Mean Shift Clustering

MACHINE LEARNING CLUSTERING

Mean Shift Clustering

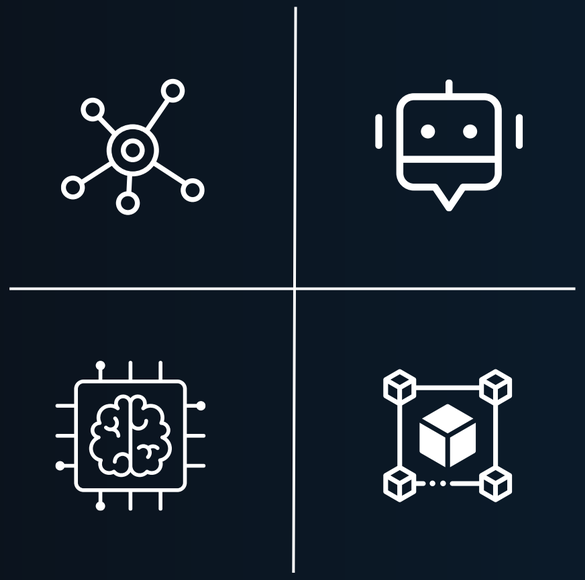
Introduction to Mean Shift Clustering

A non-parametric clustering technique that does not require prior knowledge of the number of clusters. Finds dense regions in the feature space by iteratively shifting data points towards the mode (peak) of the data distribution.

Key Concepts:

- **Bandwidth (h):** Controls the size of the window used to compute the mean shift.
- **Kernel Function:** Determines the weight of nearby points for the mean computation.
- **Convergence:** Points are shifted iteratively until they converge to the mode.

Mechanics of Mean Shift Clustering



1. **Initialization:**

   Select initial points (seeds) for the mean shift procedure.

2. **Mean Shift Vector Calculation**

3. **Update Points**

   Move each point towards the mean (mode) using the mean shift vector.

4. **Iteration:**

   Repeat the mean shift calculation and point updates until convergence (points no longer move significantly).

Application and Evaluation

Application:

- Image segmentation, object tracking, clustering in feature space.
- Steps:
  - ✓ Compute the mean shift vectors for all data points.
  - ✓ Shift data points iteratively until convergence.
  - ✓ Group points converging to the same mode into clusters.

Advantages:

- No need to specify the number of clusters.
- Can identify arbitrarily shaped clusters.
- Robust to noise and outliers.

Disadvantages:

- Computationally expensive for large datasets.
- Choice of bandwidth (h) is critical and non-trivial.
- May converge to local maxima, resulting in suboptimal clustering.

Thank You