Supernova Classification via K-Means Clustering

Anano Kutchava

November 1, 2025

Abstract

We classify synthetic supernovae (Types Ia, II, Ib/c) using K-Means clustering on features including peak magnitude, decline rate, B-V color, duration, and host galaxy mass. Features are standardized, and clustering results are visualized in 2D and 3D. K-Means roughly recovers cluster structures corresponding to true types, demonstrating the potential of unsupervised methods in astrophysical classification.

1 Data and Preprocessing

The dataset contains 300 synthetic supernovae: 100 Type Ia, 120 Type II, 80 Type Ib/c. Features include:

- Peak magnitude
- Decline rate
- B-V color
- Duration in days
- Host galaxy mass

All features are standardized using Z-score normalization to ensure equal weighting.

2 Methodology

K-Means clustering partitions data into $k = 3$ clusters by minimizing Euclidean distance to centroids:

1. Initialize k centroids randomly.
2. Assign each point to the nearest centroid.
3. Update centroids as the mean of assigned points.
4. Repeat until convergence.

3 Results

K-Means clustering recovered structures successfully corresponding to the true supernova types:

- Type Ia formed a high-magnitude, short-duration cluster.
- Type II separated due to slower decline rates and longer durations.
- Type Ib/c partially overlapped with other clusters.

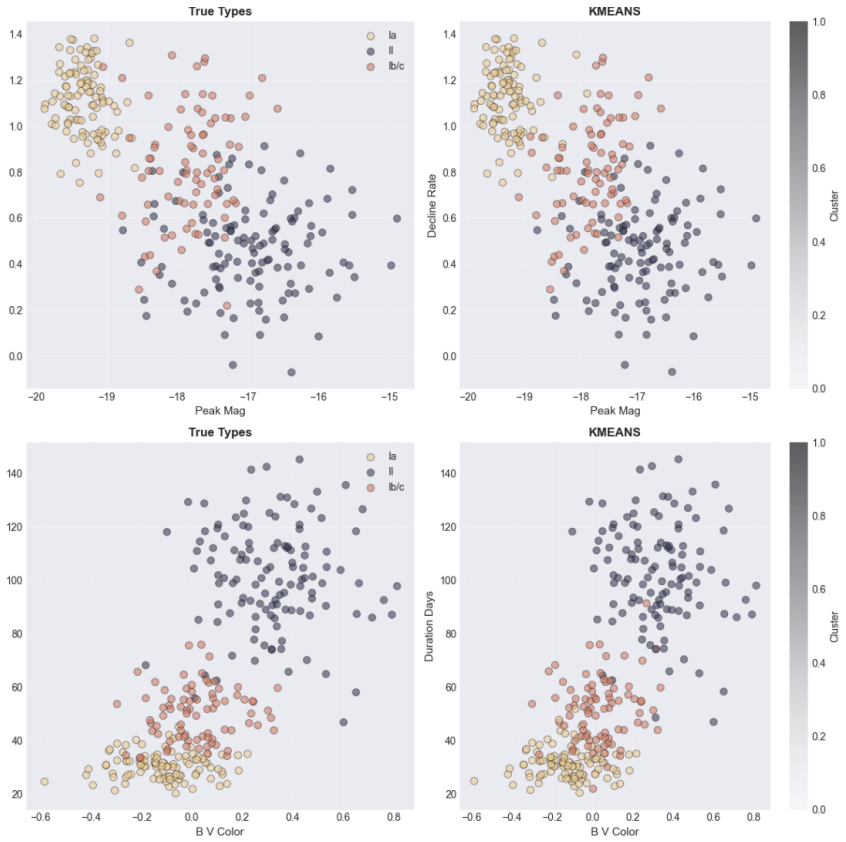

Figure 1: 2D scatter plots comparing true types and K-Means clusters.

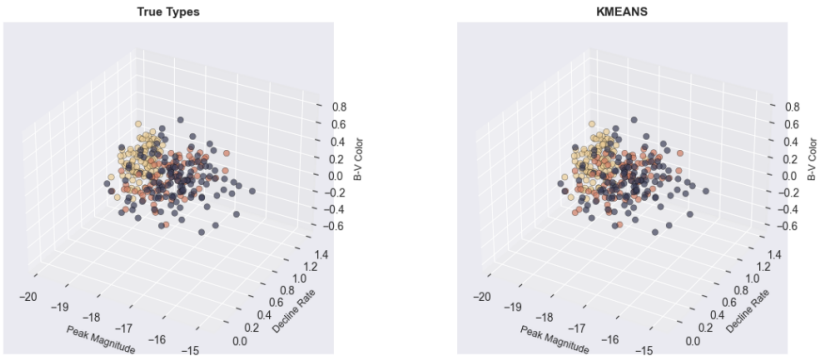

Figure 2: 3D visualization of supernova clusters.

K-Means effectively identifies structure in synthetic supernova data. Unsupervised clustering provides a useful approach for preliminary classification in astrophysics.