K-means:

Means Business

Justin Talbot,Lei Hu

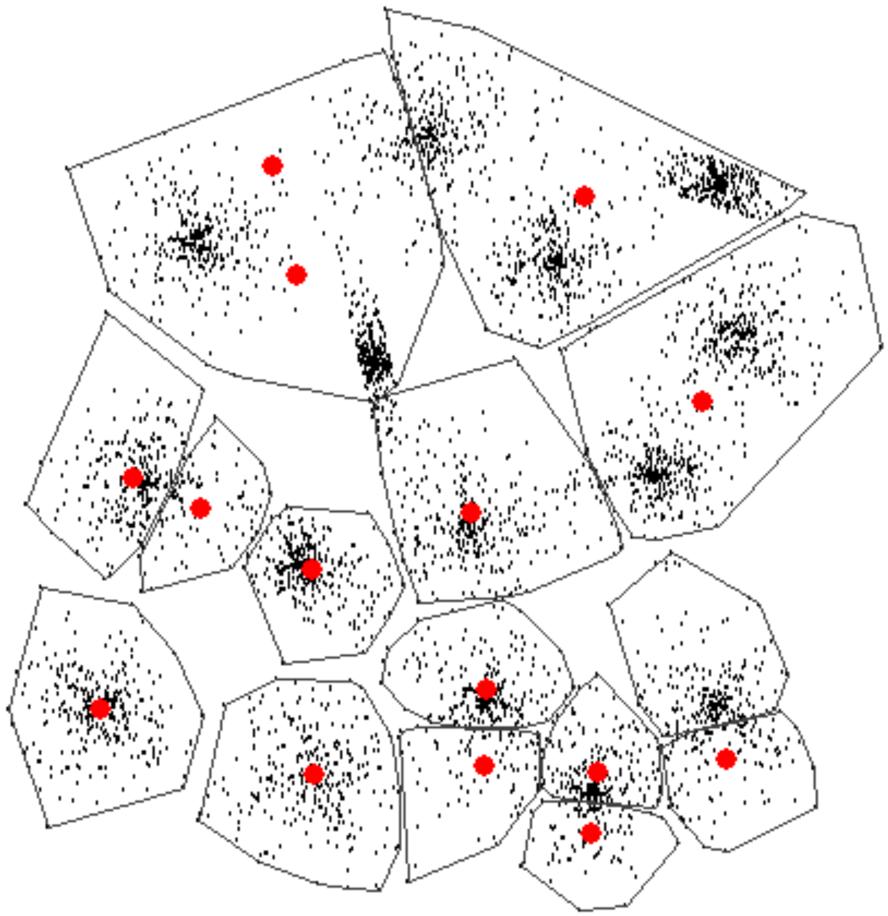
K-means

- An iterative algorithm for clustering
- An approximation to finding the maximum likelihood estimates for the means of the clusters
- Aims to minimize the within cluster sum of squares error
- No Guarantee of the global optimum

Algorithm

- Pick a number of clusters k, and the central points of the clusters
- Assign each point to the cluster with the closest mean
- Calculate the new means to be the centroid of the points in the cluster
- Repeat until convergence

Example



http://cs.joensuu.fi/sipu/

K-means++
(The advantage of Careful Seeding)

- By selecting the initial centers you can guarantee lower total error
- $E(\phi) \leq 8(\ln k + 2)\phi_{OPT}$

Algorithm

- Choose initial center c_1 uniformly at random
- Choose the next center c_i by setting c_i = to a point in the dataset x’ with probability

$$\frac{D(x')^2}{\sum_x D(x)^2}$$

- D(x) being the shortest distance from a data point x to the closest center already chosen

Fuzzy k-means clustering

Probability of cluster membership

$$\hat{P}(w_i|x_j,\hat{\theta}) = \frac{1}{d(center_k, x)}$$

$$\sum_{i=1}^{c} \hat{P}(\omega_i|\mathbf{x}_j) = 1, \qquad j = 1, \ldots, n,$$

Cost function

$$J_{fuz} = \sum_{i=1}^{c} \sum_{j=1}^{n} [\hat{P}(\omega_i | \mathbf{x}_j, \hat{\boldsymbol{\theta}})]^b ||\mathbf{x}_j - \boldsymbol{\mu}_i||^2$$

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||\mathbf{x}_n - \boldsymbol{\mu}_k||^2$$

Minimization of cost function

$$J_{fuz} = \sum_{i=1}^{c} \sum_{j=1}^{n} [\hat{P}(\omega_i|\mathbf{x}_j, \hat{\boldsymbol{\theta}})]^b ||\mathbf{x}_j - \boldsymbol{\mu}_i||^2$$

$$\sum_{i=1}^{c} \hat{P}(\omega_i|\mathbf{x}_j) = 1, \qquad j = 1, \ldots, n,$$

Solution

$$\partial J_{fuz}/\partial \boldsymbol{\mu}_i = 0 \quad \text{and} \quad \partial J_{fuz}/\partial \hat{P}_j = 0$$

$$\boldsymbol{\mu}_j = \frac{\sum_{j=1}^{n} [\hat{P}(\omega_i|\mathbf{x}_j)]^b \mathbf{x}_j}{\sum_{j=1}^{n} [\hat{P}(\omega_i|\mathbf{x}_j)]^b}$$

$$\hat{P}(\omega_i|\mathbf{x}_j) = \frac{(1/d_{ij})^{1/(b-1)}}{\sum_{r=1}^{c} (1/d_{rj})^{1/(b-1)}}$$

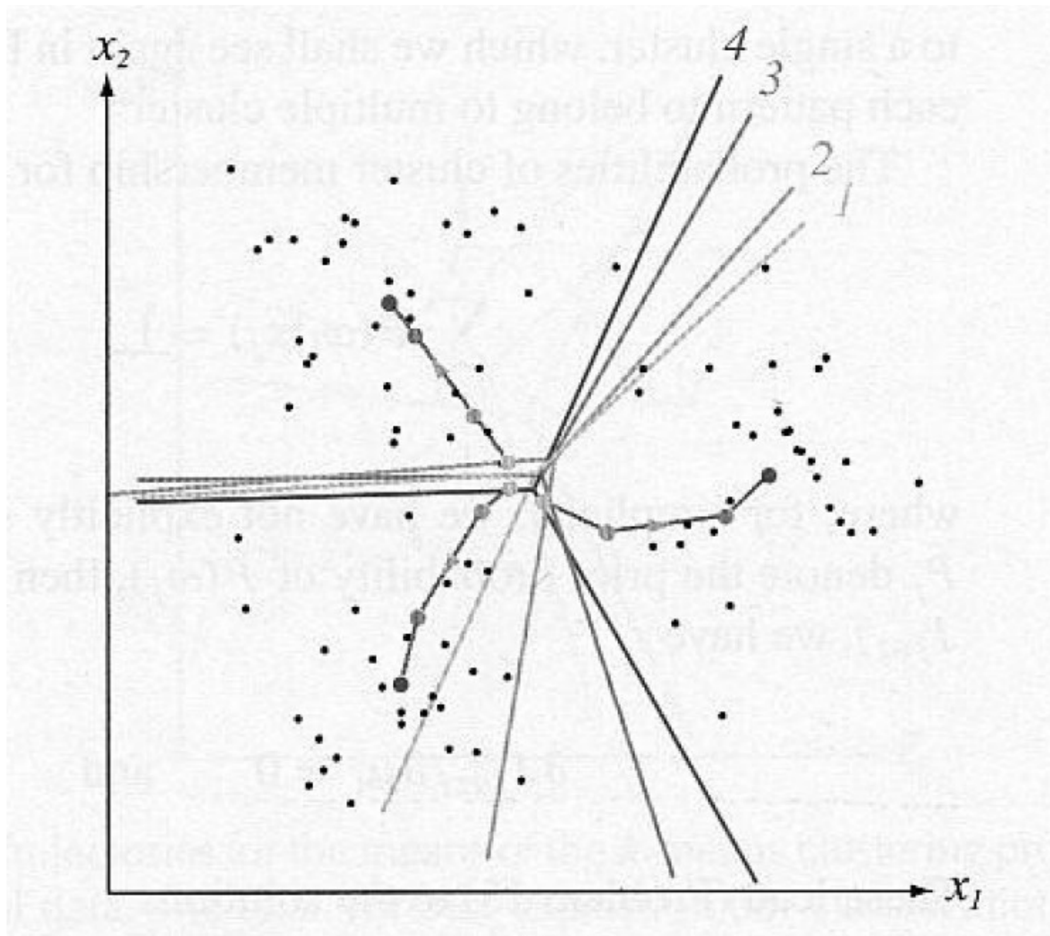
$$d_{ij} = ||\mathbf{x}_j - \boldsymbol{\mu}_i||^2$$

Algorithm

■ **Algorithm 2.** **(Fuzzy k-Means Clustering)**

begin **initialize** $n, c, b, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_c, \hat{P}(\omega_i \mid \mathbf{x}_j), i = 1 \ldots, c;\ j = 1, \ldots, n$
    normalize $\hat{P}(\omega_i \mid \mathbf{x}_j)$
    do recompute $\boldsymbol{\mu}_i$
        recompute $\hat{P}(\omega_i \mid \mathbf{x}_j)$
    until small change in $\boldsymbol{\mu}_i$ and $\hat{P}(\omega_i \mid \mathbf{x}_j)$
  return $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_c$
end

Example



References

Duda, R., Hart, P., and Stork, D. (2001).
Pattern Classication.
John Wiley and Sons, 2nd edition.

Sergios T., Konstantinos K.(2006)
Pattern Recognition.
Elsevier 3rd Edition.

Arthur D., Vassilvitski S. (2007)
k-means++: the advantages of careful seeding
Proceedings 18th Annual ACM-SIAM Symposium on Discrete Algorithms

Questions