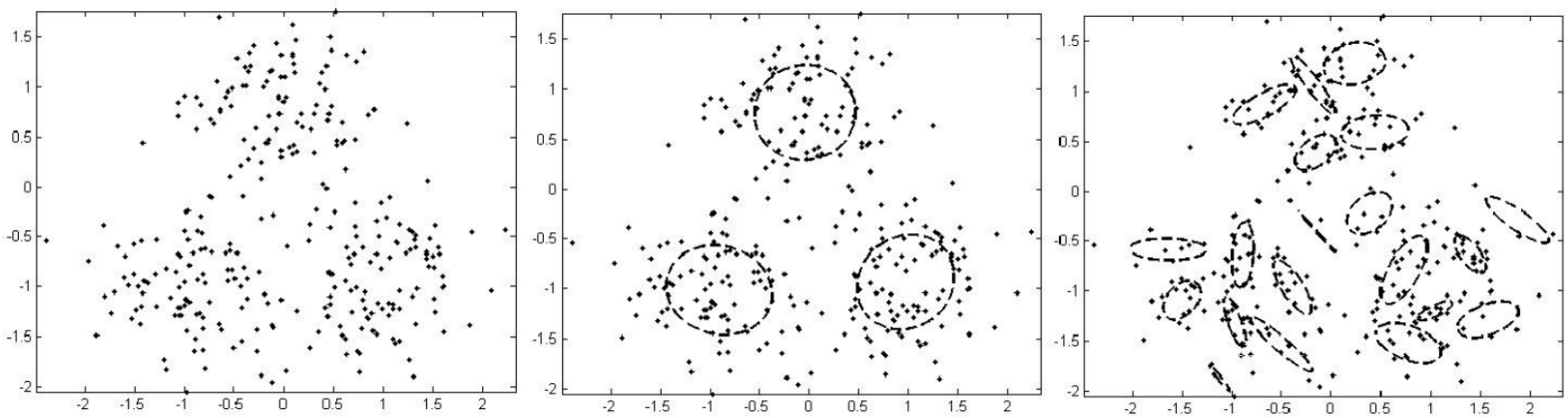
K-means, Mean Shift

DS 600 Data Mining

K-means

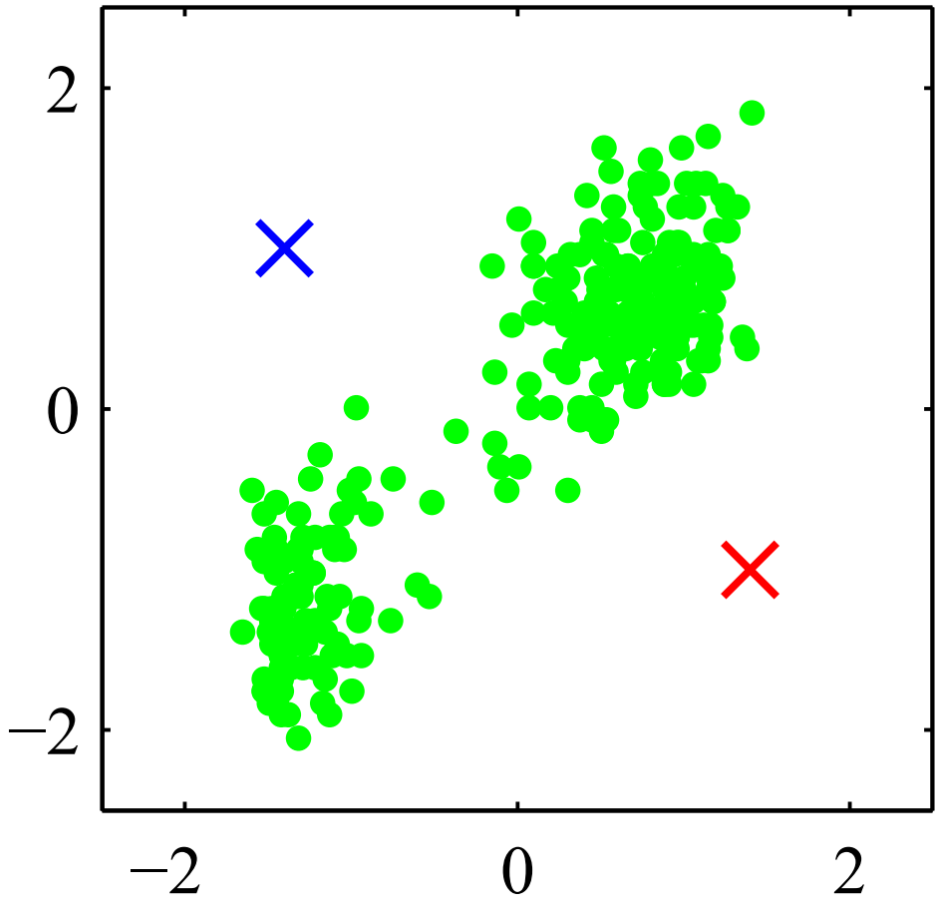
Choosing the Number of Mixtures



In clustering, there is no right or wrong way to choose the number of clusters M.

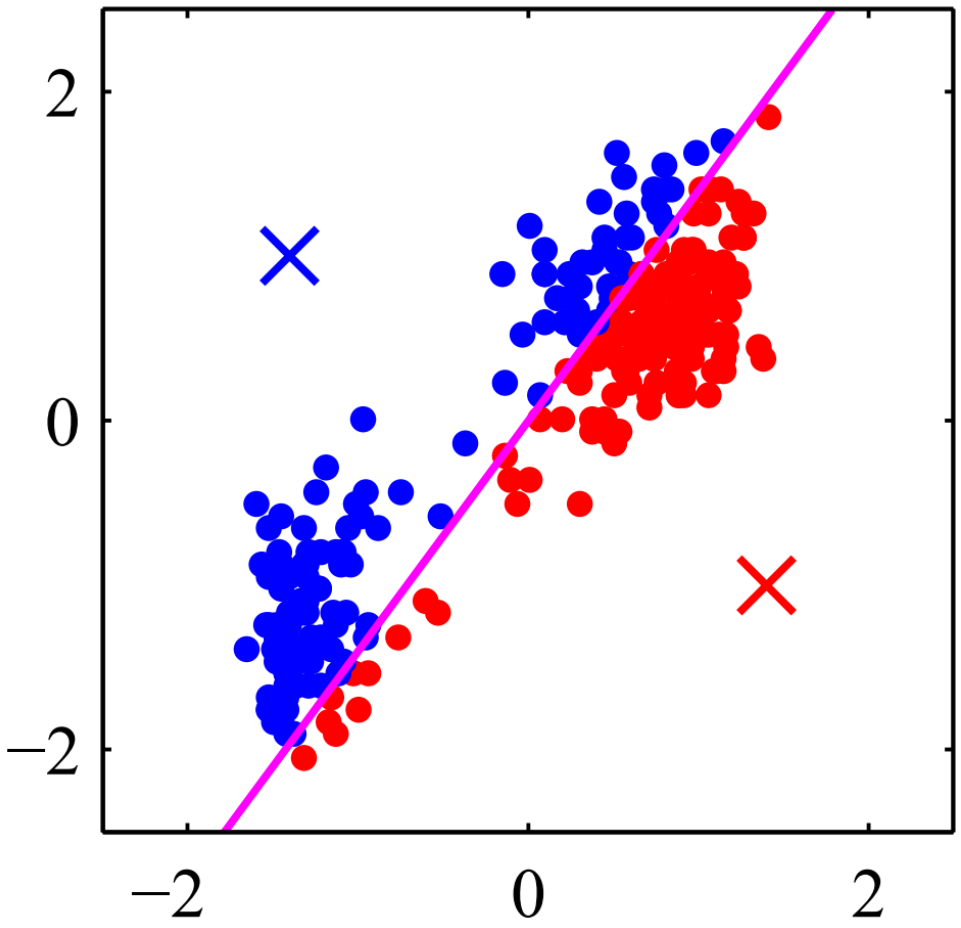
- One could let the number of clusters $M \to N$ but many of these clusters will be empty or contain very few points.
- One could try from $M = 2$ and increase the number of clusters until the log likelihood of the data stops improving.
- Information criteria based scoring: Akaike, Bayesian, etc.

K-means Demo



A random initialization

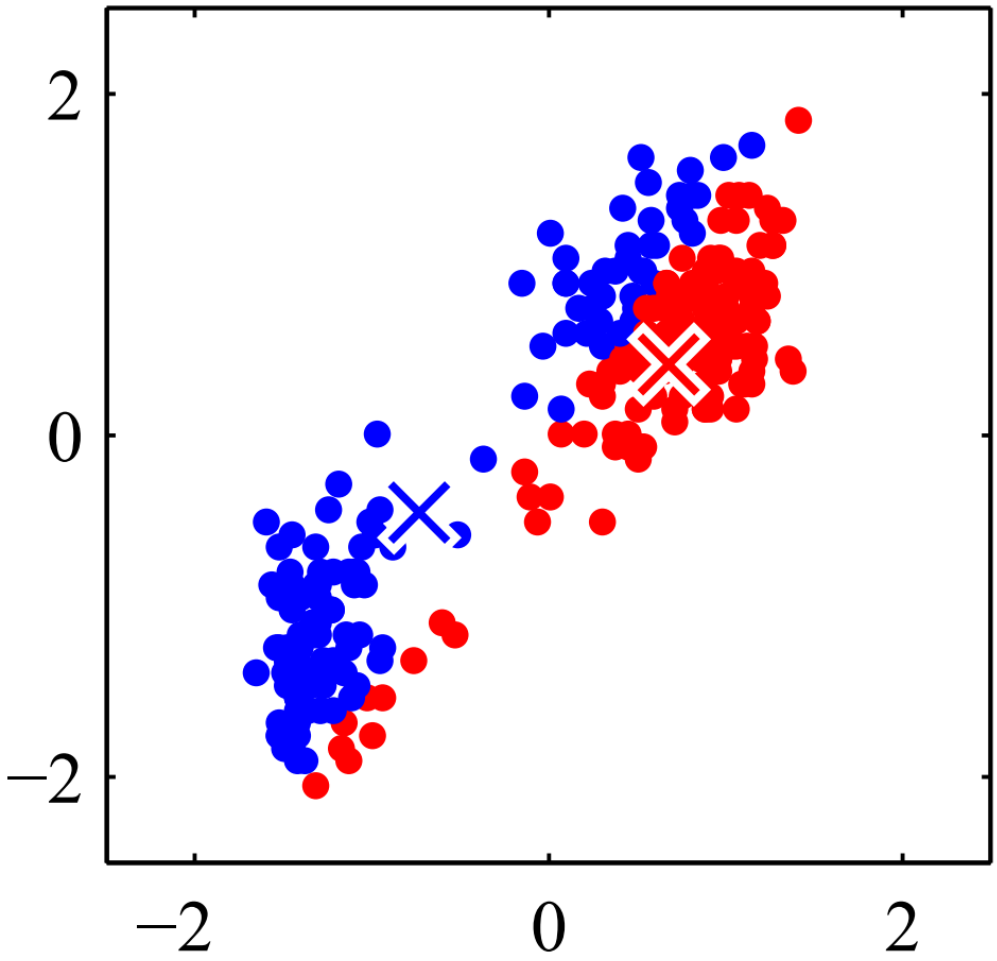
K-means Demo



Iteration 1

Assign data to clusters

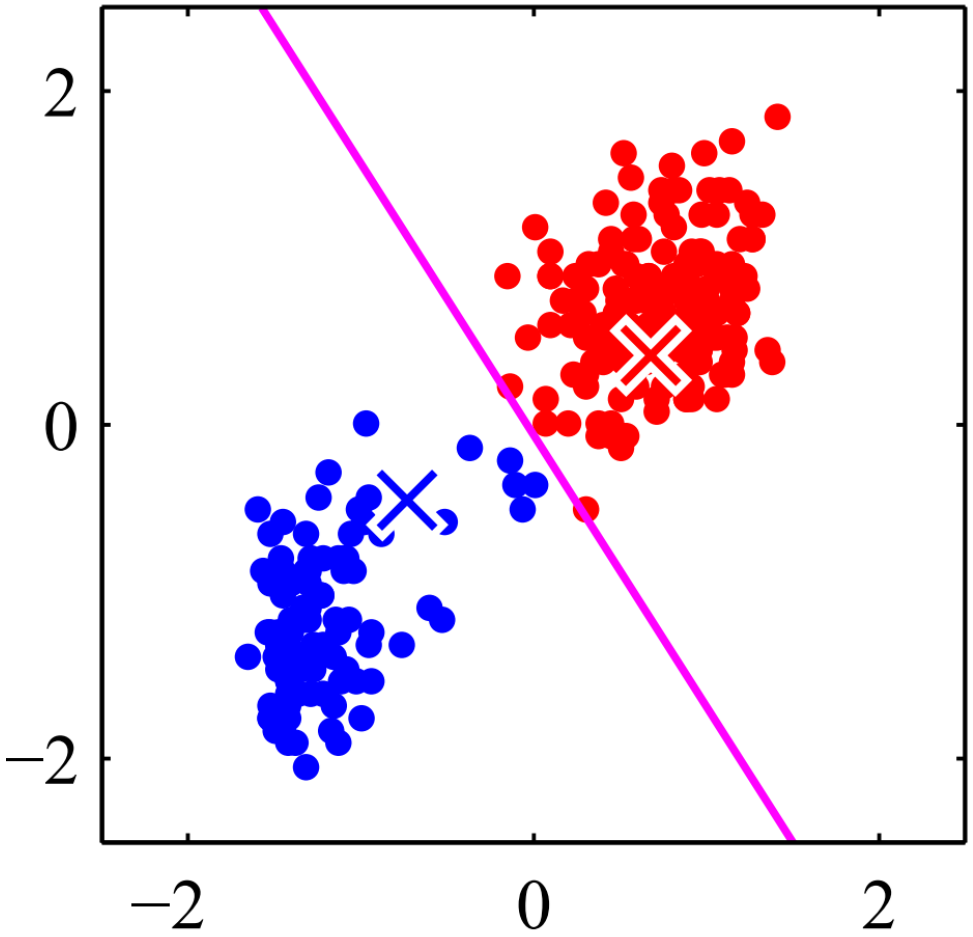
K-means Demo



Iteration 1

Update the centroids

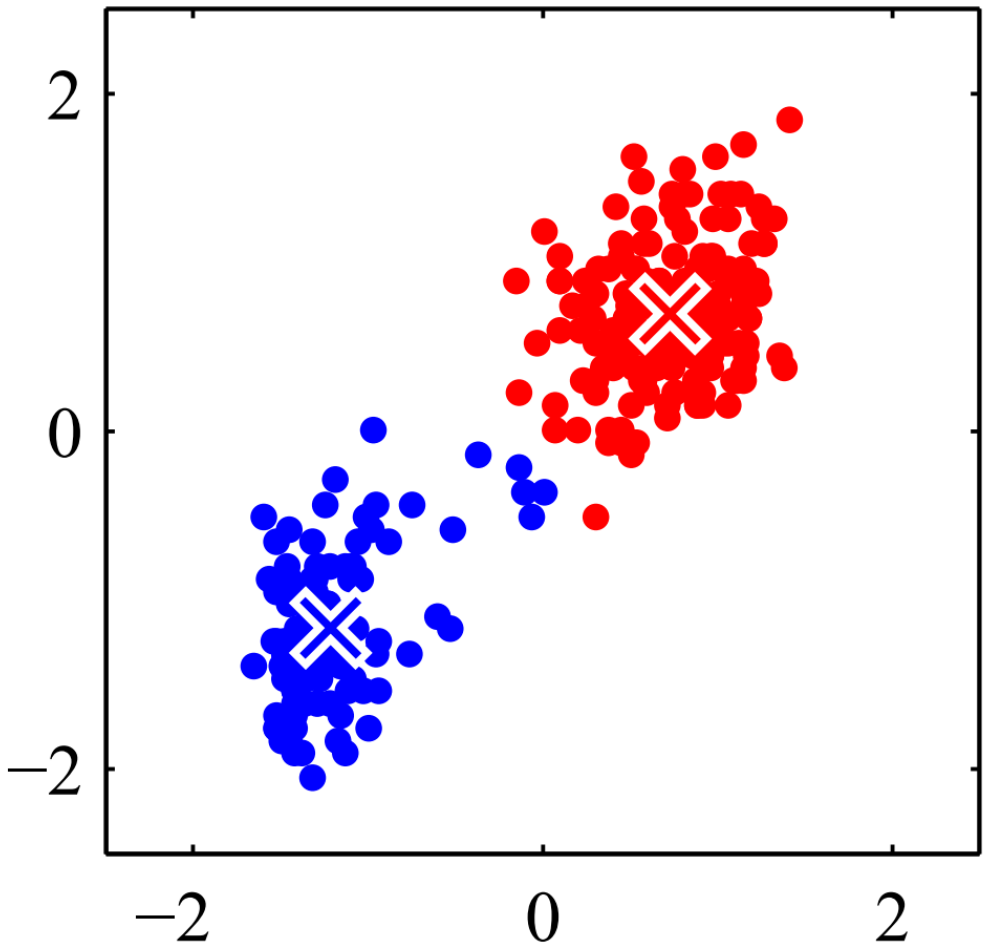
K-means Demo



Iteration 2

Assign data to clusters

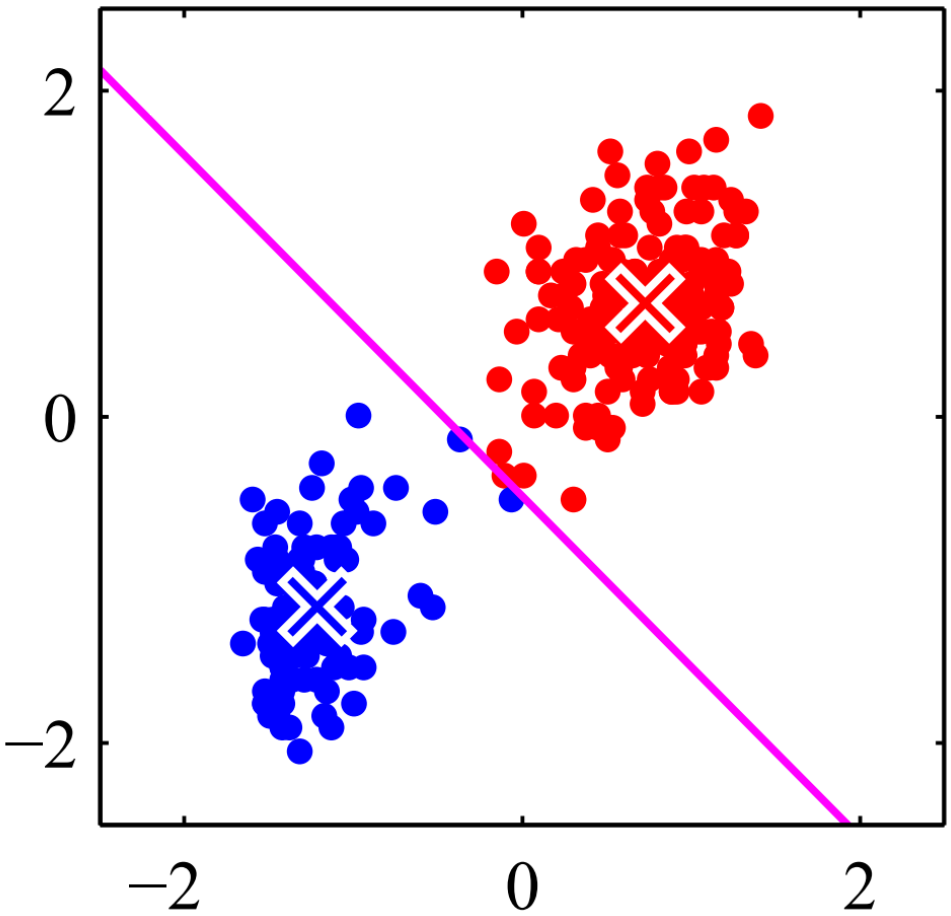
K-means Demo



Iteration 2

Update the centroids

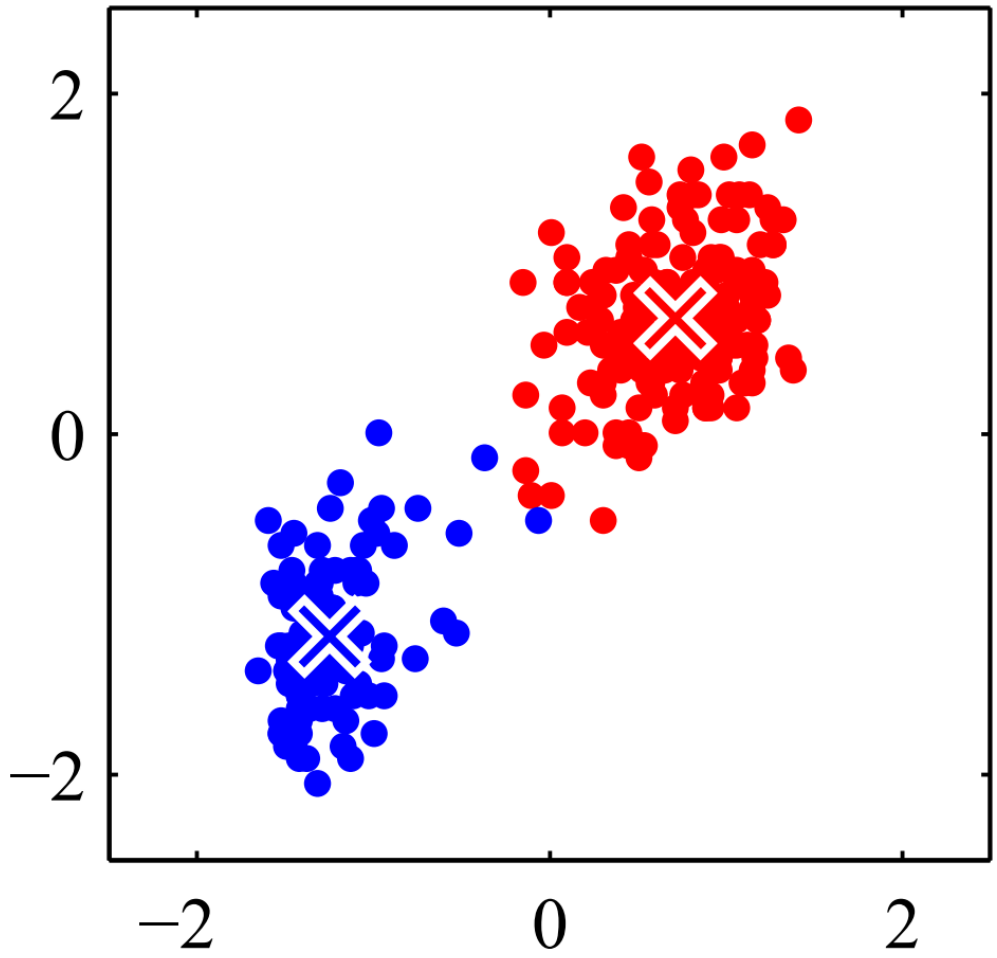
K-means Demo



Iteration 3

Assign data to clusters

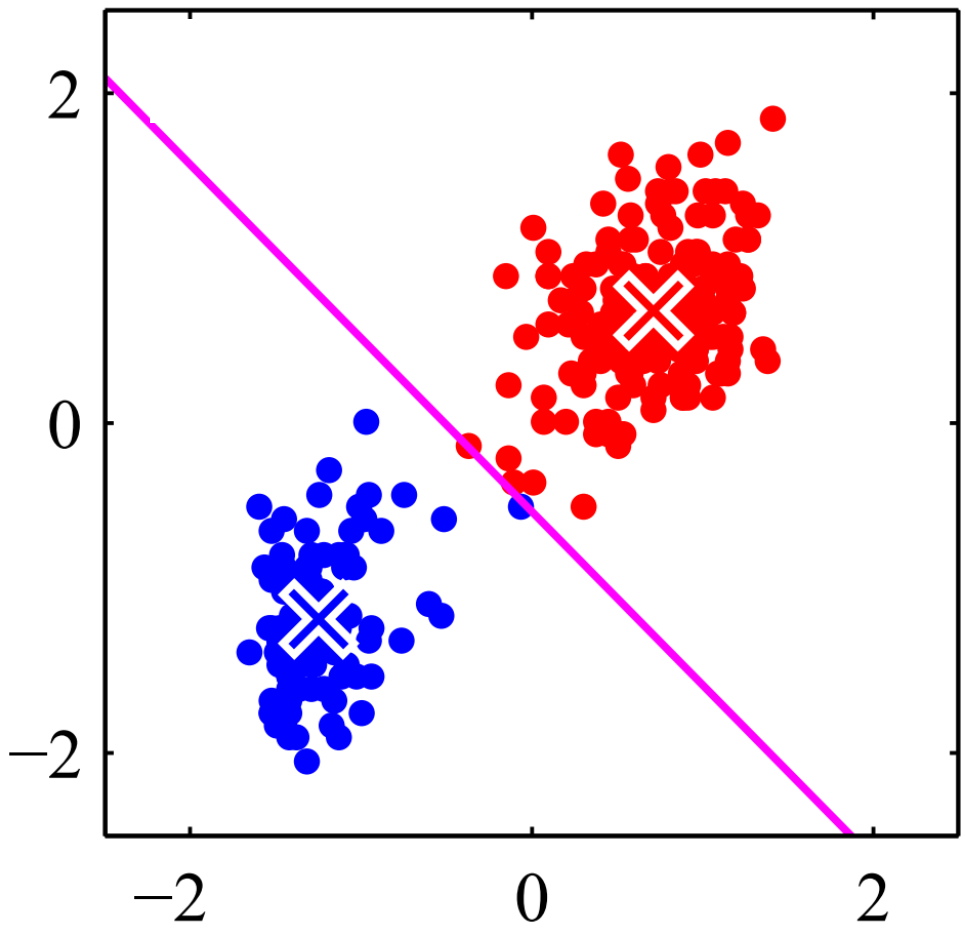
K-means Demo



Iteration 3

Update the centroids

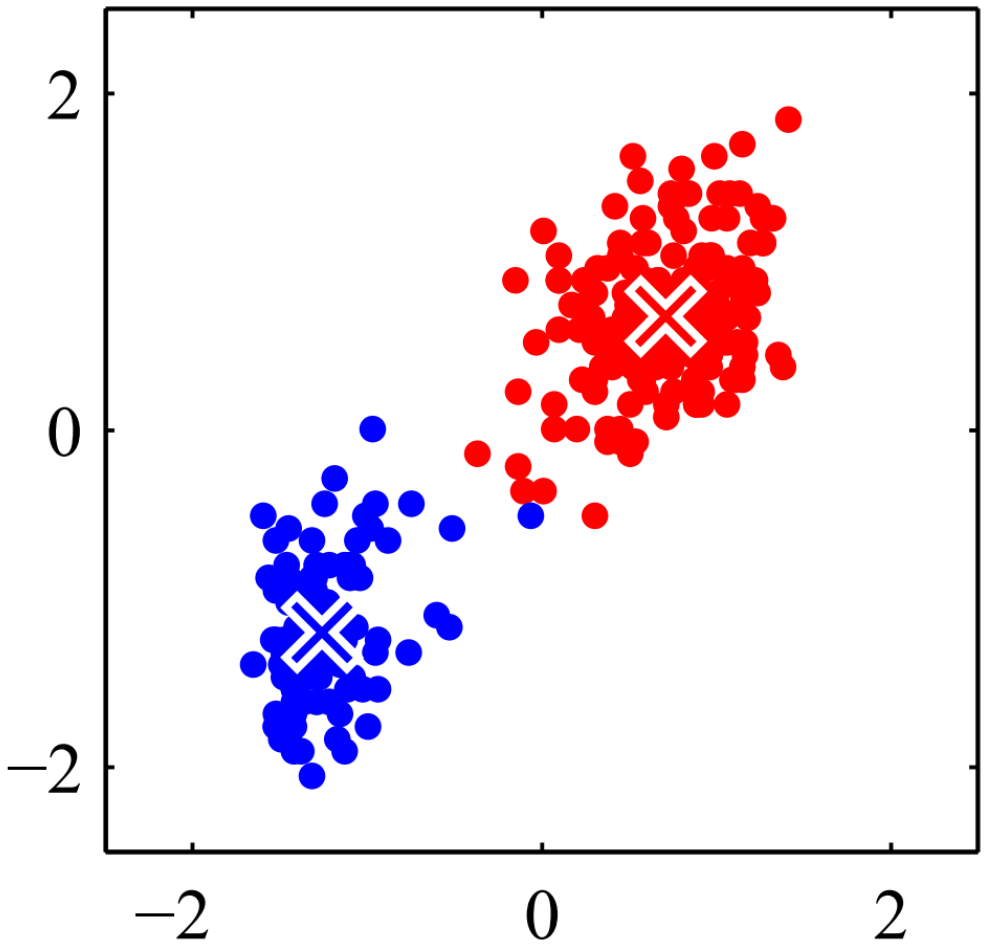
K-means Demo



Iteration 4

Assign data to clusters

K-means Demo



Iteration 4

Update the centroids

Mean Shift

Mean Shift = Gradient Ascent on Density Estimate

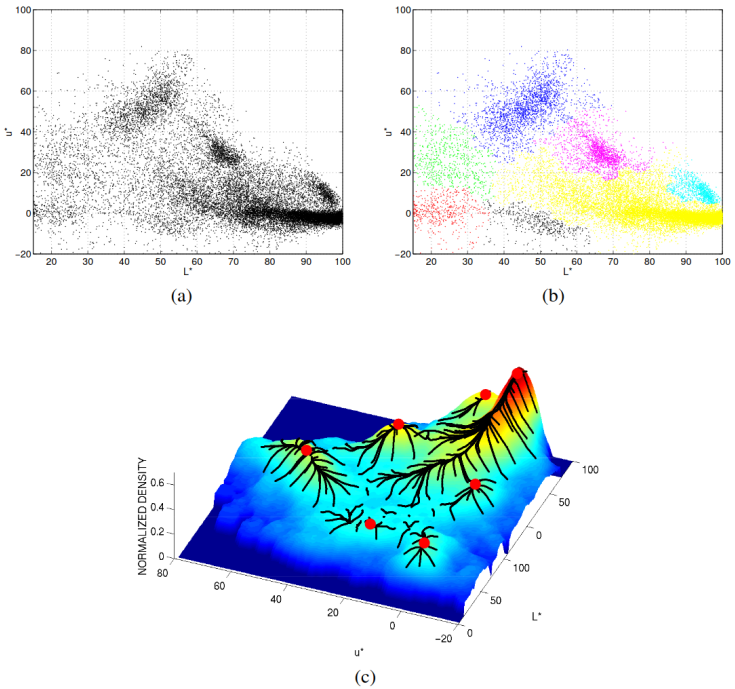
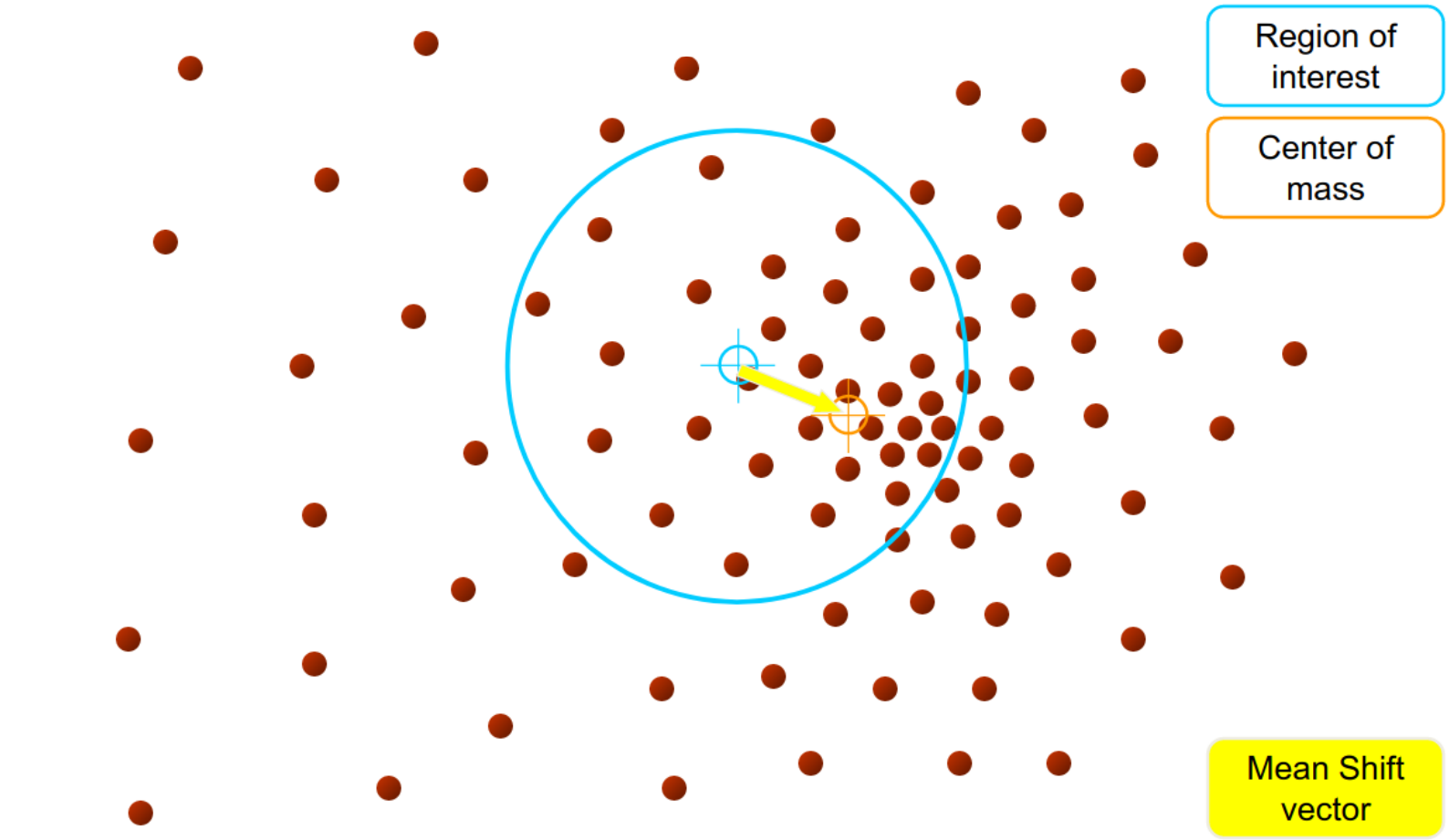


Figure 2: Example of a 2D feature space analysis. (a) Two dimensional data set of $110,400$ points representing the first two components of the $L^*u^*v^*$ space shown in Figure 1b. (b) Decomposition obtained by running 159 mean shift procedures with different initializations. (c) Trajectories of the mean shift procedures drawn over the Epanechnikov density estimate computed for the same data set. The peaks retained for the final classification are marked with red dots.

D. Comaniciu and P. Meer. "Mean shift: A robust approach toward feature space analysis". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.5 (2002), pp. 603–619

Local modes of the underlying density function can be used as cluster centers. This is the idea of mean shift method.

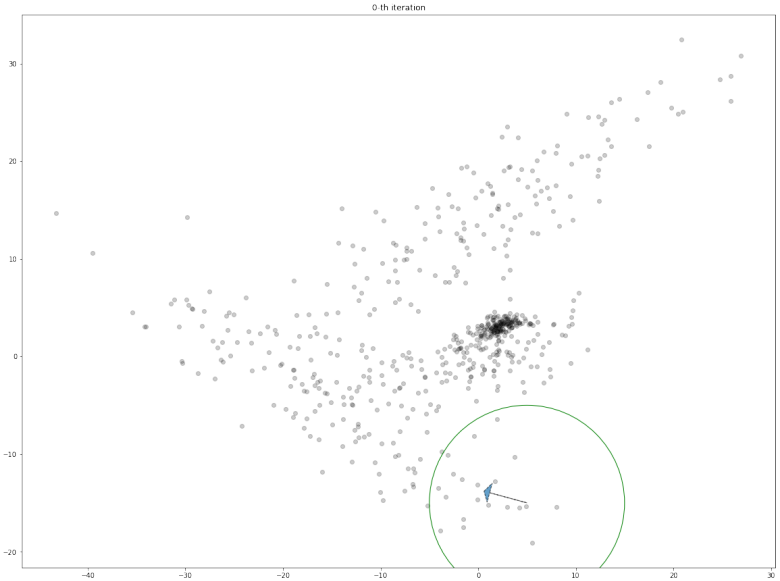
Mean Shift Clustering

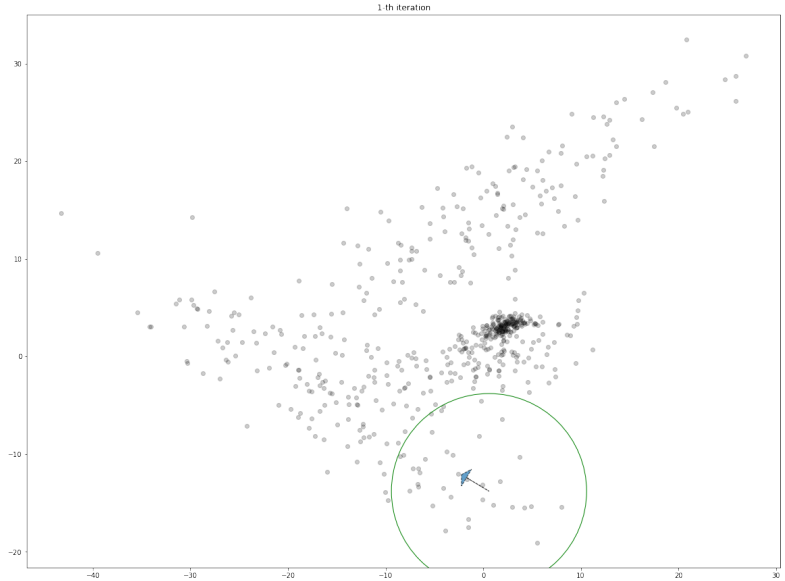


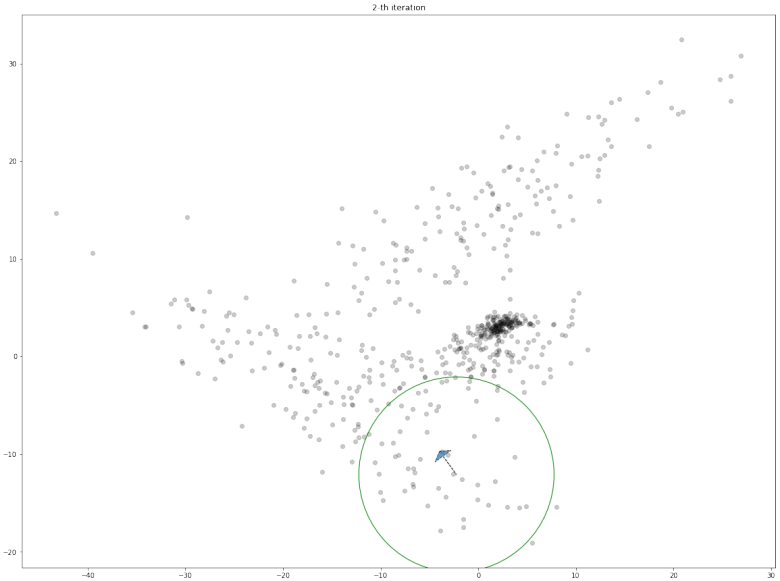
Slide by Y. Ukrainitz & B. Sarel

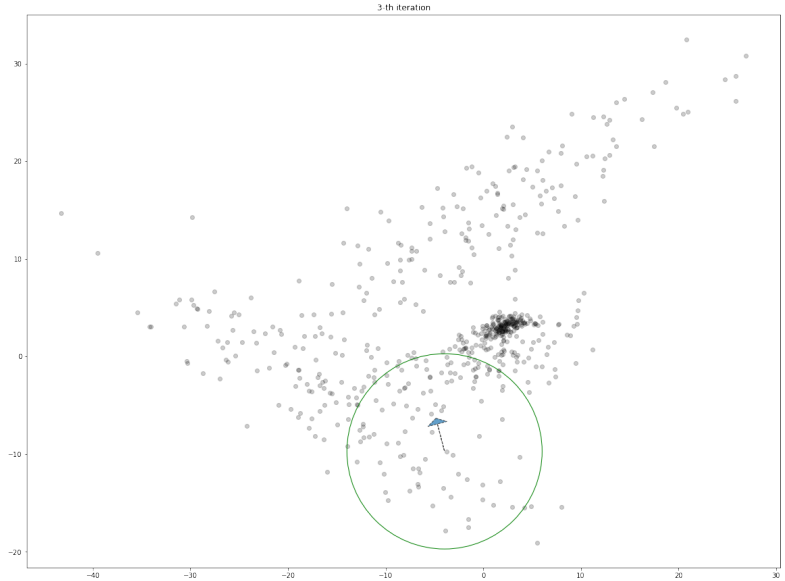
Iteratively move each point toward the center of mass in a window defined around it.

Mean Shift Demo

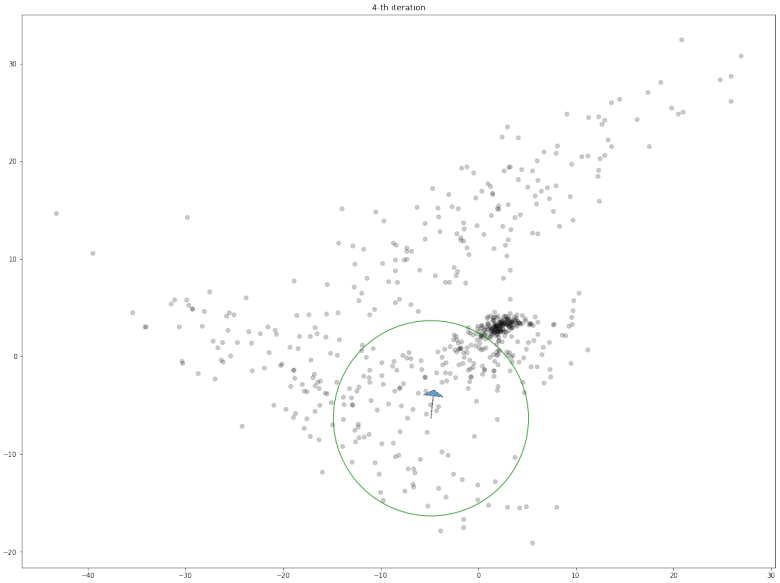


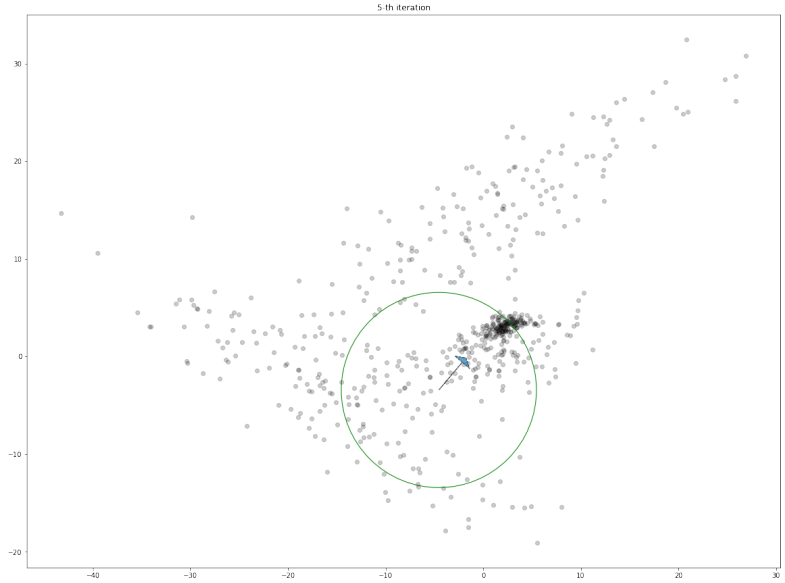


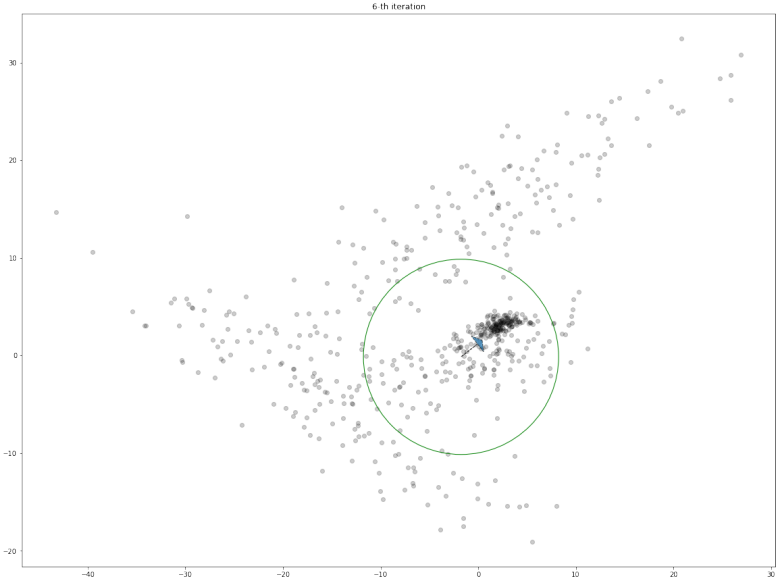


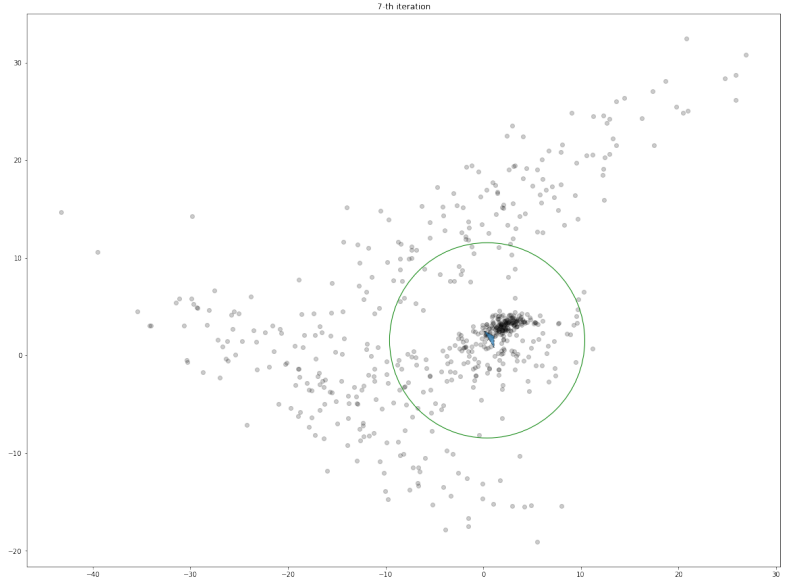


Mean Shift Demo

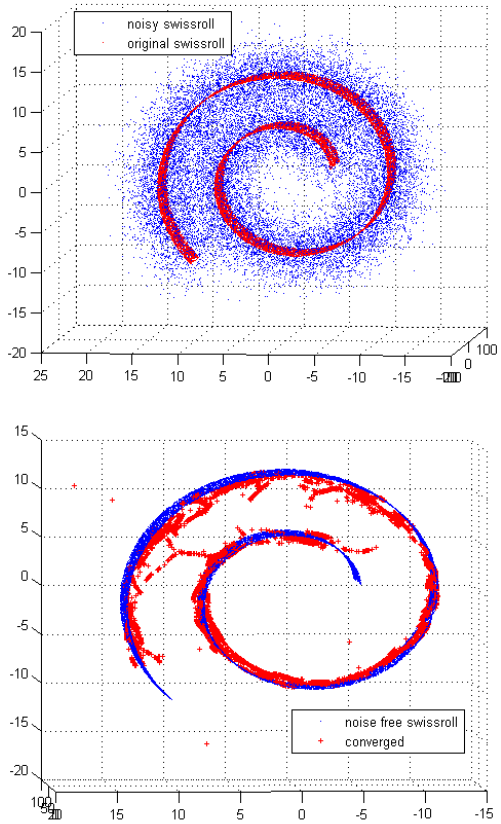








Mean Shift Example



Unrolling the swiss roll

Image segmentation

Choosing Parameter

Mean shift clustering needs the bandwidth parameter specified:

$$\mathbf{x}^{(i,k+1)} = \sum_{j=1}^{N} \frac{g\left(\frac{||\mathbf{x}^{(i,k)} - \mathbf{x}^{(j)}||^2}{h^2}\right)}{\sum_{p=1}^{N} g\left(\frac{||\mathbf{x}^{(i,k)} - \mathbf{x}^{(p)}||^2}{h^2}\right)} \mathbf{x}^{(i)}$$