Mean Shift Calibri

Mean-shift is falling under the category of a clustering algorithm in contrast of Unsupervised learning that assigns the data points to the clusters iteratively by **shifting points towards the mode** (mode is the **highest density of data points** in the region, in the context of the Mean-shift). As such, it is also known as the **Mode-seeking algorithm**. Mean-shift algorithm has applications in the field of image processing and computer vision.

Given a set of data points, the algorithm iteratively assigns each data point towards the closest cluster centroid and direction to the closest cluster centroid is determined by where most of the points nearby are at. So, each iteration each data point will move closer to where the most points are at, which is or will lead to the cluster center. When the algorithm stops, each point is assigned to a cluster.

Some step:

Step-1: First, start with the data points assigned to a cluster of their own.

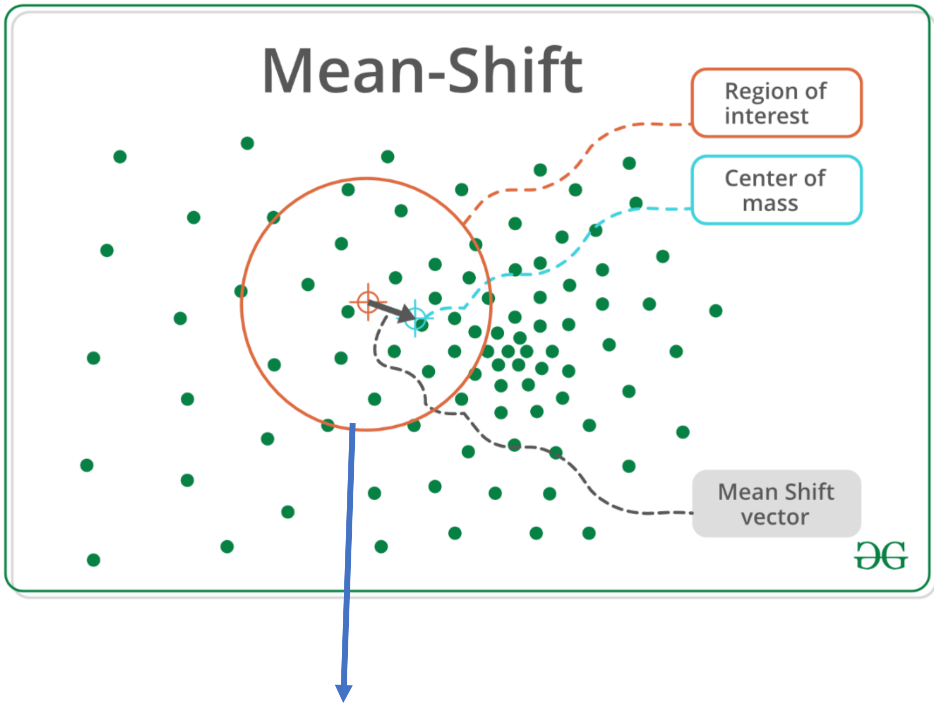
Step-2: Next, this algorithm will computer the centroids.

Step-3: In this step, location of new centroids will be updated.

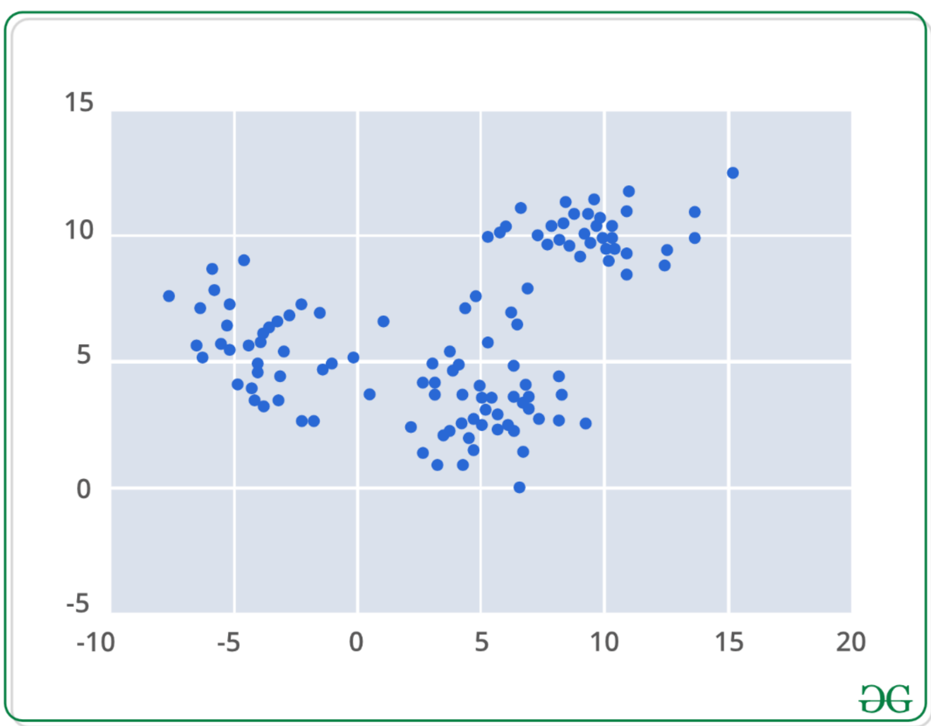
Step-4: Now, the process will be iterated and moved to the higher density region.

Step-5: At last, it will be stopped once the centroids reach at position from.

Note: mean-shift does not require specifying the number of clusters in advance. The number of clusters is determined by the algorithm with respect to the data.

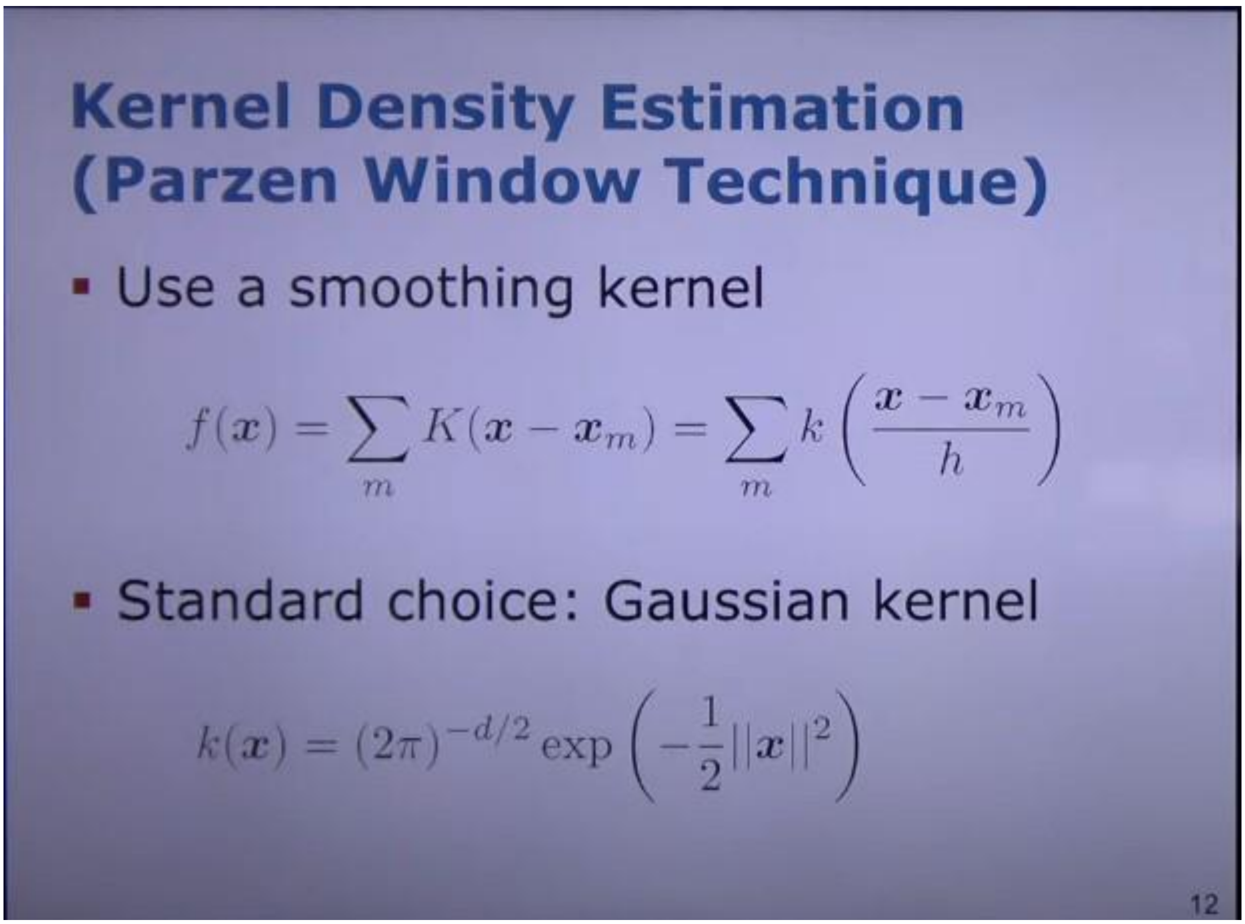

Mean-Shift
Region of interest
Center of mass
Mean Shift vector

Bandwidth


15
10
5
0
-5
-10
-5
0
5
10
15
20

Mean-shift builds upon the concept of **kernel density estimation is sort KDE**. Imagine that the above data was sampled from a probability distribution. KDE is a method to estimate the underlying distribution also called the probability density function for a set of data.

It works by placing a kernel on each point in the data set. A kernel is a fancy mathematical word for a weighting function generally used in convolution. There are many different types of kernels, but the most popular one is the Gaussian kernel. Adding up all of the individual kernels generates a probability surface example density function. Depending on the kernel bandwidth parameter used, the resultant density function will vary.



Computing the Mean Shift

Simple mean shift procedure:

- Compute mean shift vector
- Translate the Kernel window

$$\hat{\boldsymbol{x}} = \boxed{\frac{\sum_m \boldsymbol{x}_m \exp\left(-\frac{1}{2}\left|\left|\frac{\boldsymbol{x}-\boldsymbol{x}_m}{h}\right|\right|^2\right)}{\sum_m \exp\left(-\frac{1}{2}\left|\left|\frac{\boldsymbol{x}-\boldsymbol{x}_m}{h}\right|\right|^2\right)}} - \boxed{\boldsymbol{x}}$$

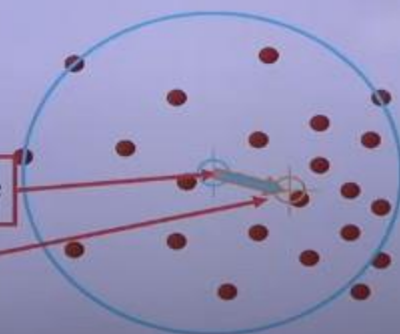
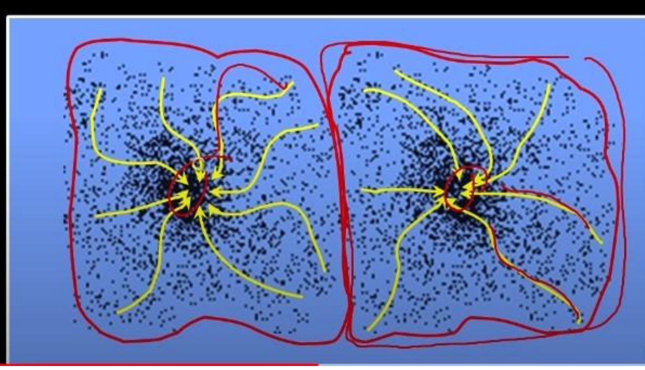
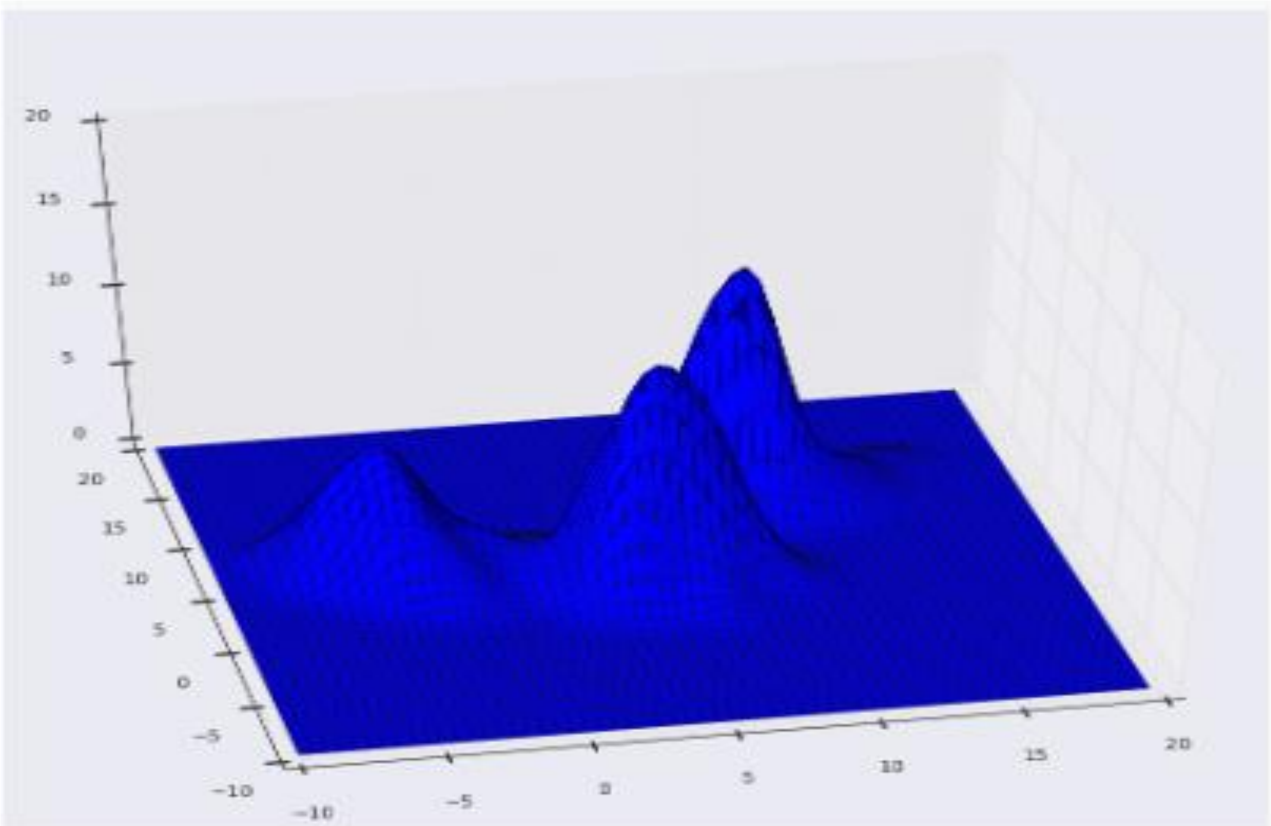


Image courtesy: Ulkrainitz & Sarel 20

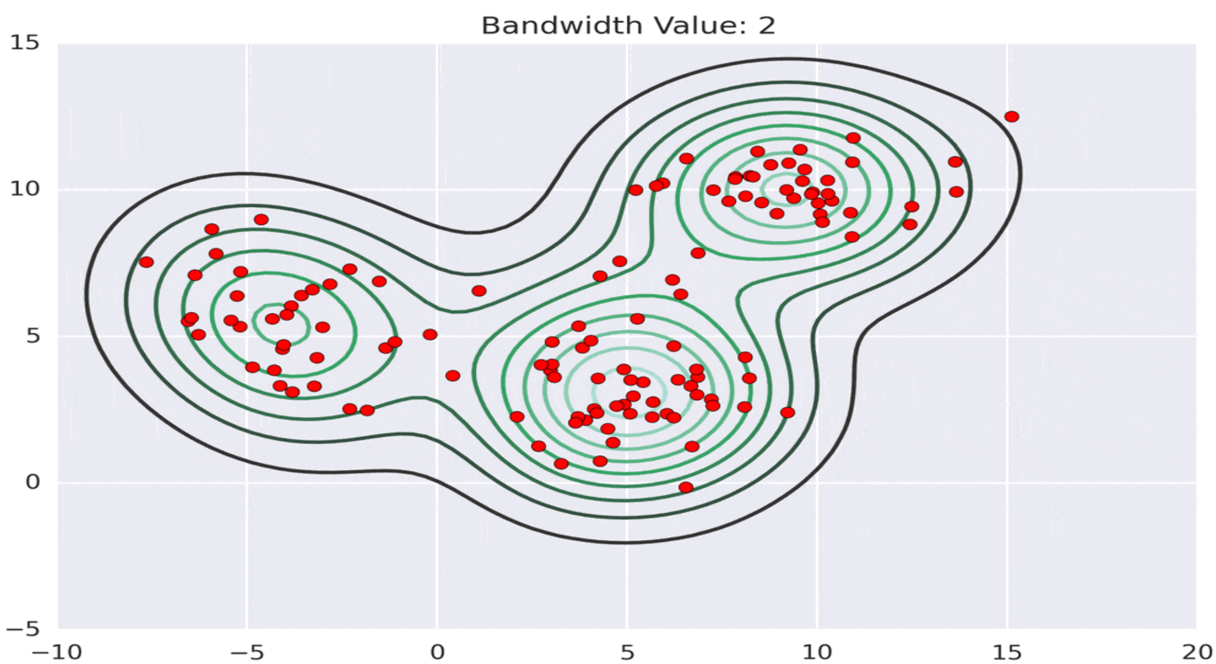
Mean shift clustering

- Cluster: all data points in the *attraction basin* of a mode
- Attraction basin: the region for which all trajectories lead to the same mode





20
15
10
5
0
20
15
10
5
0
−5
−10
−10
−5
0
5
10
15
20



Bandwidth Value: 2
15
10
5
0
−5
−10
−5
0
5
10
15
20