Machine Learning 101

Rajdeep Chatterjee, Ph.D.
Amygdala AI, Bhubaneswar, India *

March 2025

Mean Shift Clustering

Mean Shift Clustering

Mean Shift Clustering is an unsupervised machine learning algorithm used for clustering data points by finding dense areas in the feature space. It is a non-parametric algorithm that does not require prior knowledge of the number of clusters.

How Mean Shift Works

- **Initialize Points:** Start with each data point as a cluster center.
- **Kernel Density Estimation (KDE):** Use a kernel function (e.g., Gaussian) to estimate the density of data points in the feature space.
- **Mean Shift Step:** For each point, compute the weighted mean of points within a given bandwidth and shift the point to this mean.
- **Convergence:** Repeat the mean shift step until points no longer move significantly or a maximum number of iterations is reached.
- **Cluster Formation:** Points that converge to the same position are grouped into the same cluster.

Mathematical Formulation

The mean shift vector is calculated as:

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)} \tag{1}$$

Where:

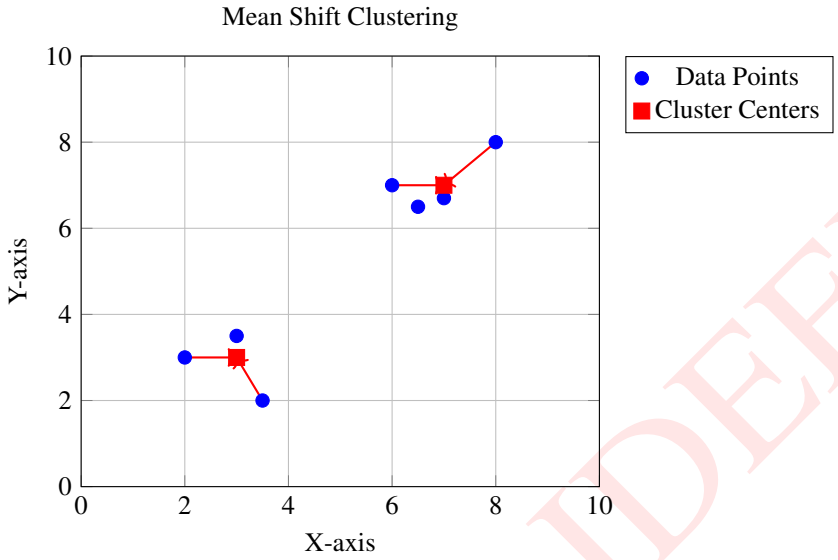
- $m(x)$ = mean shift vector

*Amygdala AI, is an international volunteer-run research group that advocates for *AI for a better tomorrow* `http://amygdalaai.org/`.

- K = kernel function (often Gaussian)
- $N(x)$ = neighborhood points within the bandwidth

Visualization

Below is a visual representation of Mean Shift clustering applied to a sample dataset.



Mean Shift is a **non-parametric**, **unsupervised** clustering algorithm that finds dense areas in a feature space. Unlike k-means, it does not require the number of clusters to be pre-specified.

1 Algorithm Steps

1. **Initialize Points:** Each data point is treated as a cluster center.
2. **Define a Kernel Density Estimator (KDE) Window:**
   - A circular window of radius h (bandwidth) is placed around each point.
   - The window size determines the search area for nearby points.
3. **Compute Mean Shift Vector:** The mean of all points within the window is computed:

$$m(x) = \frac{\sum_i K(x_i - x)x_i}{\sum_i K(x_i - x)} \tag{2}$$

   where $K(x)$ is a kernel function (e.g., Gaussian, Flat, or Epanechnikov).
4. **Shift Window:** Move the window to the computed mean.
5. **Repeat Until Convergence:** Continue shifting until movement is negligible.
6. **Assign Clusters:** All points converging to the same mode belong to the same cluster.

2 Characteristics of Mean Shift

- No need to specify the number of clusters in advance.
- Detects arbitrarily shaped clusters.
- Computationally expensive for large datasets.
- The choice of bandwidth (h) significantly affects clustering results.

3 Applications

- Image Segmentation
- Object Tracking in Computer Vision
- Anomaly Detection
- Density Estimation

4 Mean Shift Algorithm in Pseudocode

Algorithm 1 Mean Shift Clustering

```
Initialize each data point as a cluster center.
while not converged do
    for each data point x do
        Find all points within bandwidth h.
        Compute mean of these points.
        Shift x toward the mean.
    end for
end while
Assign clusters based on final convergence points.
```

5 Merits and Demerits

5.1 Merits

- Does not require prior knowledge of the number of clusters.
- Can detect arbitrarily shaped clusters.
- Works well in high-dimensional spaces.
- Robust to outliers and noise in the data.

5.2 Demerits

- Computationally expensive, especially for large datasets.
- The choice of bandwidth (h) is crucial and affects performance.
- May converge to local optima, depending on initialization.
- Difficult to scale to very large datasets due to repeated density estimation.