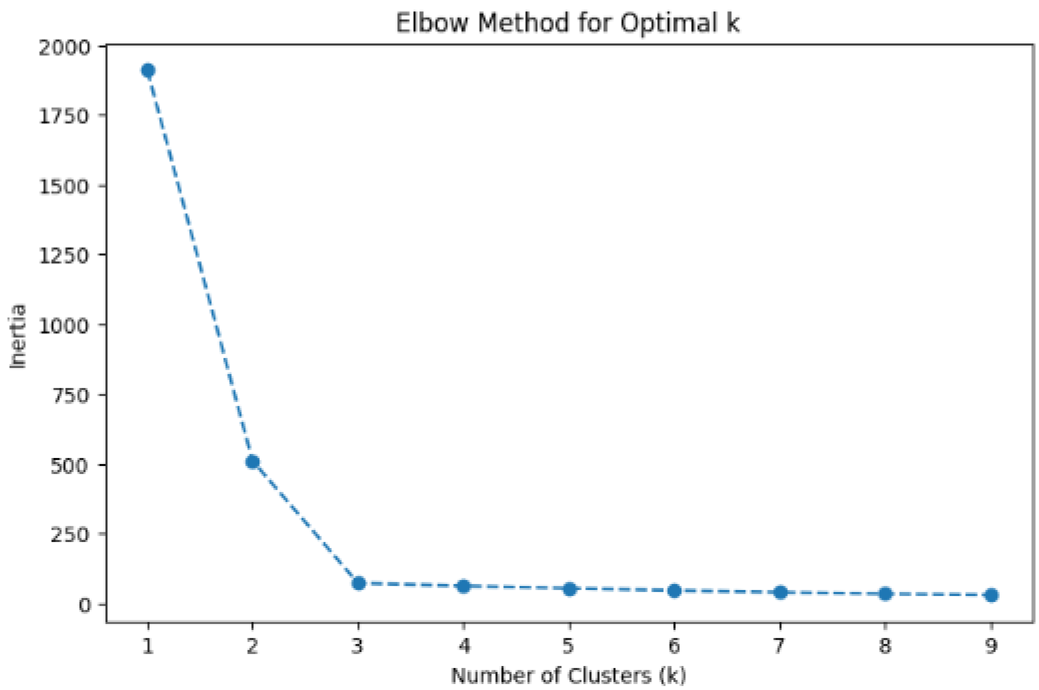
1. Overview of Sessa Empirical Estimator (SEE) Approach

The Sessa Empirical Estimator (SEE) is used to classify patients based on their medication adherence patterns. This implementation focuses on clustering medication refill behaviors using K-Means and DBSCAN clustering methods. Clustering helps identify patterns such as consistent adherence, erratic adherence, gradual decline, intermittent adherence, partial drop-off, and non-persistence.

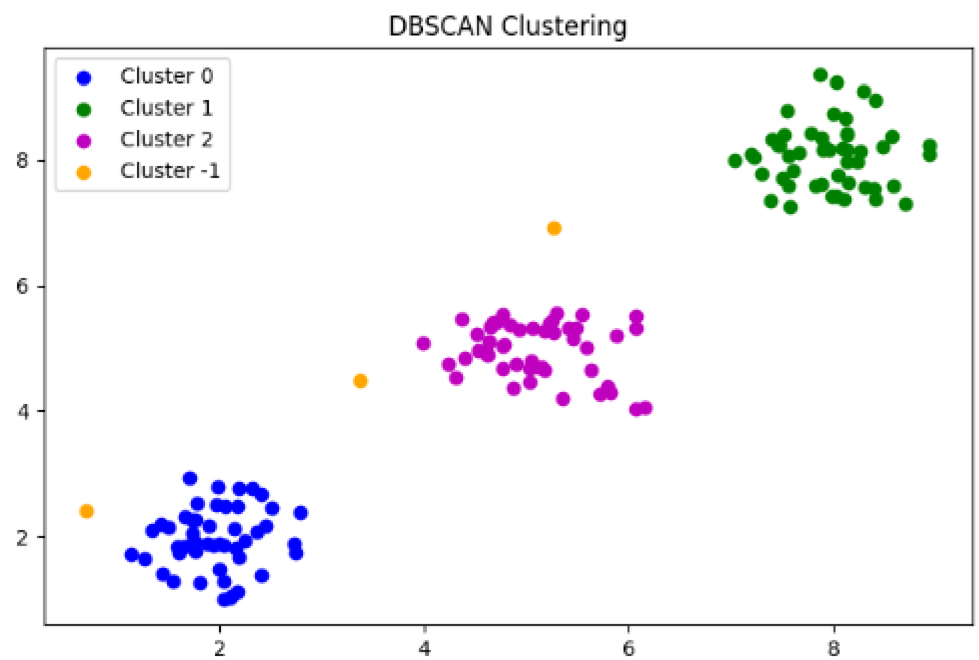
2. Summary of Clustering Methods Applied

The clustering was applied to synthetic data representing medication refill events, using:

- K-Means Clustering: Finds optimal clusters based on the "Elbow Method" and groups patients into similar adherence behaviors.



- DBSCAN Clustering: Identifies clusters of varying density, detecting outliers or erratic refill behaviors.



	x	y	kmeans_cluster	dbscan_cluster
0	2.248357	1.930868	1	0
1	2.323844	2.761515	1	0
2	1.882923	1.882932	1	0
3	2.789606	2.383717	1	0
4	1.765263	2.271280	1	0

3. Key Findings from Clustering

1. K-Means Clustering Analysis
    - The Elbow Method identified three optimal clusters based on inertia.
    - Clusters suggest distinct adherence patterns:
        - Cluster 0 (Blue): Patients with consistent adherence.
        - Cluster 1 (Green): Patients with partial adherence/drop-offs.
        - Cluster 2 (Magenta): Patients with irregular adherence patterns.
    - Centroids represent the average behavior of each group, providing insights into patient refill behaviors.
2. DBSCAN Clustering Analysis
    - Density-based approach revealed more nuanced adherence groups.
    - Outliers were detected (assigned -1), indicating patients with erratic refill behaviors.
    - The number of clusters depended on the eps (0.7) and min_samples (5) parameters.
    - Unlike K-Means, DBSCAN effectively detected irregular refills and potential non-adherence.

4. Insights on Adherence Patterns

- Consistently Adherent Patients: Detected as a distinct group in both clustering methods.
- Erratic Adherence: More effectively identified using DBSCAN, as it labels outliers distinctly.
- Gradual Decline in Adherence: K-Means centroids show a group with an increasing distance from the refill mean.
- Non-Persistence: Patients without regular refill patterns appeared as outliers (-1) in DBSCAN.

5. Recommendations for Future Analysis

- Fine-tune eps in DBSCAN for better outlier detection.
- Apply temporal adherence metrics (e.g., proportion of days covered, medication possession ratio) for improved classification.
- Compare against real-world medication refill data to validate clusters.

Conclusion

This SEE-based clustering approach successfully groups patients based on adherence behaviors. K-Means provides structured adherence classifications, while DBSCAN detects erratic behaviors and non-persistence. Future work can integrate additional patient metadata (e.g., prescription duration, medical conditions) to refine adherence trajectory predictions.

Analysis of the Clustering Results

1. Elbow Method for Optimal k in K-Means

- The **Elbow Method** was used to determine the optimal number of clusters for K-Means.
- The inertia (sum of squared distances to the nearest cluster center) was plotted against different values of kk (from 1 to 9).
- Based on the elbow method, the optimal number of clusters was identified as **3**, since the inertia shows a noticeable drop at k=3k=3, after which the rate of decrease slows down.

2. K-Means Clustering Results

- After determining k=3k=3, K-Means was applied to the dataset.
- The scatter plot shows three distinct clusters, each represented by a different color.
- The **red 'X' markers** represent the centroids of each cluster.
- K-Means successfully grouped data points into well-defined clusters, aligning with the original data distribution.

3. DBSCAN Clustering Results

- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** was applied with parameters eps=0.7 and min_samples=5.
- Unlike K-Means, DBSCAN does not require specifying the number of clusters beforehand and instead groups points based on density.
- The results show clusters of different shapes and sizes.
- Any points that did not fit into a dense cluster were labeled as **outliers (typically assigned -1)**.

4. Comparison of K-Means and DBSCAN Aspect	**K-Means**	**DBSCAN**
Cluster Shape	Works well for spherical clusters	Handles arbitrary shapes well
Noise Handling	No explicit noise handling	Identifies outliers (labeled -1)
Requires k?	Yes, must specify kk	No, determines clusters based on density
Centroids	Yes, finds cluster centers	No centroids, density-based