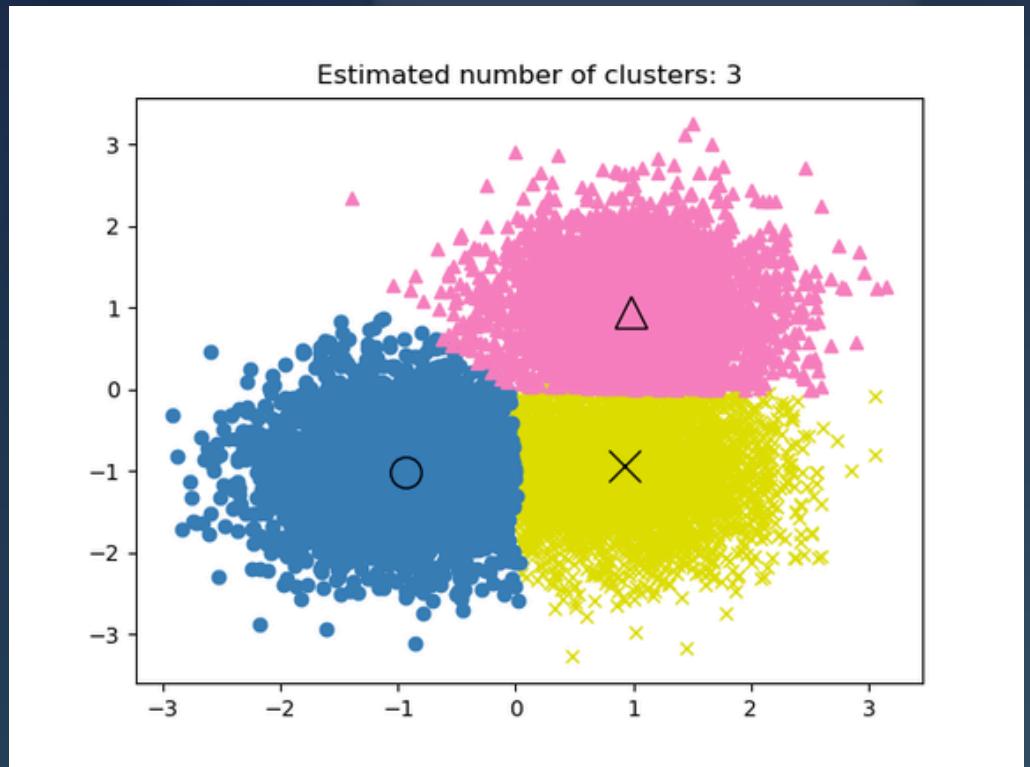


Mean Shift

PAGE 01



Automatic cluster number, but very slow and sensitive to bandwidth. Cluster forms based on the highest density of the data points.

Advantages

PAGE 03

- No Need to Specify Number of Clusters
 - Automatically finds number of clusters
- Finds Arbitrary Shape Clusters
 - Not limited to spherical clusters like K-Means
- Robust to Outliers (to some extent)
 - Dense regions attract points, sparse noise is ignored
- Simple Concept
 - Based on moving points toward highest data density (mode seeking)
- Works Well for Image Segmentation
 - Popular in computer vision tasks

Disadvantages

- Very Slow for Large Data
 - Time complexity is high ($O(n^2)$)
- Memory Intensive
 - Stores all data points and neighbors
- Bandwidth Selection is Critical
 - Wrong bandwidth → too many or too few clusters
- Not Good for High Dimensions
 - Curse of dimensionality affects density estimation
- No Probabilistic Output
 - Only gives cluster labels