Fuzzy Clustering
Algorithm: hard c-means (HCM)
(also known as k means)
Plot of tiles by frequencies (logarithms). The whole tiles (o) seem well separated from the cracked tiles (*). The **objective** is to find the two clusters.
1. Place two cluster centres (x) at random.
2. Assign each data point (* and o) to the nearest cluster centre (x)
1. Compute the new centre of each class
2. Move the crosses (x)
Iteration 2

Tiles data: o = whole tiles, * = cracked tiles, x = centres
Iteration 3

Tiles data: o = whole tiles, * = cracked tiles, x = centres
Iteration 4 (then stop, because no visible change)
Each data point belongs to the cluster defined by the nearest centre
Membership matrix $M$

$$m_{ik} = \begin{cases} 1 & \text{if } \| u_k - c_i \|^2 \leq \| u_k - c_j \|^2 \\ 0 & \text{otherwise} \end{cases}$$
A c-partition is defined as follows:

- All clusters $C$ together fill the whole universe $U$:
  $$\bigcup_{i=1}^{c} C_i = U$$

- Clusters do not overlap:
  $$C_i \cap C_j = \emptyset \quad \text{for all } i \neq j$$

- A cluster $C$ is never empty and it is smaller than the whole universe $U$:
  $$\emptyset \subset C_i \subset U \quad \text{for all } i$$

- There must be at least 2 clusters in a c-partition and at most as many as the number of data points $K$:
  $$2 \leq c \leq K$$
Minimise the total sum of all distances

\[ J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \left( \sum_{k, u_k \in C_i} \| u_k - c_i \|^2 \right) \]
Algorithm: fuzzy c-means (FCM)
Each data point belongs to two clusters to different degrees
1. Place two cluster centres
2. Assign a fuzzy membership to each data point depending on distance
1. Compute the new centre of each class
2. Move the crosses (x)
Tiles data: o = whole tiles, * = cracked tiles, x = centres
Tiles data: o = whole tiles, * = cracked tiles, x = centres
Tiles data: o = whole tiles, * = cracked tiles, x = centres

Iteration 10
Iteration 13 (then stop, because no visible change)
Each data point belongs to the two clusters to a degree
Fuzzy membership matrix $\mathbf{M}$

- Point $k$'s membership of cluster $i$:
  \[ m_{ik} = \frac{1}{\sum_{j=1}^{c} \left( \frac{d_{ik}}{d_{jk}} \right)^{2/(q-1)}} \]

  - Distance from point $k$ to current cluster centre $i$:
  \[ d_{ik} = \| \mathbf{u}_k - \mathbf{c}_i \| \]

  - Distance from point $k$ to other cluster centres $j$:
  \[ d_{jk} \]

  - Fuzziness exponent:
  \[ 2/(q-1) \]
Fuzzy membership matrix $\mathbf{M}$

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left( \frac{d_{ik}}{d_{jk}} \right)^{2/(q-1)}}$$

$$= \frac{1}{\left( \frac{d_{ik}}{d_{1k}} \right)^{2/(q-1)} + \left( \frac{d_{ik}}{d_{2k}} \right)^{2/(q-1)} + \cdots + \left( \frac{d_{ik}}{d_{ck}} \right)^{2/(q-1)}}$$

Gravitation to cluster $i$ relative to total gravitation
Fuzzy c-partition

All clusters $C$ together fill the whole universe $U$.
Remark: The sum of memberships for a data point is 1, and the total for all points is $K$.

$$\bigcup_{i=1}^{c} C_i = U$$

$C_i \cap C_j = \emptyset$ for all $i \neq j$

$\emptyset \subset C_i \subset U$ for all $i$

$2 \leq c \leq K$

A cluster $C$ is never empty and it is smaller than the whole universe $U$.

Not valid: Clusters do overlap.

There must be at least 2 clusters in a c-partition and at most as many as the number of data points $K$. 

Links


◆ PapSmear tutorial. URL http://fuzzy.iau.dtu.dk/smear/