The fuzzy k-means

In the k-means procedure, any feature vector $\mathbf{x}$ either is or is not a member of a particular cluster

Clusters produced by the k-means algorithm are sometimes referred as ”hard” or ”crisp”

In ”soft” or ”fuzzy” clusters, each $\mathbf{x}$ can have *a degree of membership in each cluster*

The membership function is selected based on heuristics

However, the algorithm can be interpreted probabilistically

- select arbitrarily the first cluster centers $\mathbf{z}_1(1), \mathbf{z}_2(1), \ldots, \mathbf{z}_K(1)$
- at the k^th iteration, find the degree of membership $u(j,i)$ of $\mathbf{x}_j$ in cluster *i*. For example:

$$u(j,i) = \frac{\exp(-\|\mathbf{x}_j - \mathbf{z}_i(k)\|^2)}{\sum_{r=1}^{K} \exp(-\|\mathbf{x}_j - \mathbf{z}_r(k)\|^2)}$$

- compute the new cluster centers $\mathbf{z}_1(k+1), \mathbf{z}_2(k+1), \ldots, \mathbf{z}_K(k+1)$:

$$\mathbf{z}_i(k+1) = \frac{\sum_{j=1}^{N} u(j,i)^2 \mathbf{x}_j}{\sum_{j=1}^{N} u(j,i)^2}$$

- if the cluster means are unchanged, terminate

Often this makes intuitively more sense than having to assign each point into some cluster

Fuzzy k-means is a heuristic procedure

Probabilistically, the degree of membership can be interpreted as the posteriori probability of cluster *i* given $\mathbf{x}$, i.e. $p(i|\mathbf{x})$

This brings us to the concept of *mixture densities,* which model an arbitrary pdf as a linear combination of *K* component densities

We will look at the *Gaussian mixture model* in detail later on in this course

Hierarchical clustering

- Agglomerative hierarchical clustering:
- Start with all the samples in own clusters, i.e. the number of clusters equals the number of data points
- At each iteration, compute the distances between all remaining clusters
- Merge the two closests clusters
- If the desired number of clusters has been obtained, terminate