Unsupervised clustering with growing self-organizing neural network

A comparison with non-neural approach

Martin Hynar, Michal Burda, Jana Šarmanová

Department of Computer Science, VŠB – Technical University of Ostrava,
17. listopadu 15, 708 00, Ostrava – Poruba, Czech Republic

Dateso 2005

Outline

- K-means based methods
- CLASS method - flexible k-means
- Self-Organizing Map
- Growing Neural Gas
- Examples and comparison
- Conclusions

Introduction

- k-means method belongs to the most used ones in DM.
- It must be given the number of expected clusters.
- What to do if it could not be determined?

1. Make multiple computations with varying settings.
2. Adapt the algorithm to determine the count of clusters by itself.

The goal

- Describe the "classical" approach of determining clusters using k-means based methods.
- Describe the solution using self-organizing neural network.
- Compare both approaches.

K-means based methods

Phases

1. *Choose typical points.*
2. *Clustering.*
3. *Recompute typical points.*
4. *Check termination condition.*

CLASS method - flexible k-means

- Tries to determine number of clusters on-line.
- During the clustering process it performs splitting of large clusters.
- The very first step is one k-means clustering iteration. It divide patterns into base clustering.
- Each iteration starts with exclusion of small clusters
- Excessively variable clusters are dispersed.

CLASS method - phases

Phases

1. *Excluding small clusters.*
2. *Splitting clusters.*
3. *Revoking clusters.*

CLASS method - splitting clusters

- The splitting threshold is determined with equation

$$S_m = S_{m-1} + \frac{1 - S_0}{GAMA}$$

- Then for each cluster two average deviations are computed
  - for points on the left side of the typical point
  - for the right side

$$D_{jc} = \frac{1}{k_c} \sum_{i=1}^{k_c} d_{ij} \quad c \in \{l, r\}$$

- Using these deviations we compute splitting control parameters a_1 and a_2 (relative ratios). If then:
  - Number of clusters $> 2K$
  - $a_1 > S_m$ or $a_2 > S_m$
  - Number of processed patterns $> 2(THETAN + 1)$

  we split the cluster according to j^{th} attribute.

CLASS method - revoking clusters

- Determine average minimum distance of h current clusters

$$TAU = \frac{1}{h} \sum_{i=1}^{h} D_i$$

- D_i is the minimum distance of i^{th} typical point to others.
- If for some i holds $D_i < TAU$ and $h > \frac{K}{2}$ we revoke i^{th} cluster.
- The clustering ends in $GAMA^{th}$ iteration.

Self-Organizing Map

- A set $\mathcal{A}$ of neurons mutually interconnected, forming some topological grid.
- The pattern is presented to the net to determine the winner.

$$c = arg\,\min_{a \in \mathcal{A}} \{ || \vec{x} - \vec{w}_a || \}$$

- The weight vectors of the winner and its neighbours are adapted

$$w_{ji}(t+1) = \begin{cases} w_{ji}(t) + h_{cj}(t)(x_i(t) - w_{ji}(t)) & j \in N(c) \\ w_{ji}(t) & \text{otherwise.} \end{cases}$$

- The SOM network preserves topology so neurons are placed in the most dense regions.

Growing Neural Gas

- Introduced by Bernd Fritzke
- Motivation:
  - The net can have variable size.
  - Neurons are added and/or replaced according to proportions in the net.
  - Impermanent connections between neurons.
  - The resulting net could be in fact set of independent nets.

GNG - phases

Phases

1. *Competition.*
2. *Adaptation.*
3. *Removing.*
4. *Inserting new neurons.*
5. *Check termination condition.*

GNG - competition

- Determine the two nearest neurons s_1 and s_2 to the pattern $\vec{x}$.
- If does not exists add a connection between these two neurons.
  - The age of the connection is set to 0.
- The local error variable of the winner is increased by squared distance to the pattern.

$$\Delta E_{s_1} = ||\vec{x} - \vec{w}_{s_1}||^2$$

GNG - adaptation & removing

Adaptation

- Weight vectors of neuron s_1 and its topological neighbours are adapted by fractions ϵ_b and ϵ_n.

$$\begin{aligned} \Delta w_{s_1} &= \epsilon_b(\vec{x} - \vec{w}_{s_1}) \\ \Delta w_i &= \epsilon_n(\vec{x} - \vec{w}_i) \qquad \forall i \in N_{s_1} \end{aligned}$$

- The age of all winner's outgoing edges is increased by 1.

Removing

- All connections with age greater than age_{max} are removed.
- All standalone neurons are removed.

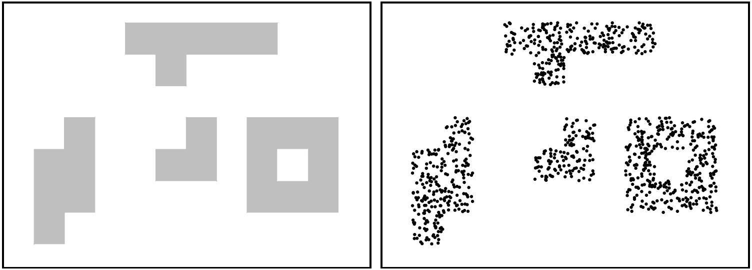
GNG - inserting new neurons

- New neurons are added every λ^{th} step using this procedure:

1. Determine neuron p with largest accumulated local error and its neighbour r with largest accumulated local error.
2. Create new neuron q and set its weight to the mean of p and r neurons weights.
3. Remove connection between p and r and add new between p and q and q and r.
4. Local accumulated errors of neurons p and r are decreased by fraction α and local accumulated error of neuron q is set to the mean of p and r errors.
5. Local accumulated errors of all other neurons are decreased by fraction β.

Examples and comparison - the basis

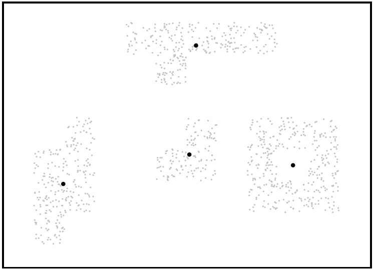
- The set of 1000 patterns with given distribution.
- The k-means and CLASS methods use discrete points, SOM and GNG use continuosly generated points from same distribution.



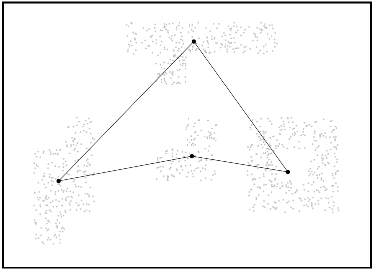
(a) Distribution (b) Objects from distribution

Examples and comparison - k-means and SOM

- Test if both methods will produce similar partitioning with number of units equal to number of clusters.



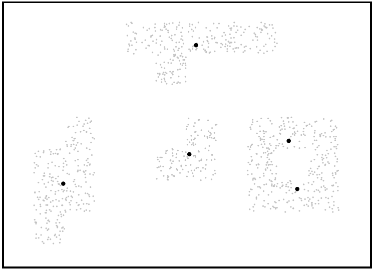
(c) k-means with $K = 4$



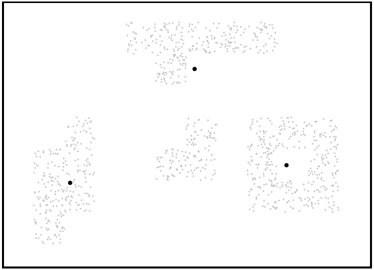
(d) SOM with 4 neurons

k-means and SOM

- The dangerous situation occurs when:
  - Number of representatives is very slightly higher or lower
- Result is hardly interpretable - i.e. typical points does not represent clusters.



(e) k-means with $K = 5$



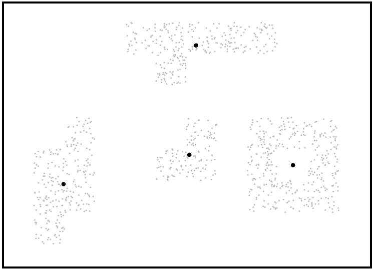
(f) k-means with $K = 3$

Examples and comparison - CLASS and GNG

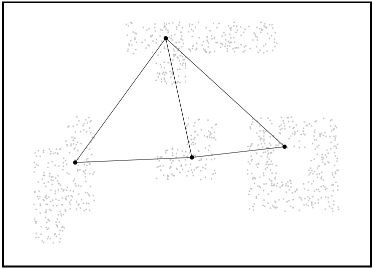
- Compare results reached with both methods.
- Both methods modify number of clusters using different approaches
  - compare them when they have identical cluster's count.
  - in the early iterations (few representatives) – 4
  - little more representatives – 9
  - enough representatives – 25

CLASS and GNG – 4 representatives

- Both results represent rough partitioning.
- Representatives are near centers - covering clusters as a whole.



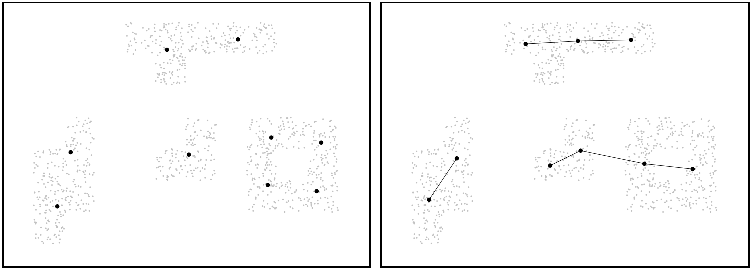
(g) CLASS



(h) GNG

CLASS and GNG – 9 representatives

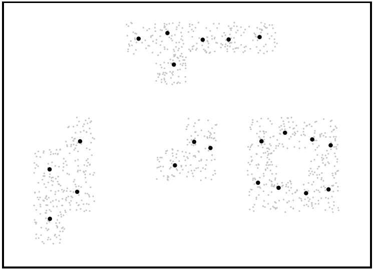
- More fine grained partitioning - an effort to cover smaller parts of clusters.
- GNG expresses the topology of clusters using connections.
- GNG's result could be interpreted as "three clusters", but ...



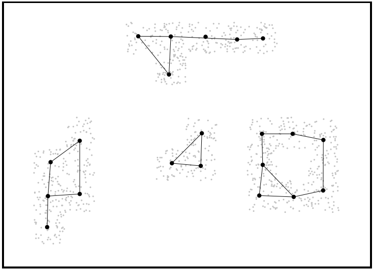
(i) CLASS (j) GNG

CLASS and GNG – 25 representatives

- Dislocation of representatives looks similar.
- GNG's result is nicely interpretable - 4 clusters with some topology.



(k) CLASS



(l) GNG

Conclusions

- Both approaches produce similar results.
- Suitable interpretation of connections could make results clearer.
- Good feature of GNG - set of independent sets of neurons.
  - Gives additional useful information.
  - Need to be interpreted with care.
- Situation in n-dimensional space - future work.

That's all, thank you for your attention.

Questions welcome.