Types of Algorithms

[https://images-na.ssl-images-amazon.com/images/M/MV5BOTU5NDkzNTM5MV5BMl5BanBnXkFtZTgwMDU4ODE5MDE@._V1_QL50_.jpg]

- The Good: Polynomial Time Complexity, $\mathcal{O}(n^k)$
  Sorting, DFT, FFT
- The Bad: Exponential Time Complexity, $\mathcal{O}(k^n)$

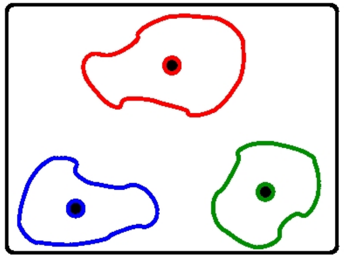
  NP–Hard

  NP–Complete

  Boolean functions of n variables: minterms, combinations of minterms; TSP
- The Ugly: Opt converge with prob 1, but infinite time

The Goal

- N points $\{\mathbf{x}_1 \ldots \mathbf{x}_N\}$ to be grouped into K clusters
- $\boldsymbol{\mu}_j$: prototype of the jth cluster (e.g., centre)
- such that the sum of squared distances of each data point to its closest vector $\boldsymbol{\mu}_j$ is a minimum
- Vector Quantisation: $\boldsymbol{\mu}_j$: codebook vectors
- Objective Function (Distortion Measure):



$$J \triangleq \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij} \, ||\mathbf{x}_i - \boldsymbol{\mu}_j||^2$$

- Binary Indicator variable $r_{ij} \triangleq 1$: ith data point assigned to jth cluster, & $\triangleq 0$, otherwise
- Find $\{r_{ij}\}$ & $\{\boldsymbol{\mu}_j\}$ such that J is minimised

Home Page
Title Page
Contents
Page 6 of 32
Go Back
Full Screen
Close
Quit

One way to solve this: two phases, EM-framework

Choose K initial cluster centres $\{\boldsymbol{\mu}_j\}$

1. Fixed $\{\boldsymbol{\mu}_j\}$, find assignment $\{r_{ij}\}$: Expectation
2. Keep $\{r_{ij}\}$ fixed, find $\{\boldsymbol{\mu}_j\}$: Maximisation

REPEAT till convergence/fixed max# of iterations

Convergence: each step provably convergent!

- Expectation Step: $\{\boldsymbol{\mu}_j\}$ fixed, minimise J wrt $\{r_{ij}\}$

$J \triangleq \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij} \, ||\mathbf{x}_i - \boldsymbol{\mu}_j||^2$ This is linear in $\{r_{ij}\}$.
For $\mathbf{x}_i$, assign it to the closest cluster centre.
Put $\{r_{ij}\}$ = 1 for $j = arg\ min_l ||\mathbf{x}_i - \boldsymbol{\mu}_l||$, 0, otherwise

- Minimisation Step: $\{r_{ij}\}$ fixed, minimise J wrt $\{\boldsymbol{\mu}_j\}$

$J \triangleq \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij} \, ||\mathbf{x}_i - \boldsymbol{\mu}_j||^2$ This is quadratic in $\{\boldsymbol{\mu}_j\}$.

$$\frac{\partial J}{\partial \{\boldsymbol{\mu}_j\}} = 0 \implies \sum_{i=1}^{N} 0 + 0 + \ldots + 2 r_{ij}(\mathbf{x}_i - \{\boldsymbol{\mu}_j\}) + \ldots = 0$$

$$\implies \sum_{i=1}^{N} r_{ij}(\mathbf{x}_i - \{\boldsymbol{\mu}_j\}) = 0 \implies$$

$$\sum_{i=1}^{N} r_{ij}\mathbf{x}_i = \left(\sum_{i=1}^{N} r_{ij}\right)\boldsymbol{\mu}_j \implies \boxed{\boldsymbol{\mu}_j = \frac{\sum_{i=1}^{N} r_{ij}\mathbf{x}_i}{\sum_{i=1}^{N} r_{ij}}}$$

Sum of points in cluster j / # of points in cluster j

= Mean of all points in cluster j!

ALGORITHM K-Means
INITIALISATION: Fix $\boldsymbol{\mu}_j$
1. [E-Step] $\{\boldsymbol{\mu}_j\}$ (centres) fixed, find $\{r_{ij}\}$ (assignment) Assign points to closest cluster prototype ($J \downarrow$)
1. [M-Step] $\{r_{ij}\}$ (assignment) fixed, find $\{\boldsymbol{\mu}_j\}$ (centres) Recompute cluster centres ($J \downarrow$)
REPEAT till no change in assign't/max iterations

Illustration of K-Means

[C. M. Bishop, Pattern Recognition and Machine Learning]