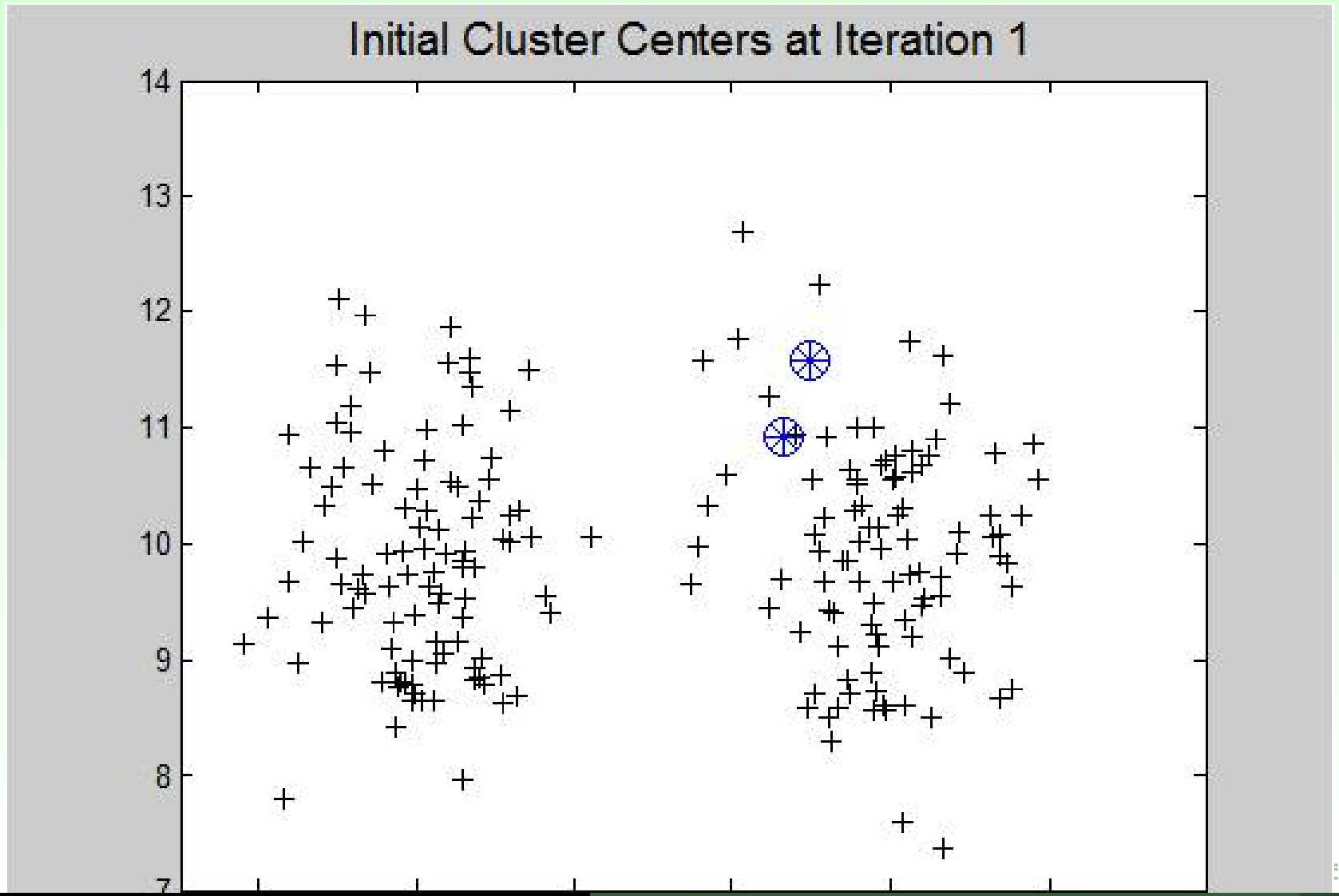
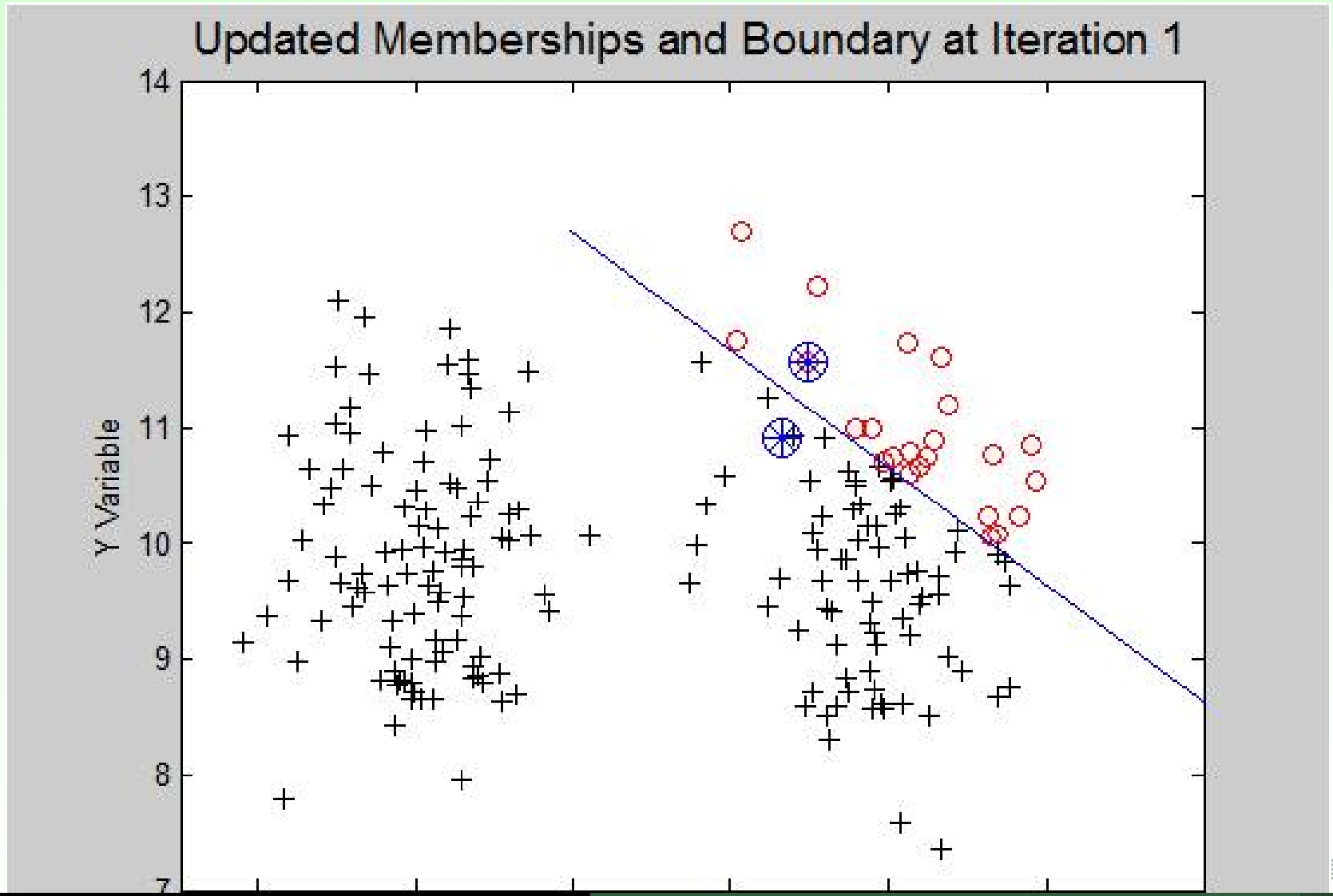
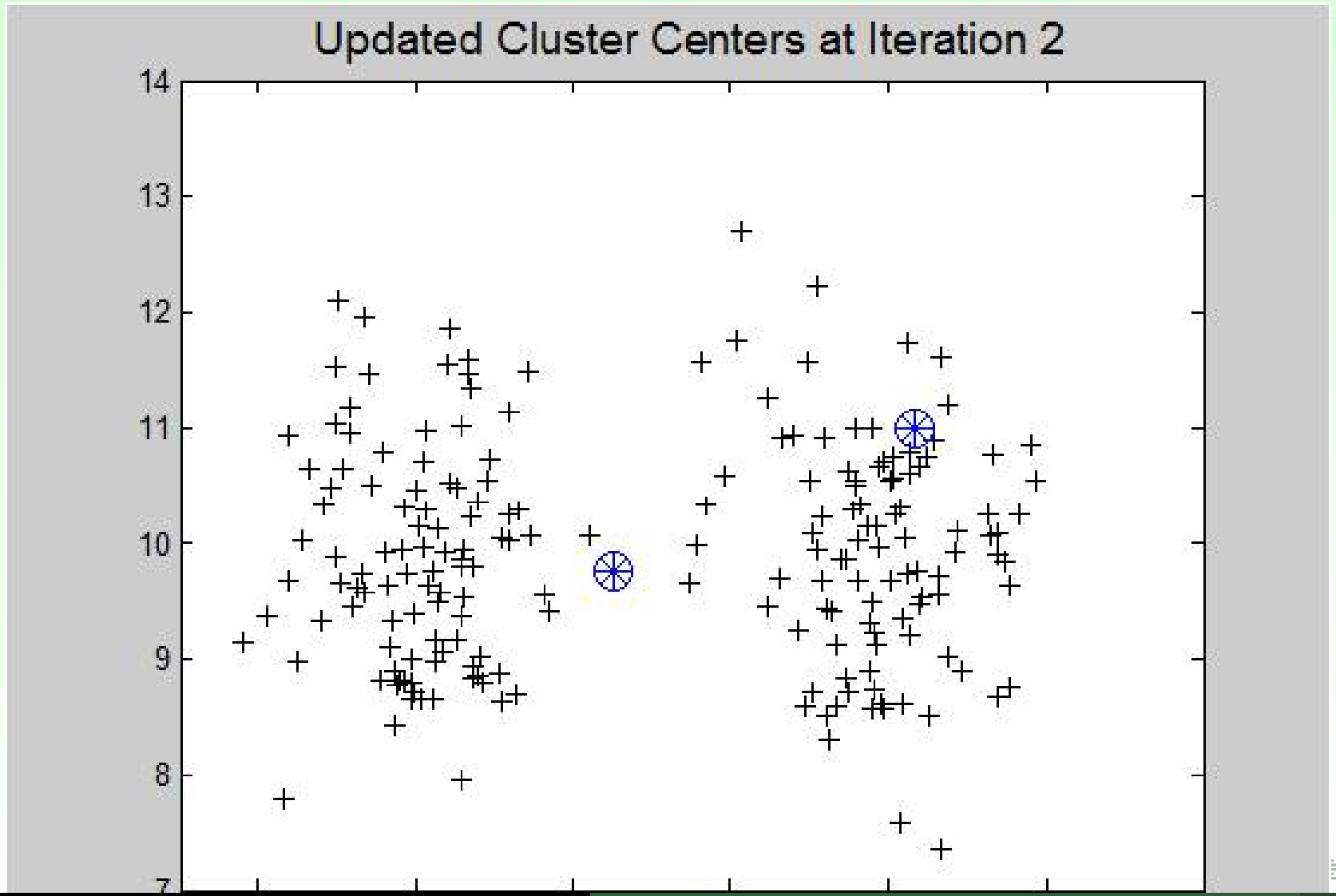
K-means illustration by Padhraic Smyth



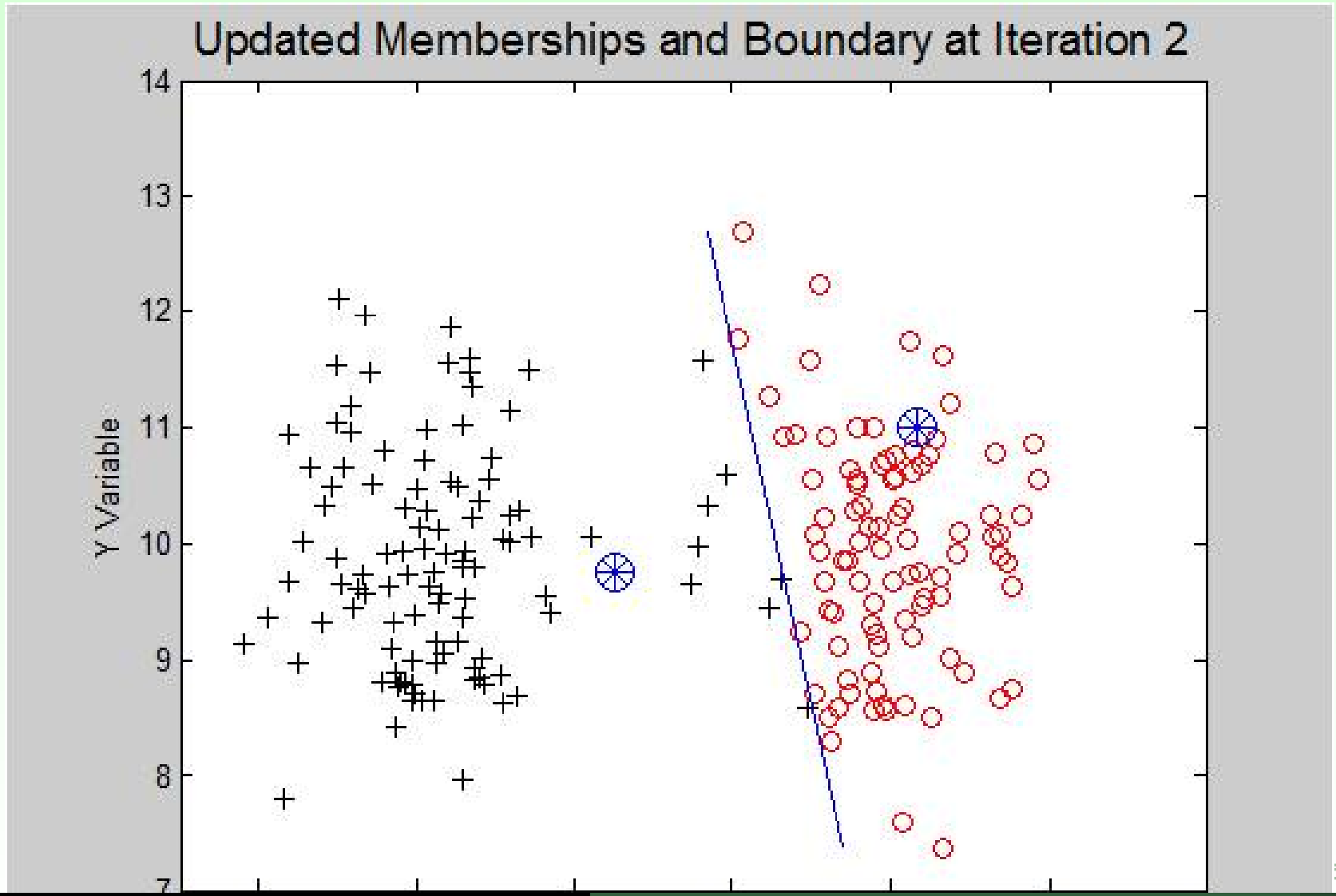
K-means illustration by Padhraic Smyth



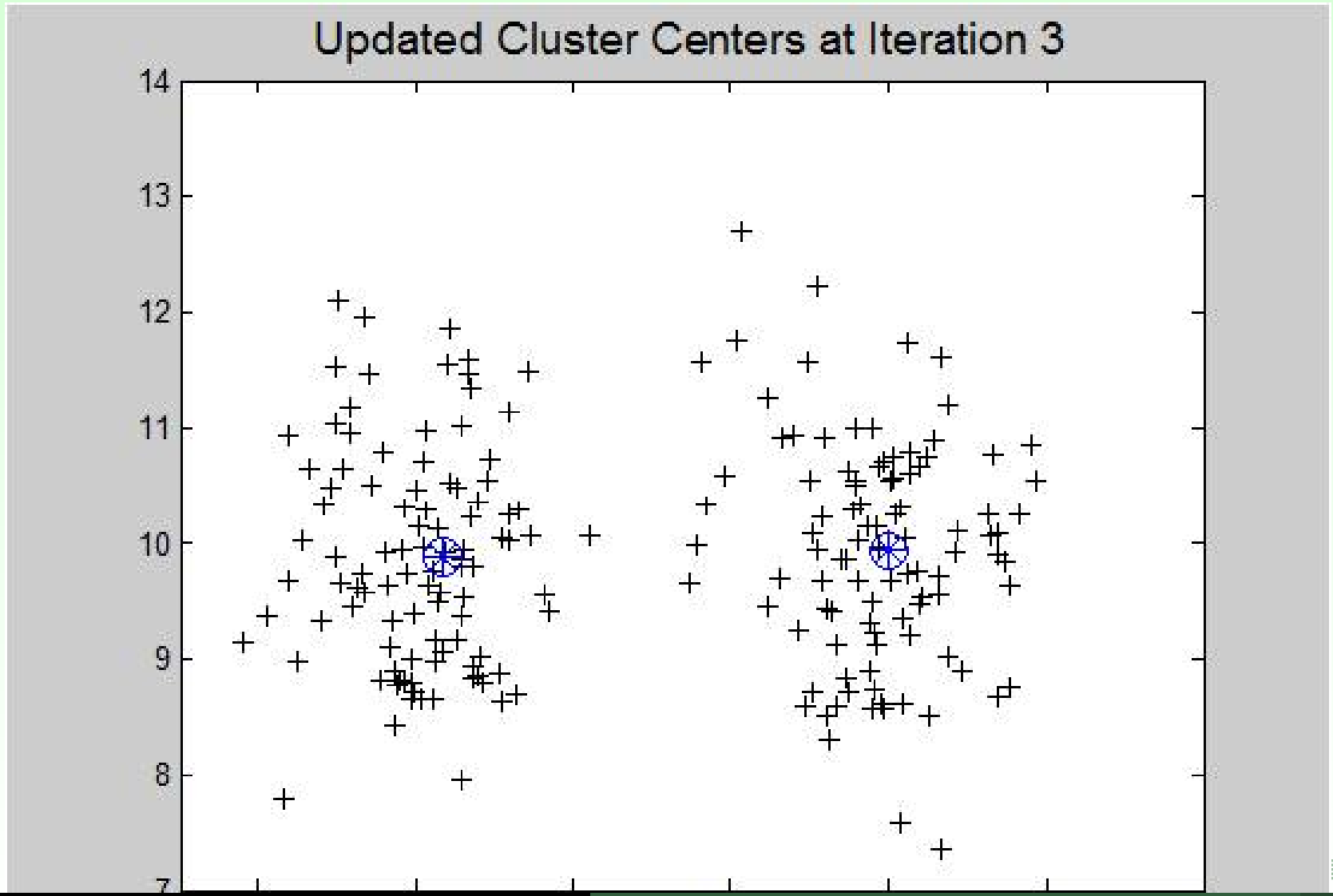
K-means illustration by Padhraic Smyth



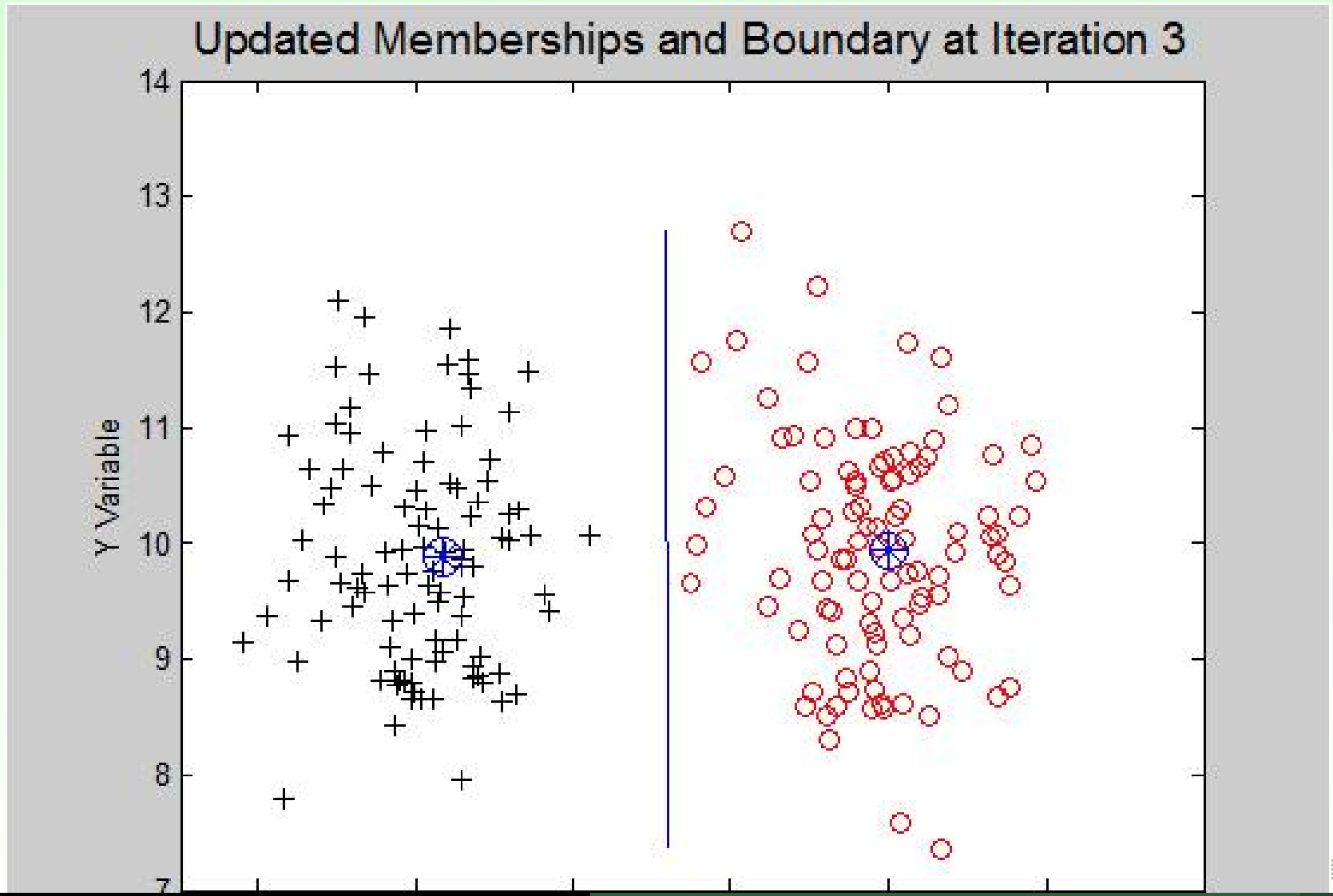
K-means illustration by Padhraic Smyth



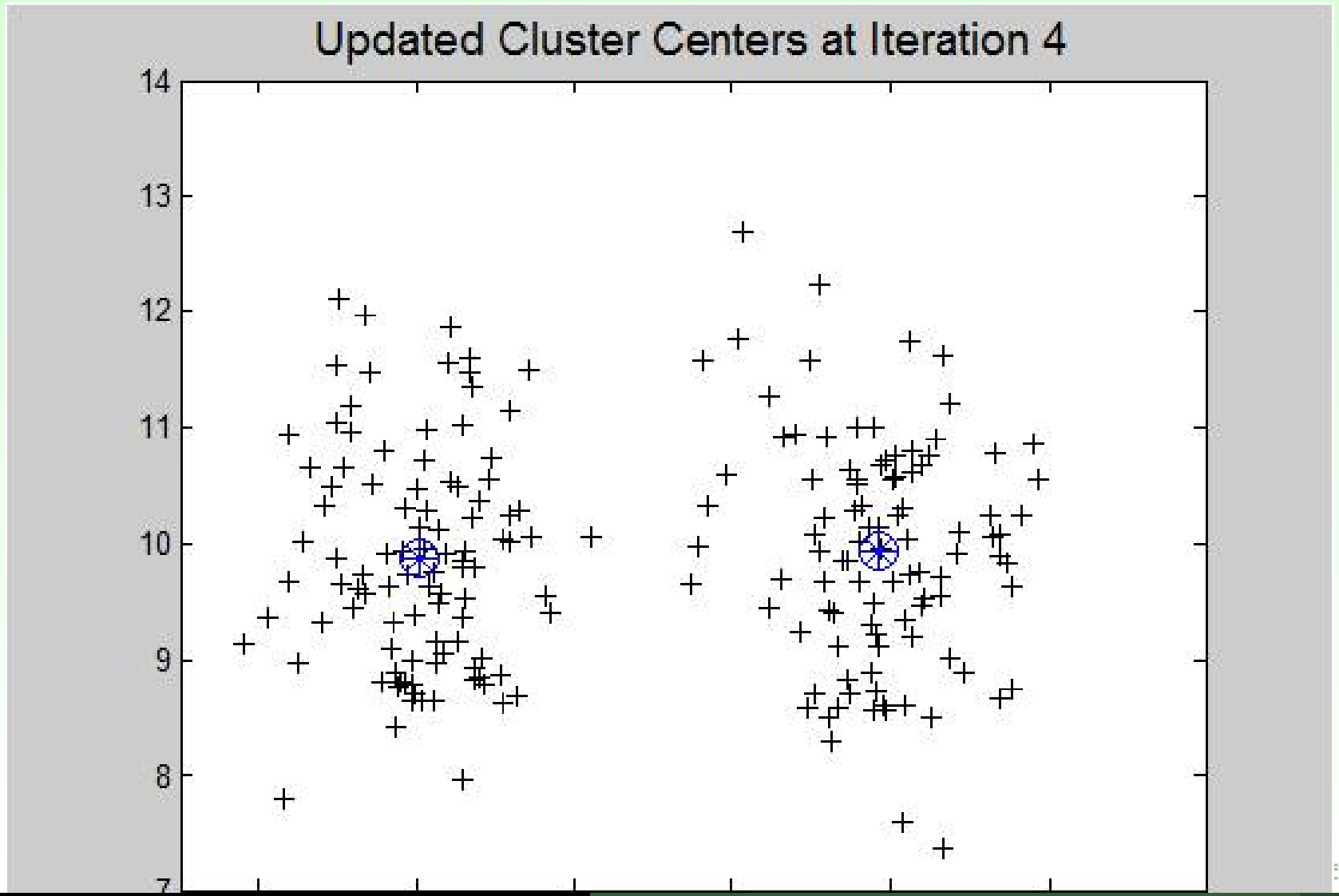
K-means illustration by Padhraic Smyth



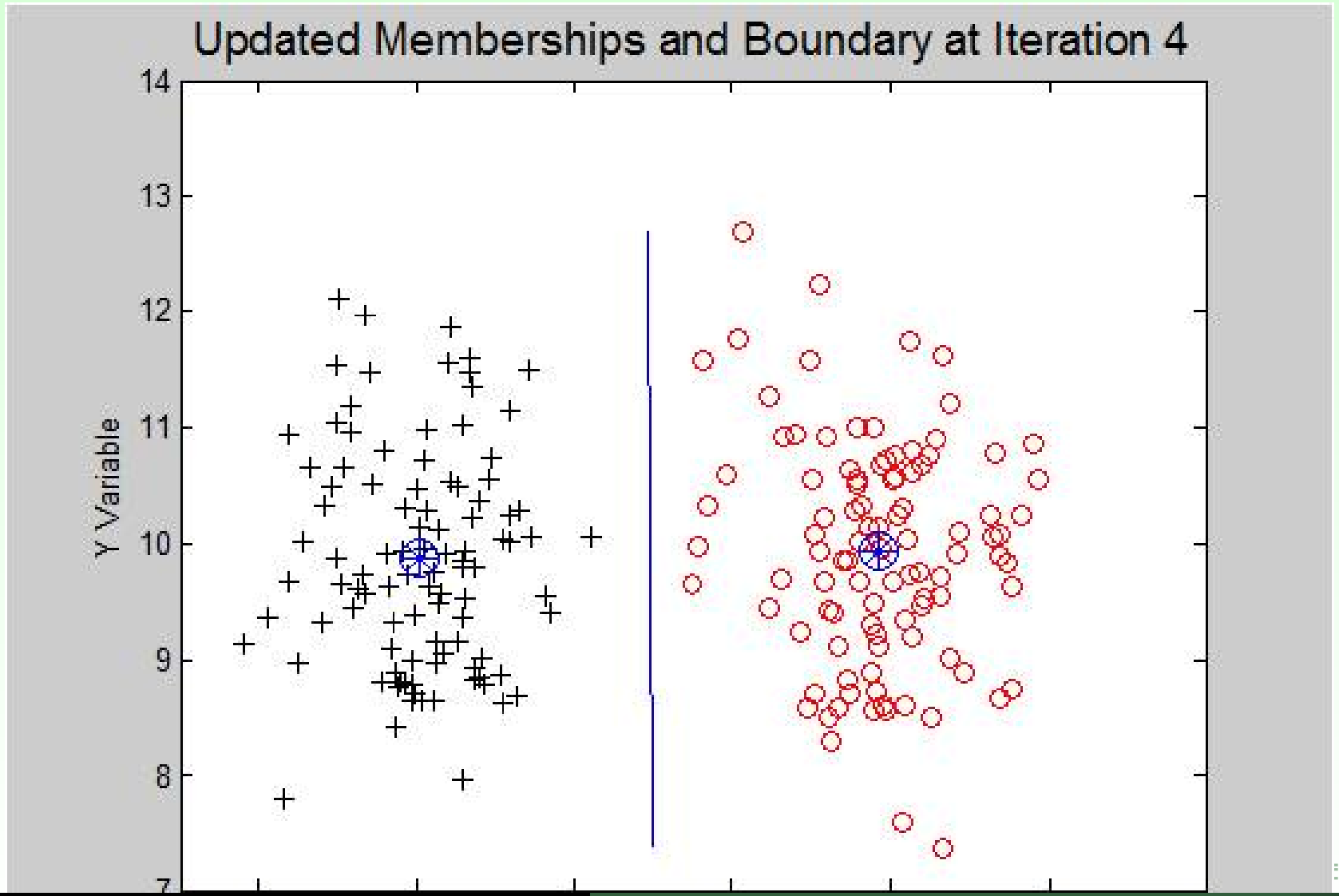
K-means illustration by Padhraic Smyth



K-means illustration by Padhraic Smyth



K-means illustration by Padhraic Smyth



K-means

- Let $X = \{\vec{x}_1, \ldots, \vec{x}_n\}$ be a finite set of objects.
- Assume that the objects are represented by m-dimensional vectors.
- A clustering scheme groups n objects into k clusters $C = \{\vec{c}_1, \ldots, \vec{c}_k\}$.
- The process begins by randomly choosing k objects as the centroids of the k clusters.
- The objects are assigned to one of the k clusters based on the minimum value of the distance $d(\mathbf{v}, \mathbf{x})$ between the object vector

K-means continued

- After the assignment of all the objects to various clusters, the new centroid vectors of the clusters are calculated as:

$$x_j = \frac{\sum_{\mathbf{v} \in \mathbf{x}} v_j}{\text{Size of cluster } \mathbf{x}}, \text{where } 1 \leq j \leq m. \quad (1)$$

- reassign the objects to new clusters
- The process stops when the centroids of clusters stabilize