**Mean Shift Clustering: A Non-Parametric Clustering Technique for Unsupervised Learning**



**Mean shift clustering** is a popular **unsupervised machine learning technique** for clustering data points. It is a **non-parametric method**, which means it does not assume any particular distribution of data. Instead, it uses a kernel density estimation to find the clusters. The main idea behind mean shift clustering is to find the modes of a kernel density function that is based on the data points. The modes are the points where the density function is at its maximum.

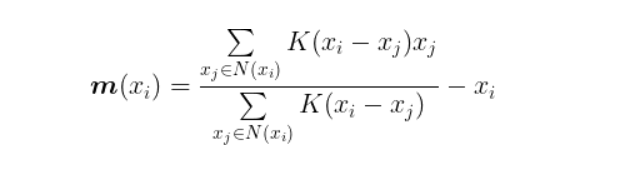
# What is Mean Shift Clustering?

Mean shift clustering is a density-based clustering algorithm that identifies the modes of a density function, which represent the clusters. In other words, it finds the areas of the dataset where the probability density function is the highest and clusters the data points in those areas together.

# How does Mean Shift Clustering work?

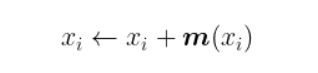
The algorithm starts by initializing a window or kernel around each data point. The kernel can be any type of function that decreases in value as the distance from the center of the kernel increases. The most common kernel function used in mean shift clustering is the Gaussian kernel.

The algorithm then computes the mean shift vector for each data point. The mean shift vector represents the direction in which the density function is increasing the most, and its magnitude represents the rate of increase. The mean shift vector is computed as follows:



where **m**(xᵢ) is the mean shift vector for data point xᵢ, K is the kernel function, and N(xᵢ) is the set of data points within the window or kernel centered at xᵢ.

The mean shift algorithm then updates the position of each data point by shifting it in the direction of the mean shift vector:



The algorithm repeats the mean shift vector and position updates until convergence, which is typically defined as a small change in the position of the data points between iterations.

Once the algorithm converges, the clusters are defined as the set of data points that converge to the same mode of the density function.