A Family of Topology-preserving Mappings for Data Visualisation

Colin Fyfe

The University of Paisley,

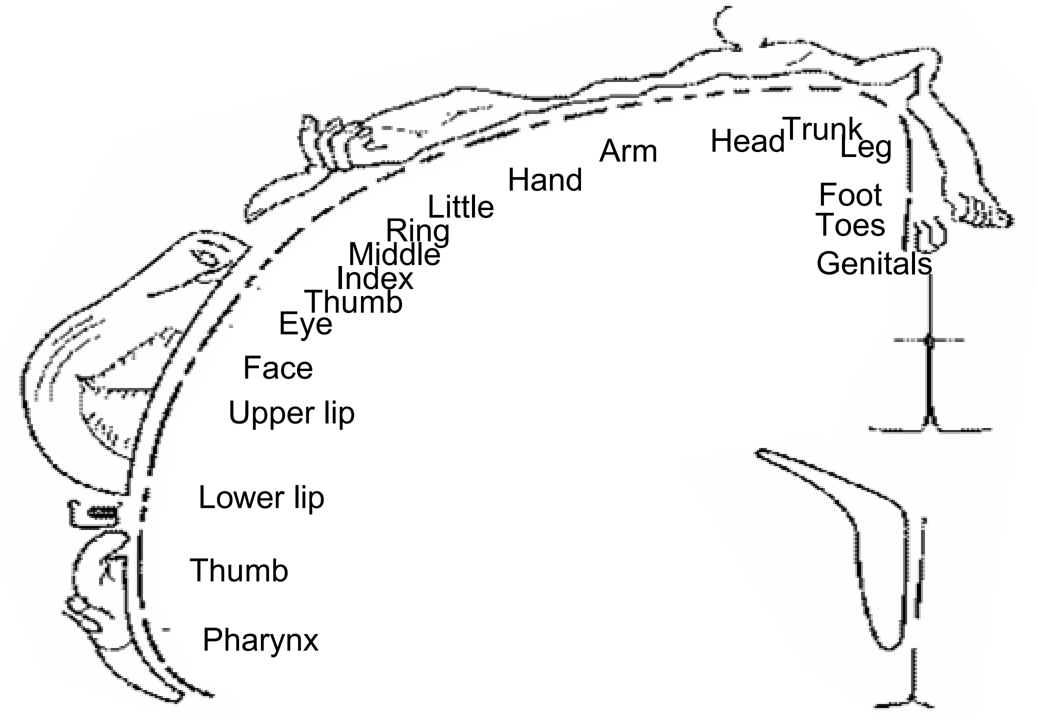
Scotland

Students: Marian Pena, Wesam Barbakh

Outline

- Topographic clustering.
- Topographic Product of Experts, ToPoE
- Simulations
- Products and mixtures of experts.
- Harmonic Topographic Mapping, HaToM
- 2 Varieties of HaToM
- IKToM

The somatosensory homunculus

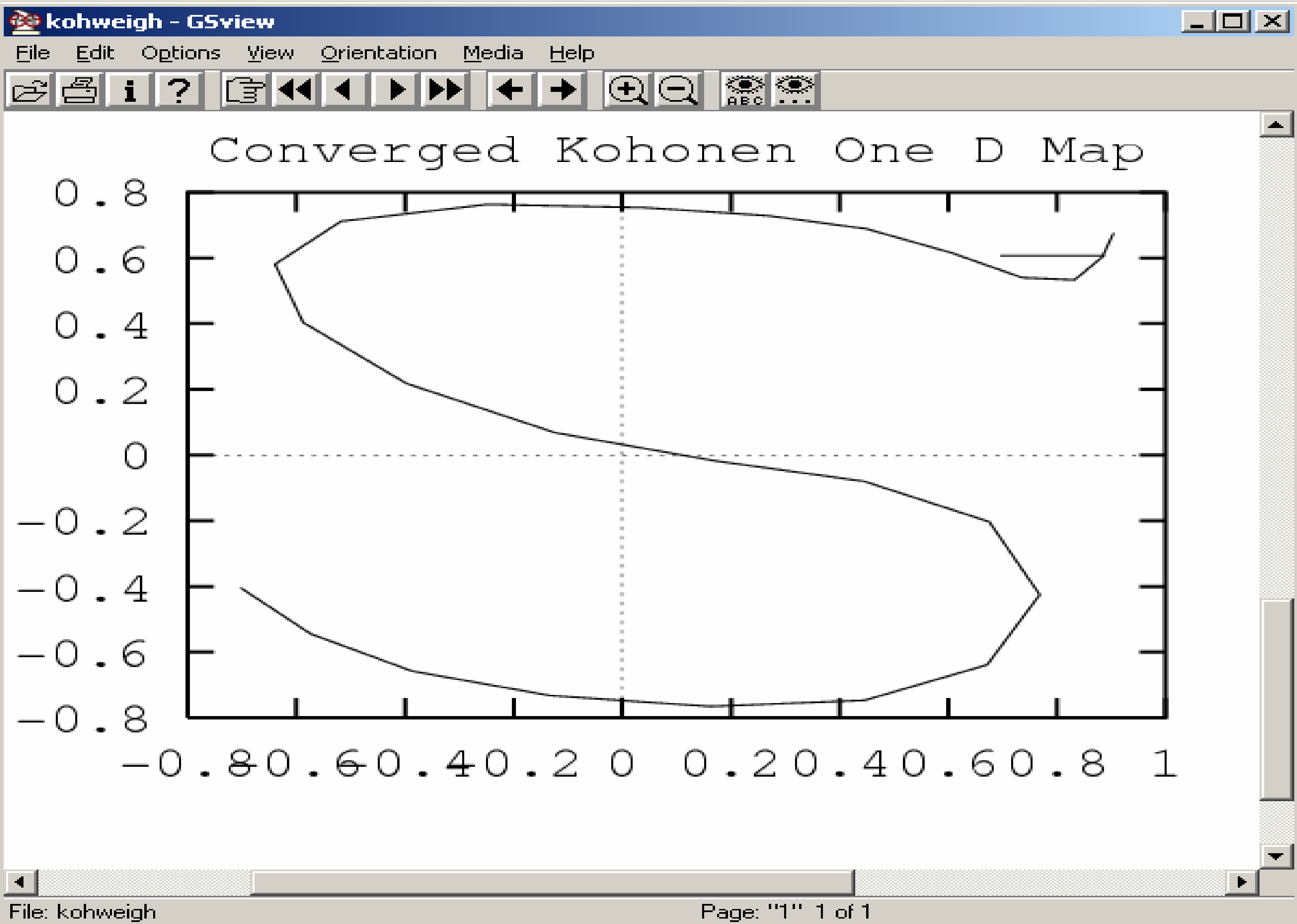


- Larger area of cortex for more sensitive body parts.

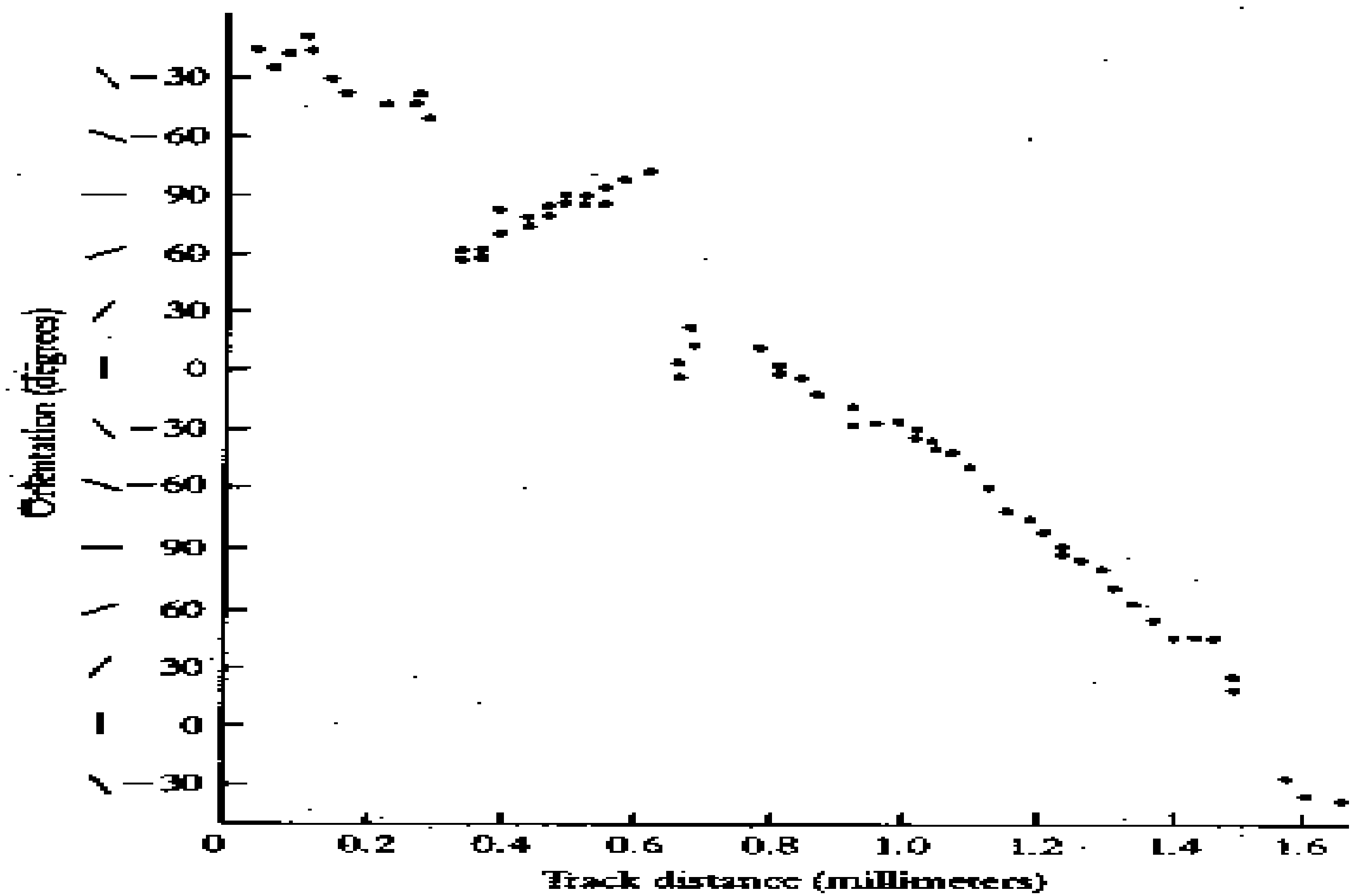
Topology Preservation and Data Visualisation

- **Data space** **Feature Space**

- Nearby → Nearby
- Distant → Distant (**)
- Nearby ← Nearby (**)
- Distant ← Distant


kohweigh - GSview
File Edit Options View Orientation Media Help
Converged Kohonen One D Map
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−0.8 0.6 0.4 0.2 0 0.2 0.4 0.6 0.8 1
File: kohweigh
Page: "1" 1 of 1

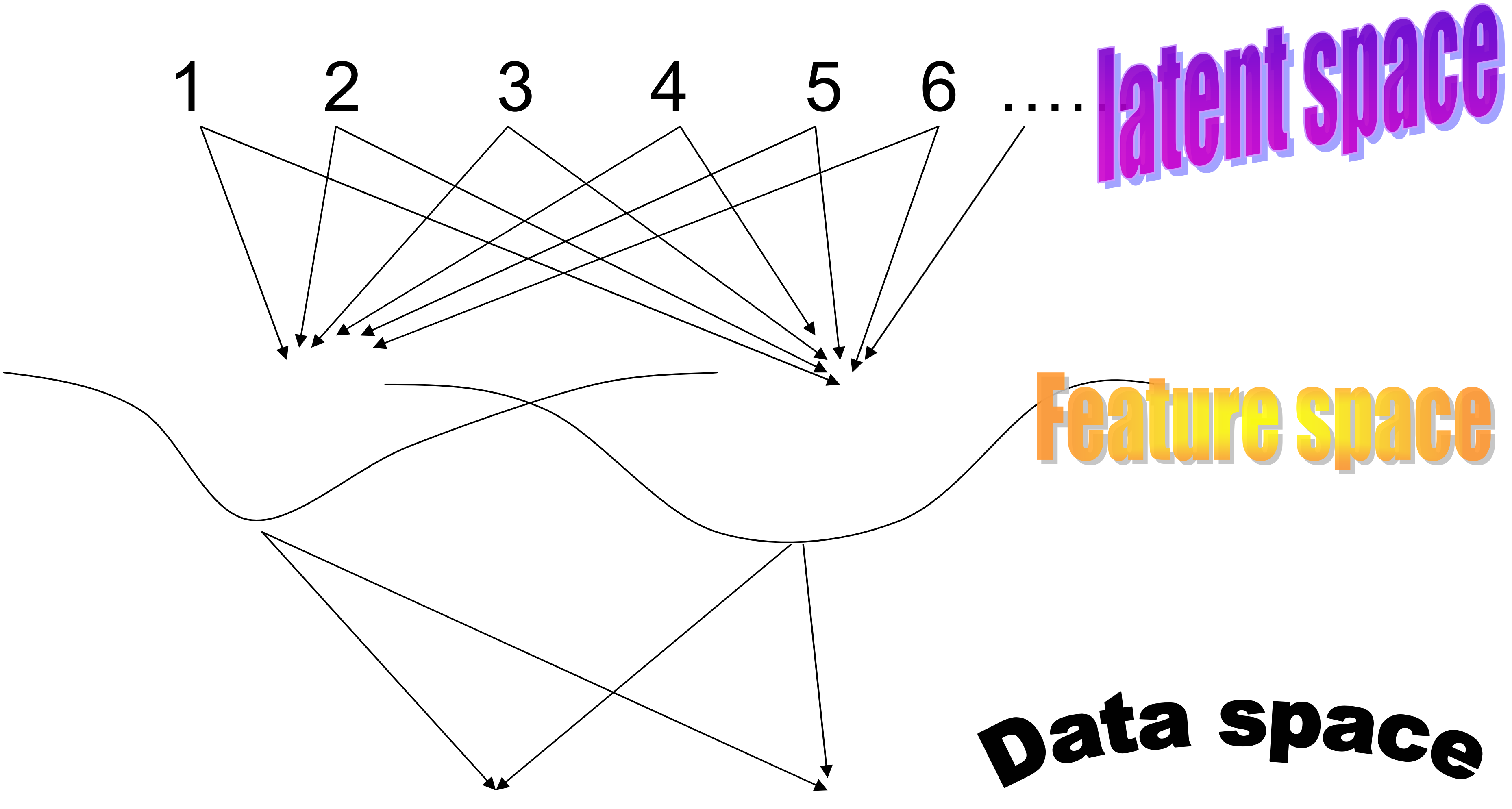
Orientation selectivity



The Model

- K latent points in a latent space with some structure.
- Each mapped through M basis functions to feature space.
- Then mapped to data space to K points in data space using W matrix (M by D)
- Aim is to fit model to data to make data as likely as possible by adjusting W

Mental Model



Details

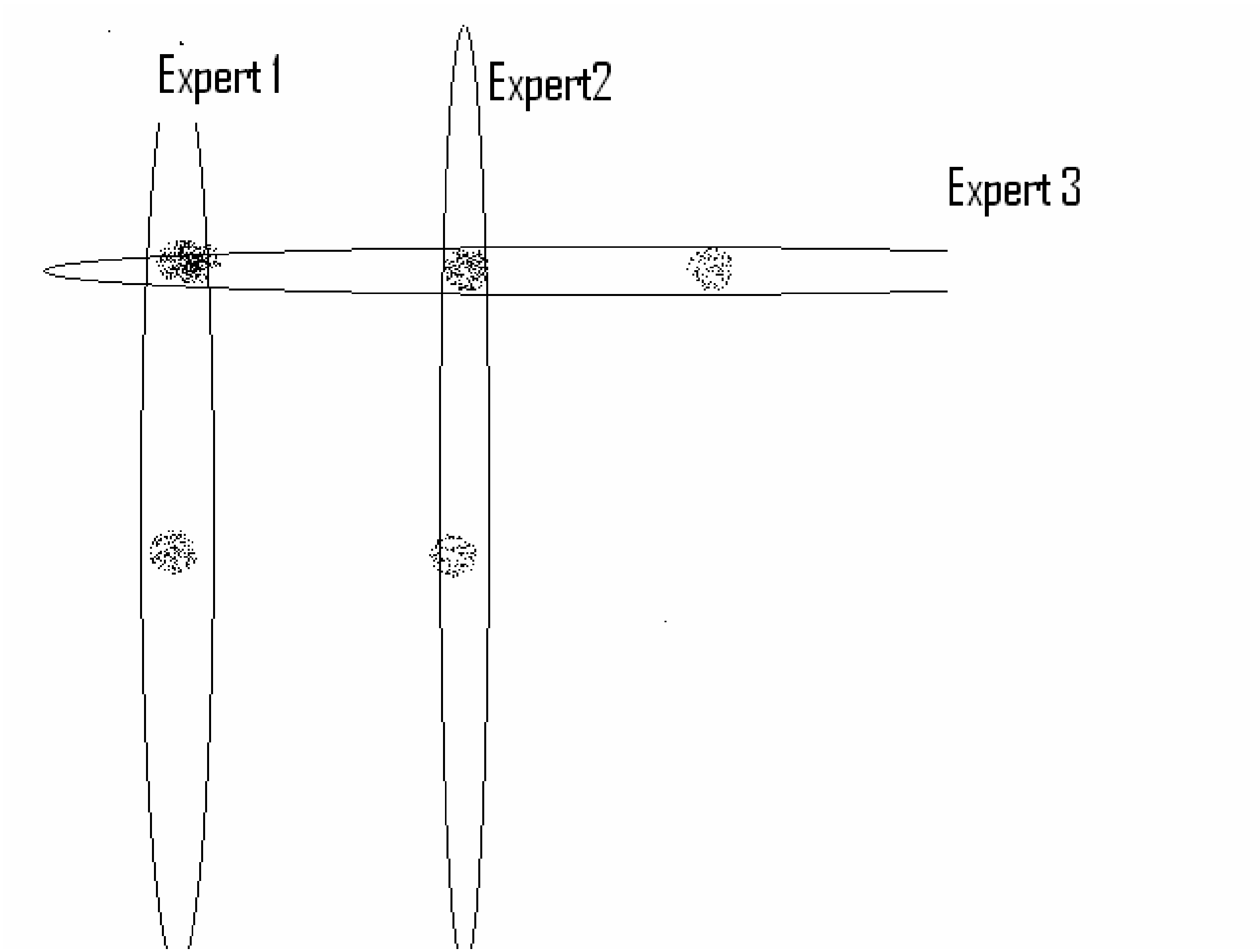
- $t_1, t_2, t_3, \ldots, t_K$ (points in latent space)
- $f_1(), f_2(), \ldots, f_M()$ (basis functions creating feature space)
- Matrix Φ (K by M), where $\varphi_{km} = f_m(t_k)$, projections of latent points to feature space.
- Matrix W (M by D) so that ΦW maps latent points to data space. $t_k \longrightarrow m_k$

Products of Gaussian Experts

$$p(x_n) \propto \prod_{k=1}^{K} \exp(-\frac{\beta}{2} \| m_k - x_n \|^2)$$

$$p(x_n) \propto \exp(-\frac{\beta}{2} \sum_{k=1}^{K} (\| m_k - x_n \|^2))$$

Products of Experts



Maximise the likelihood of the data under the model

$$-\log(p(x_n)) \propto \sum_{k=1}^{K} \left\| x_n - m_k \right\|^2$$

$$\Delta_n w_{md} = \eta \sum_{k=1}^{K} \phi_{km} (x_d^{(n)} - m_d^{(k)})$$

$$m_d^{(k)} = \sum_{m=1}^{M} w_{md} \phi_{km}$$

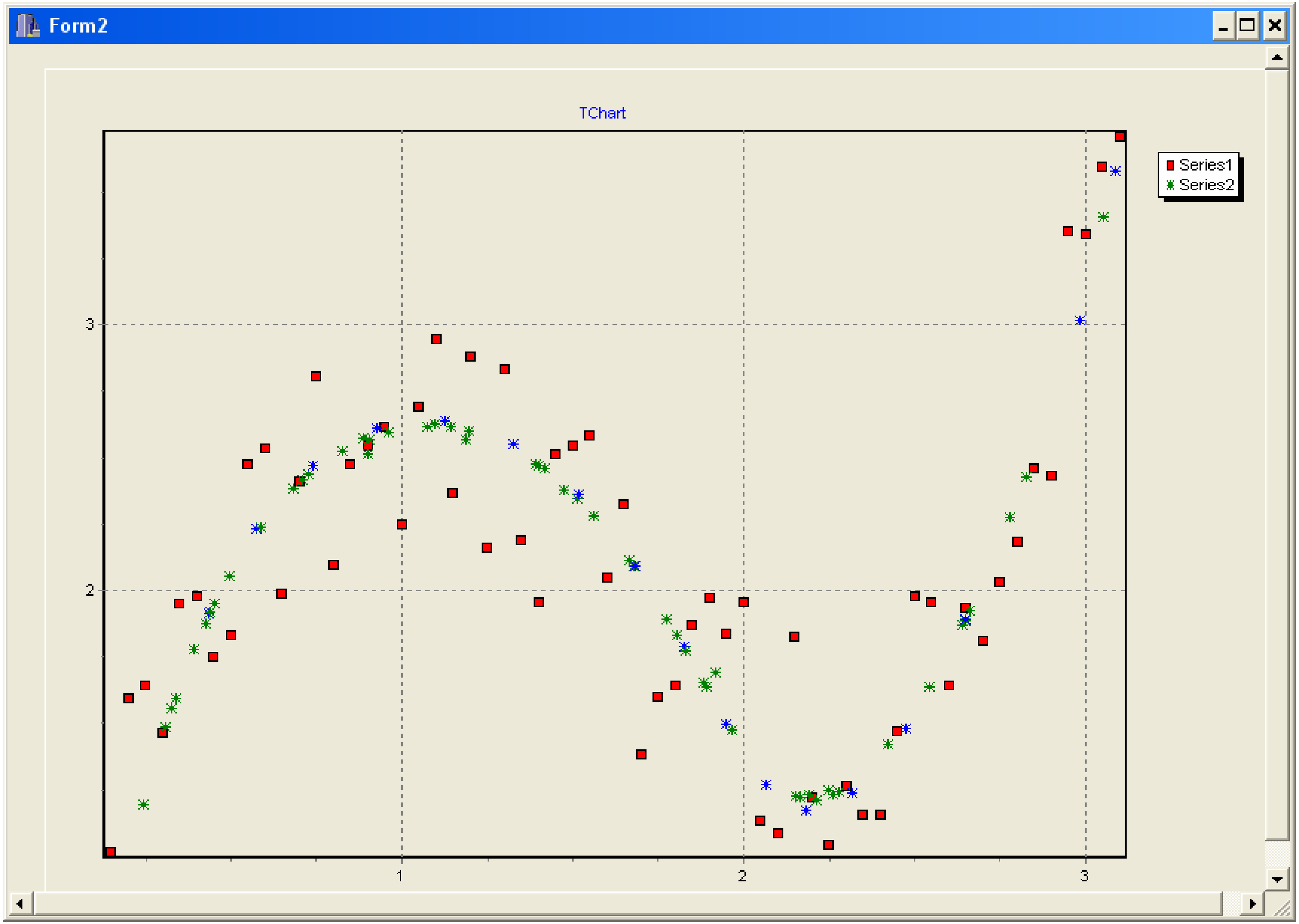
Using Responsibilities

$$\Delta_n w_{md} = \eta \sum_{k=1}^{K} \phi_{km} (x_d^{(n)} - m_d^{(k)}) r_{kn}$$

$$m_d^{(k)} = \sum_{m=1}^{M} w_{md} \phi_{km}$$

$$r_{kn} = \frac{\exp(-\gamma d_{kn}^2)}{\sum_j \exp(-\gamma d_{jn}^2)}$$

$$d_{jn} = \| x_n - \sum_{m=1}^{M} w_m \phi_{jm} \| = \| x_n - m_j \|$$


Form2
TChart
Series1
Series2
3
2
1
2
3

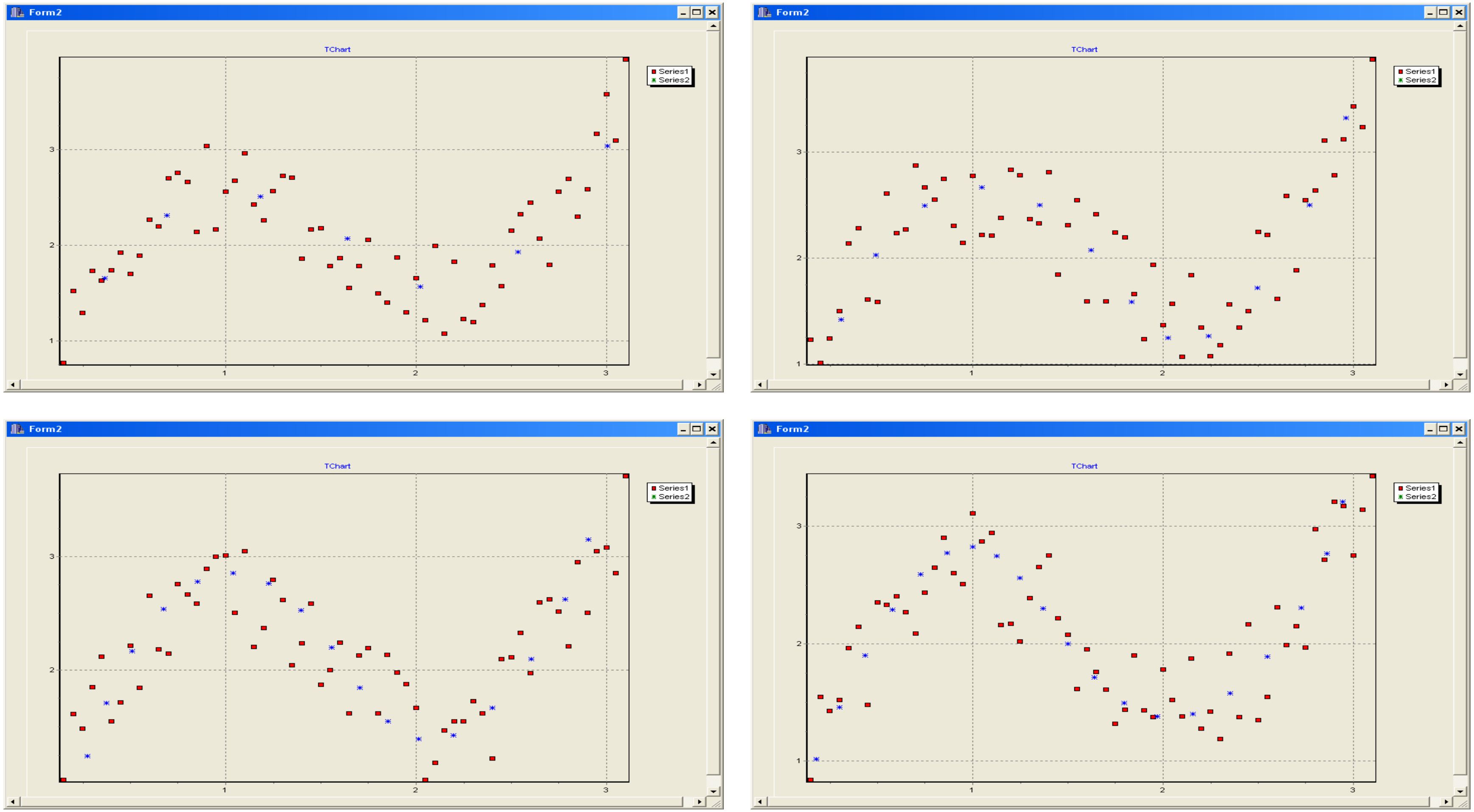
Comparison with GTM

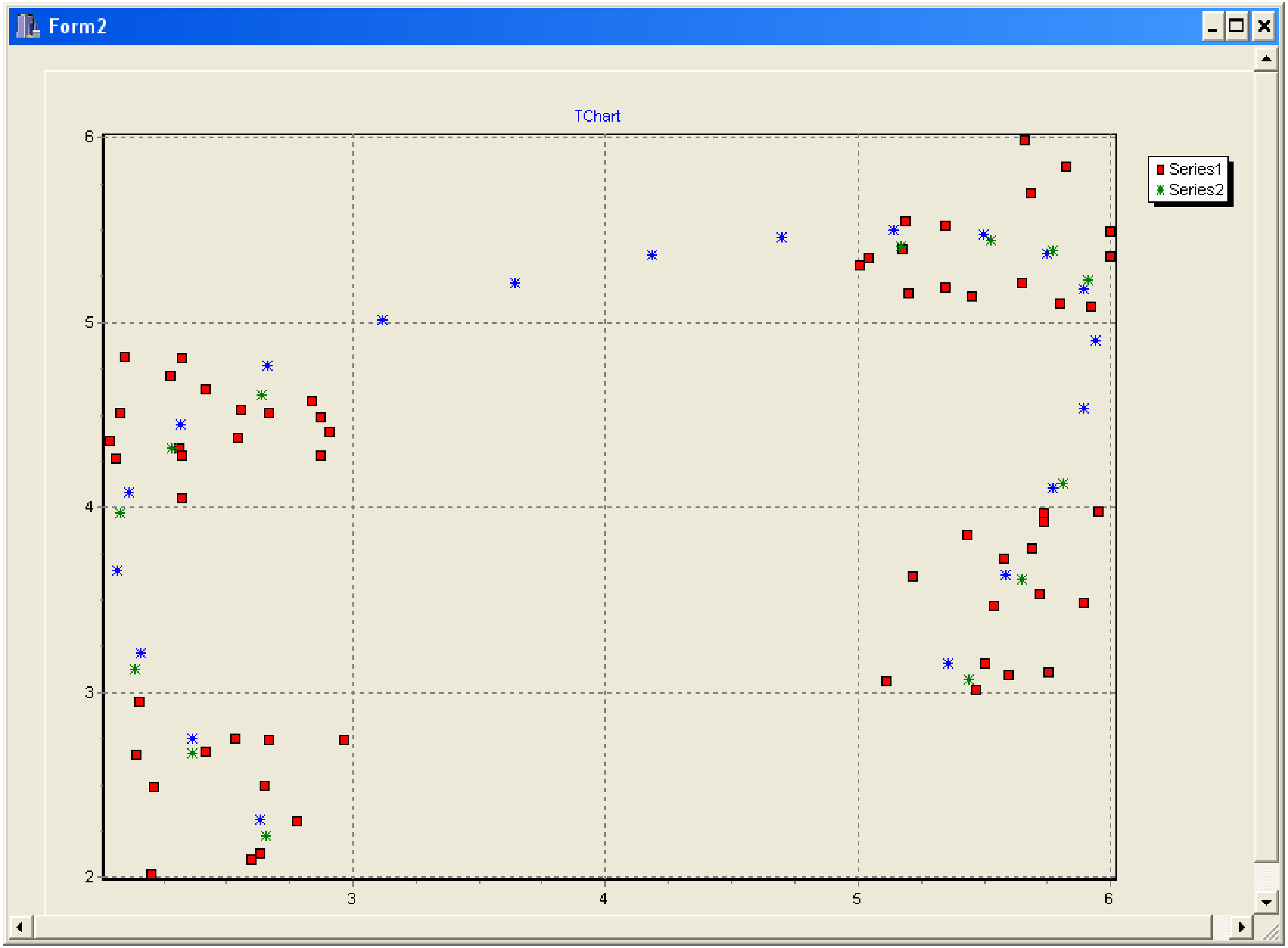
$$p(x_n) = \sum_k p(k)\,p(x_n \mid k) = \sum_k \frac{1}{K} \left(\frac{\beta}{2\pi}\right)^{\frac{D}{2}} \exp\left(-\frac{\beta}{2} \| x_n - m_k \|^2\right)$$

$$p(x_n) = \left(\frac{\beta}{2\pi}\right)^{\frac{D}{2}} \exp\left(-\frac{\beta}{2} \sum_{k=1}^{K} \left(\| m_k - x_n \|^2 \; r_{kn} \right)\right)$$

$$\beta_{k|x=x_n} = \beta r_{kn} = \beta \frac{\exp(-\gamma d_{nk}^2)}{\sum_t \exp(-\gamma d_{nt}^2)}$$

Growing ToPoEs

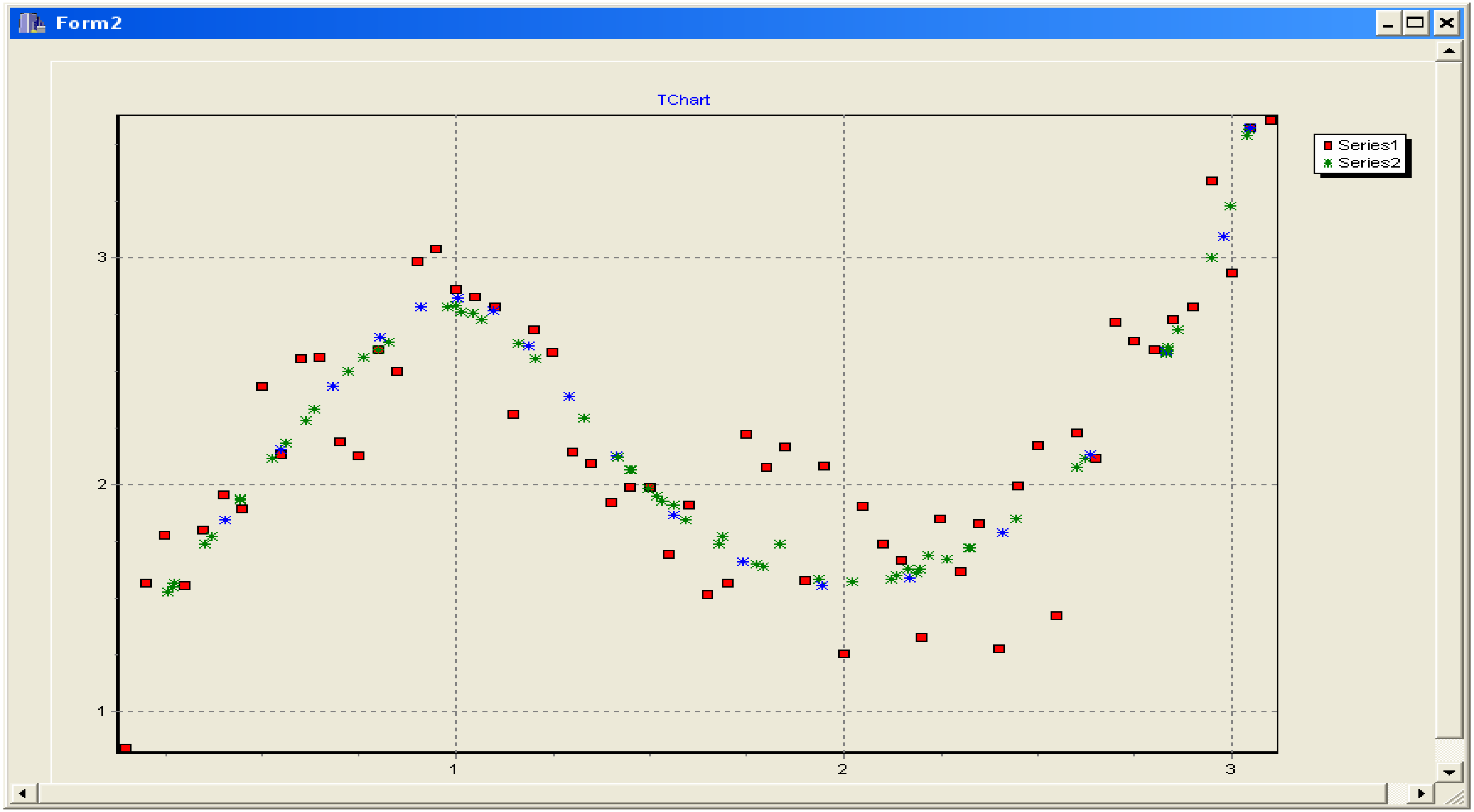



Form2
TChart
6
5
4
3
2
3
4
5
6
Series1
Series2

Advantages

- Growing : need only change Φ which goes from K by M to (K+1) by M.
- W is approximately correct and just refines its learning.
- Pruning uses the responsibility: if a latent point is never the most responsible point for any data point, remove it.
- Keep all other points at their positions in latent space and keep training.

$$p(x_n) = \frac{1}{Z}\exp(-\frac{\beta}{2}\sum_{k=1}^{K}(|m_k - x_n|_1 r_{kn})$$



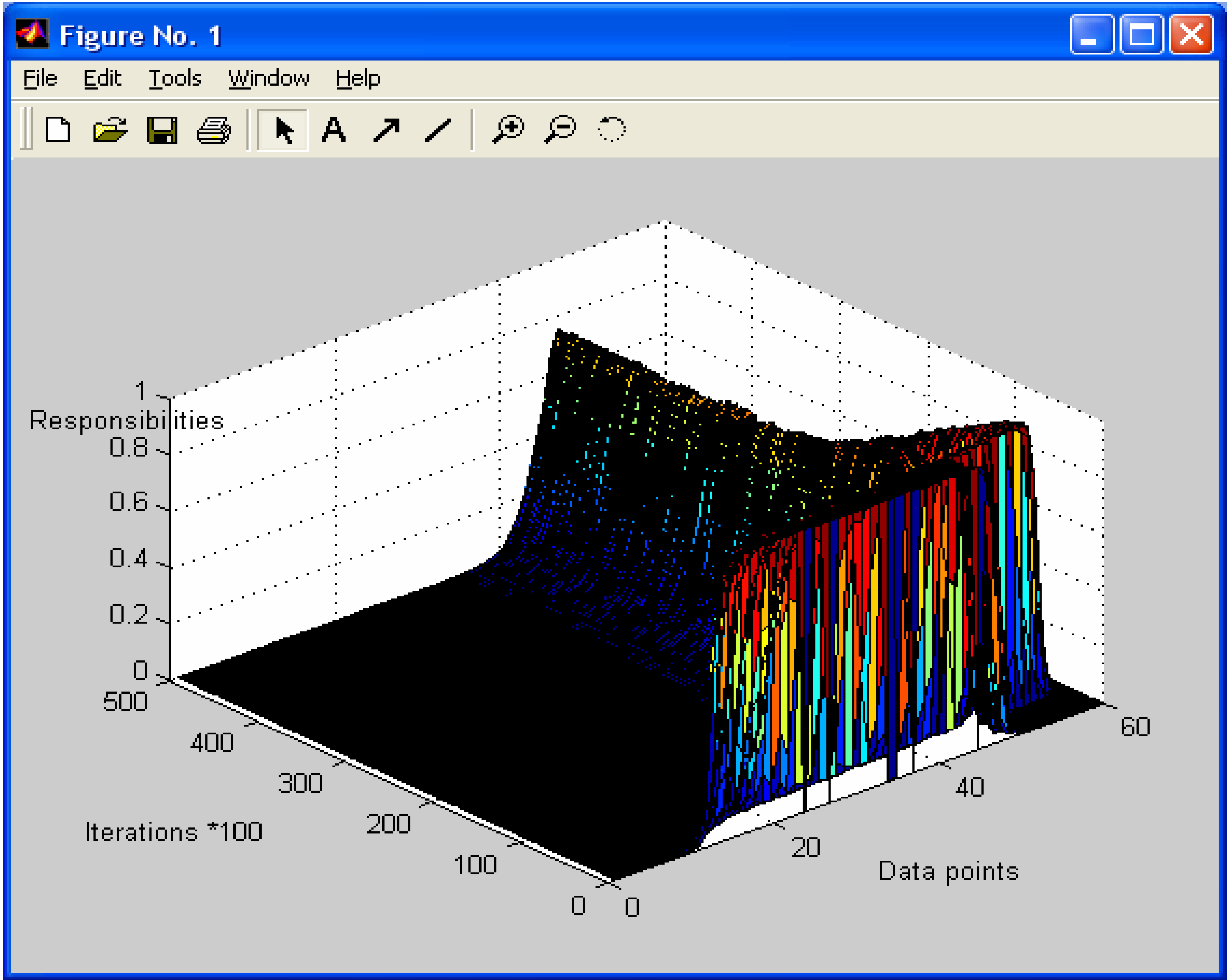


Figure No. 1
File Edit Tools Window Help
1
Responsibilities
0.8
0.6
0.4
0.2
0
500
400
300
200
100
0
Iterations *100
0
20
40
60
Data points

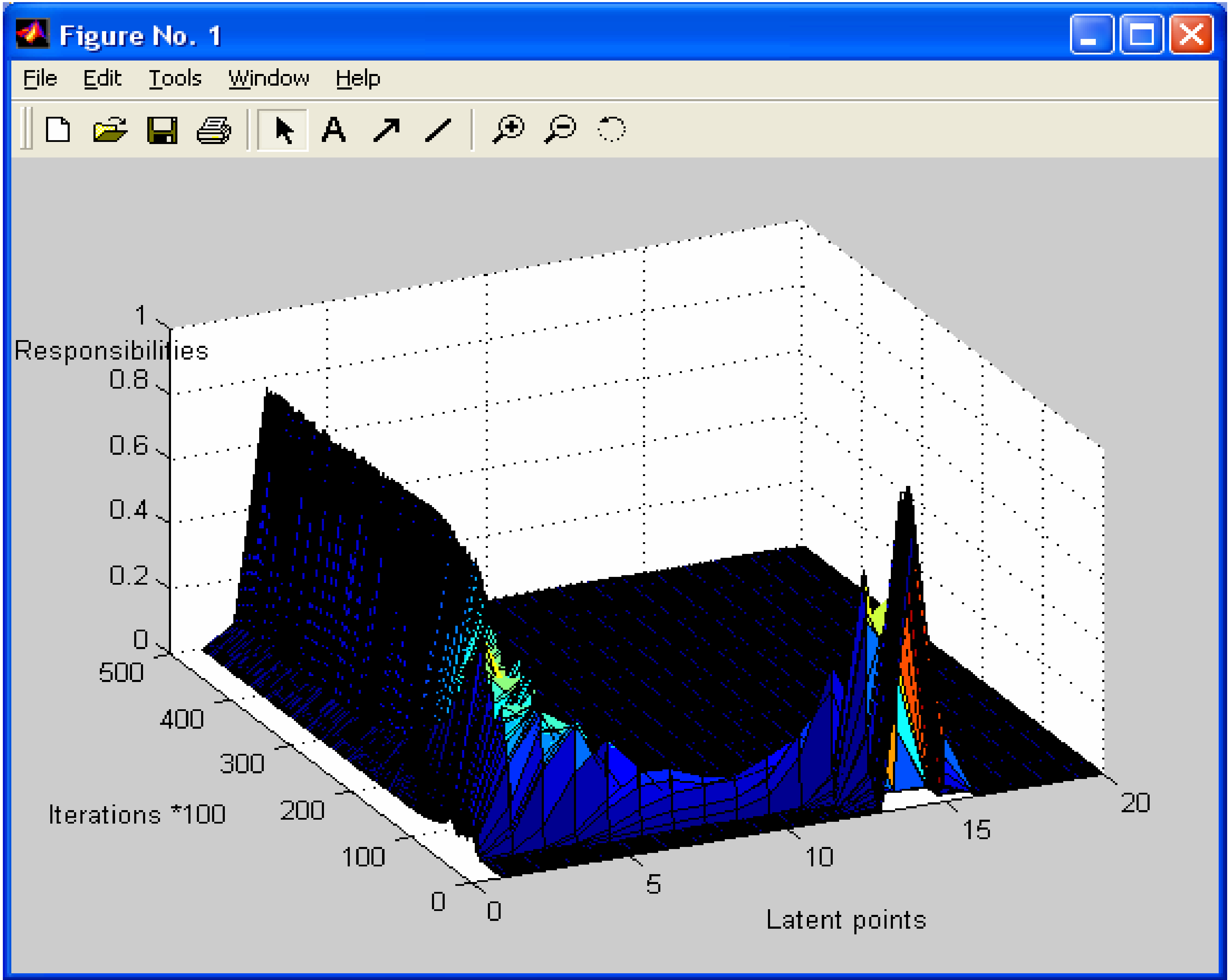
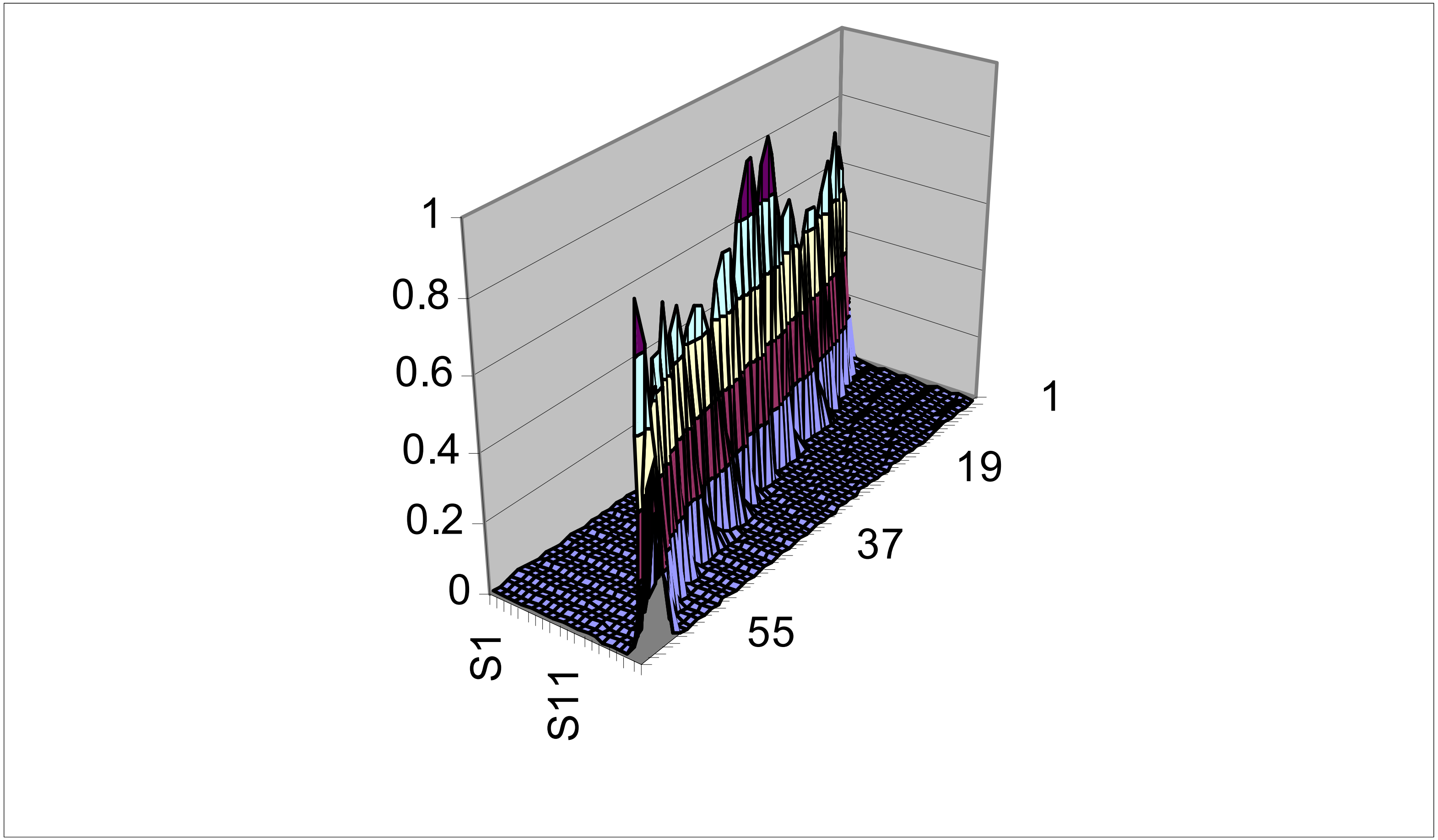


Figure No. 1
File Edit Tools Window Help
Responsibilities
1
0.8
0.6
0.4
0.2
0
500
400
300
200
100
0
Iterations *100
0
5
10
15
20
Latent points

Responsibilities with Tanh()



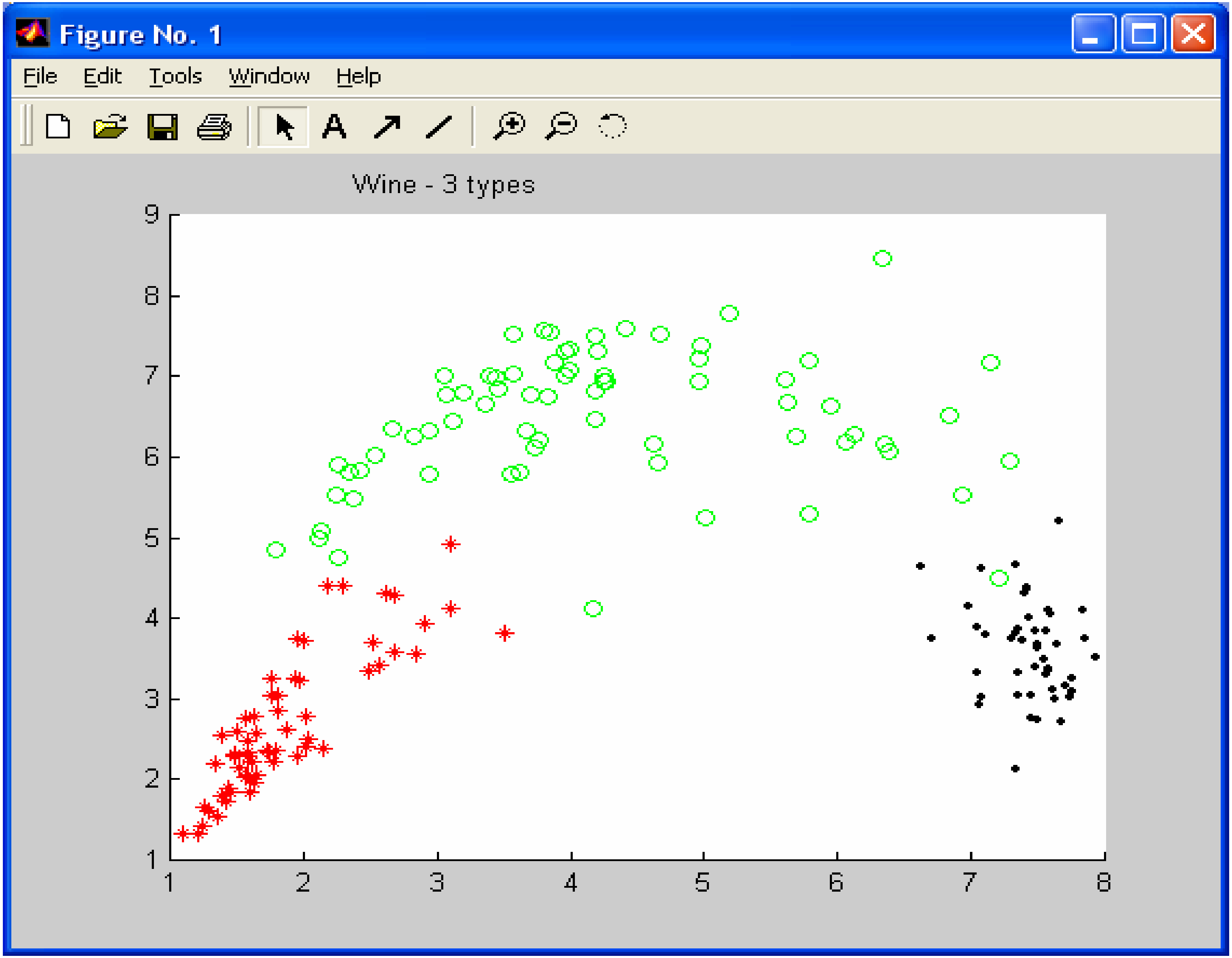
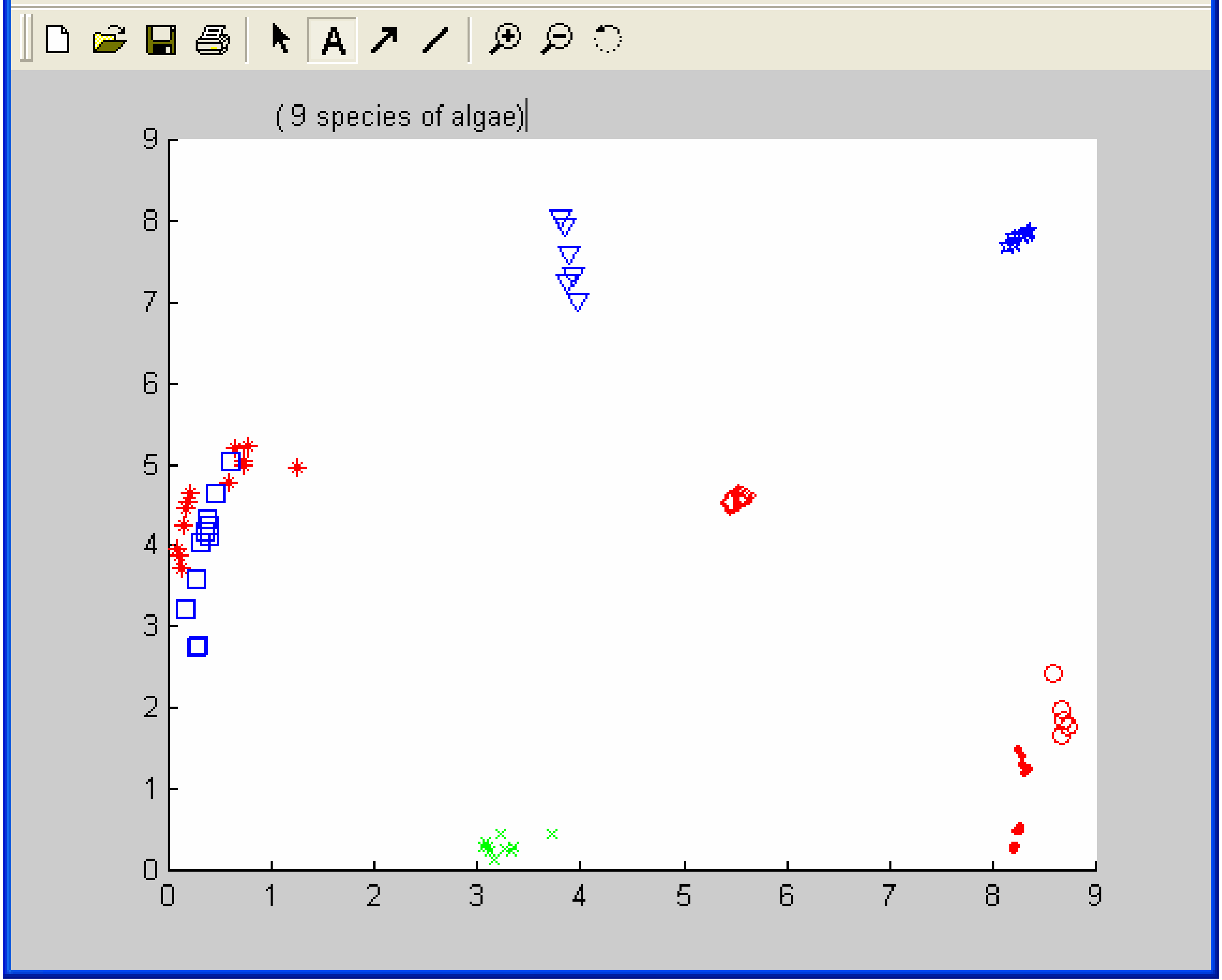


Figure No. 1
File Edit Tools Window Help
Wine - 3 types
9
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8



(9 species of algae)
0
1
2
3
4
5
6
7
8
9

A close up

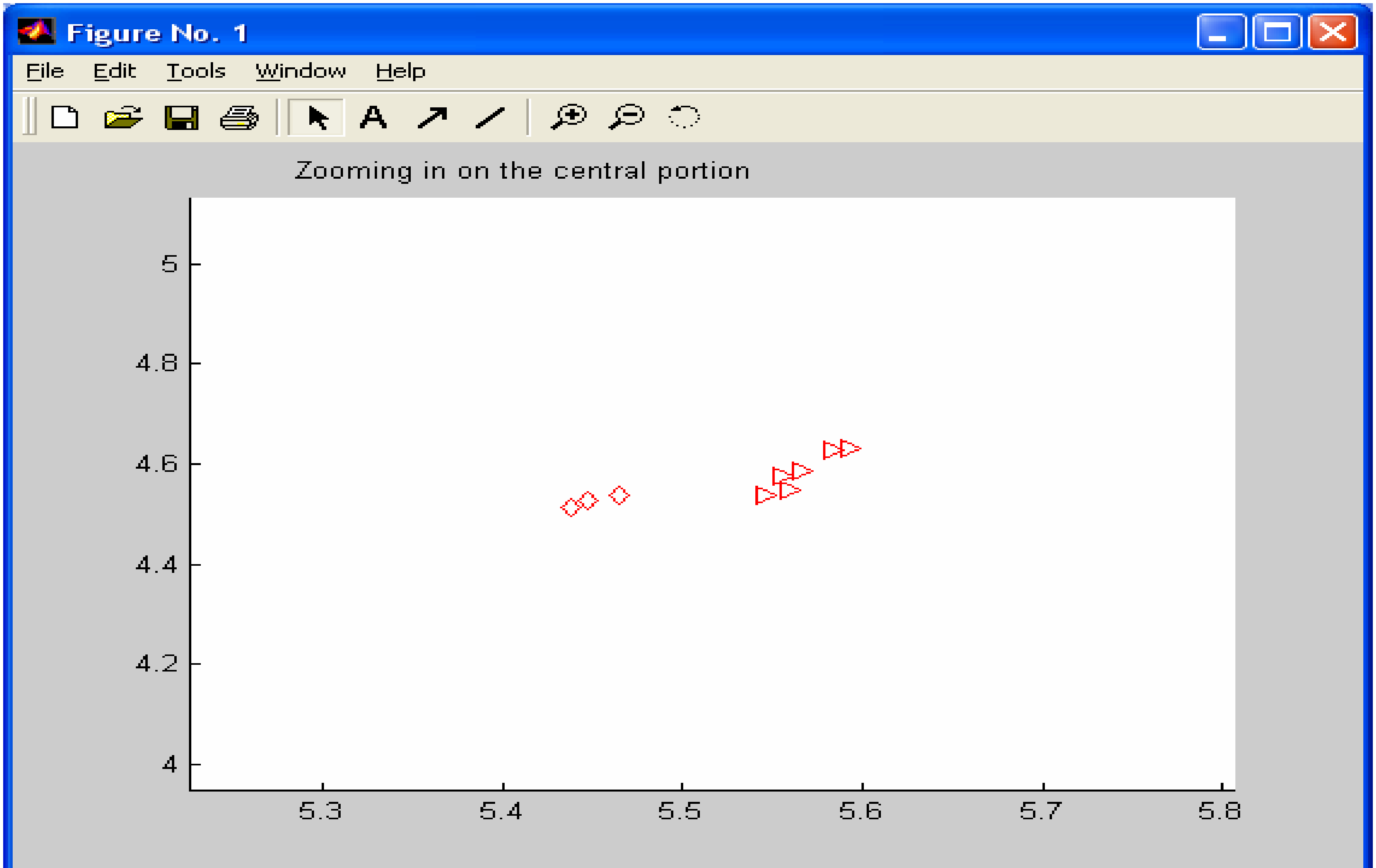


Figure No. 1
File Edit Tools Window Help
Zooming in on the central portion
5
4.8
4.6
4.4
4.2
4
5.3
5.4
5.5
5.6
5.7
5.8

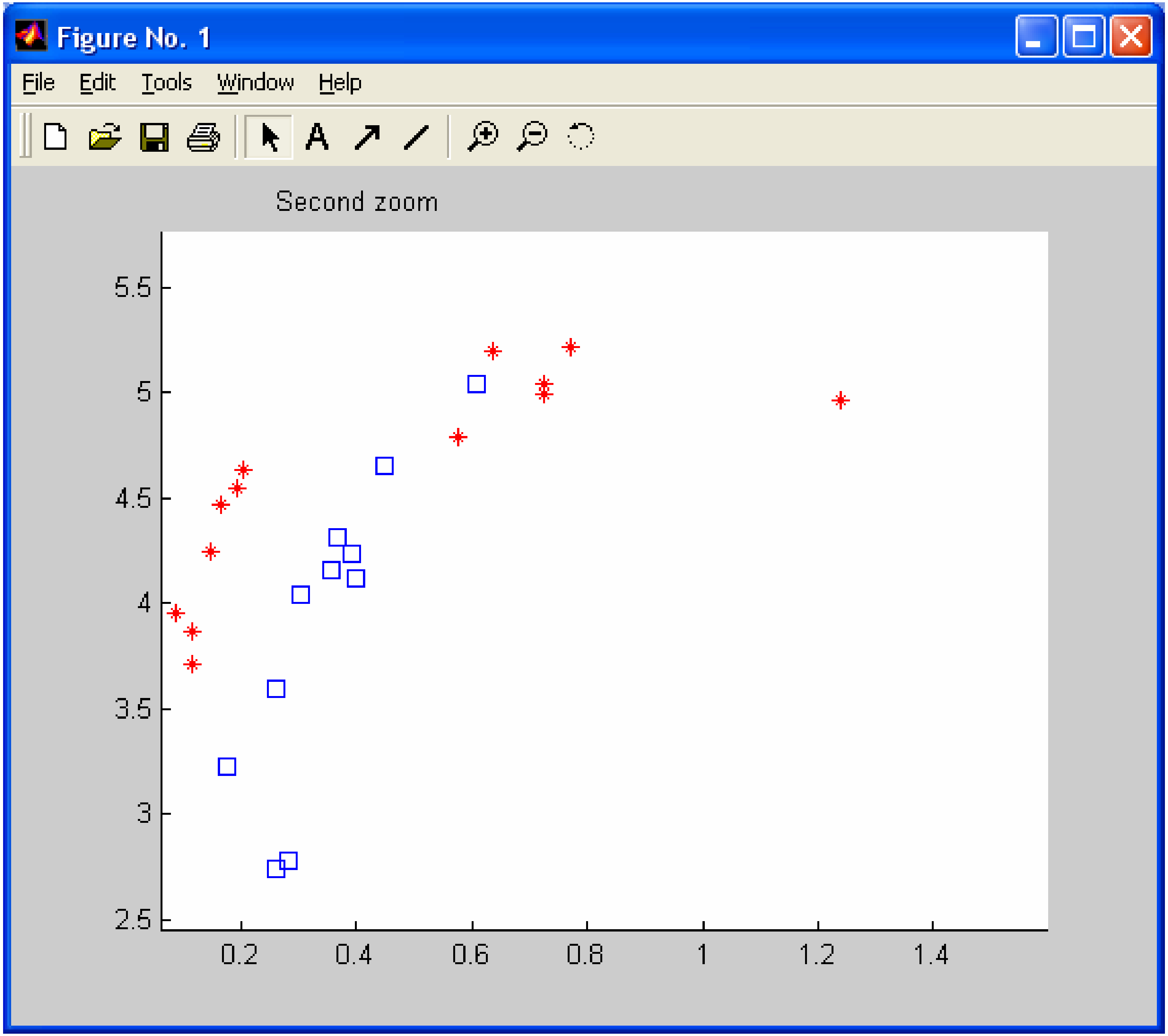


Figure No. 1
File Edit Tools Window Help
Second zoom
5.5
5
4.5
4
3.5
3
2.5
0.2
0.4
0.6
0.8
1
1.2
1.4

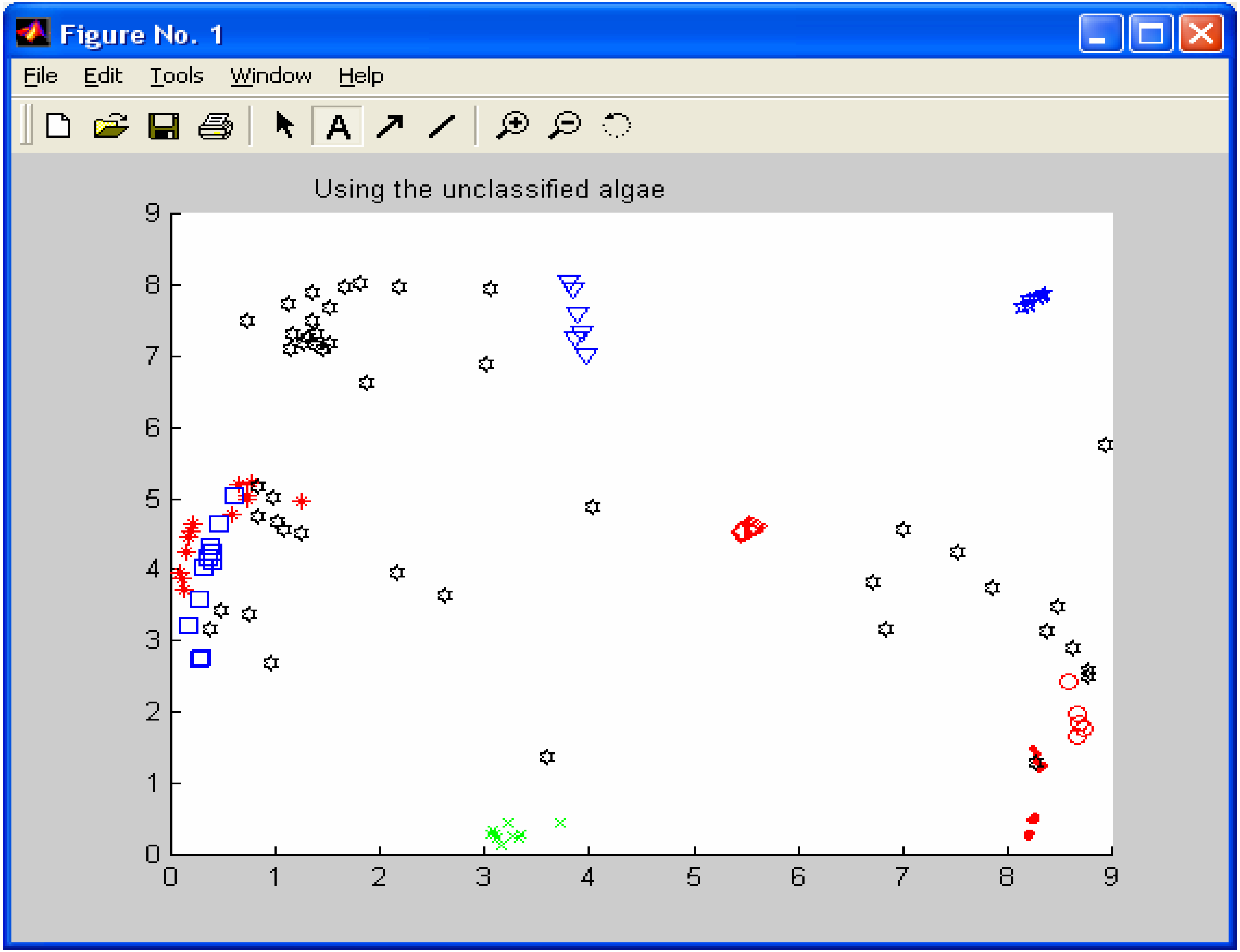


Figure No. 1
File Edit Tools Window Help
Using the unclassified algae
9
8
7
6
5
4
3
2
1
0
0 1 2 3 4 5 6 7 8 9

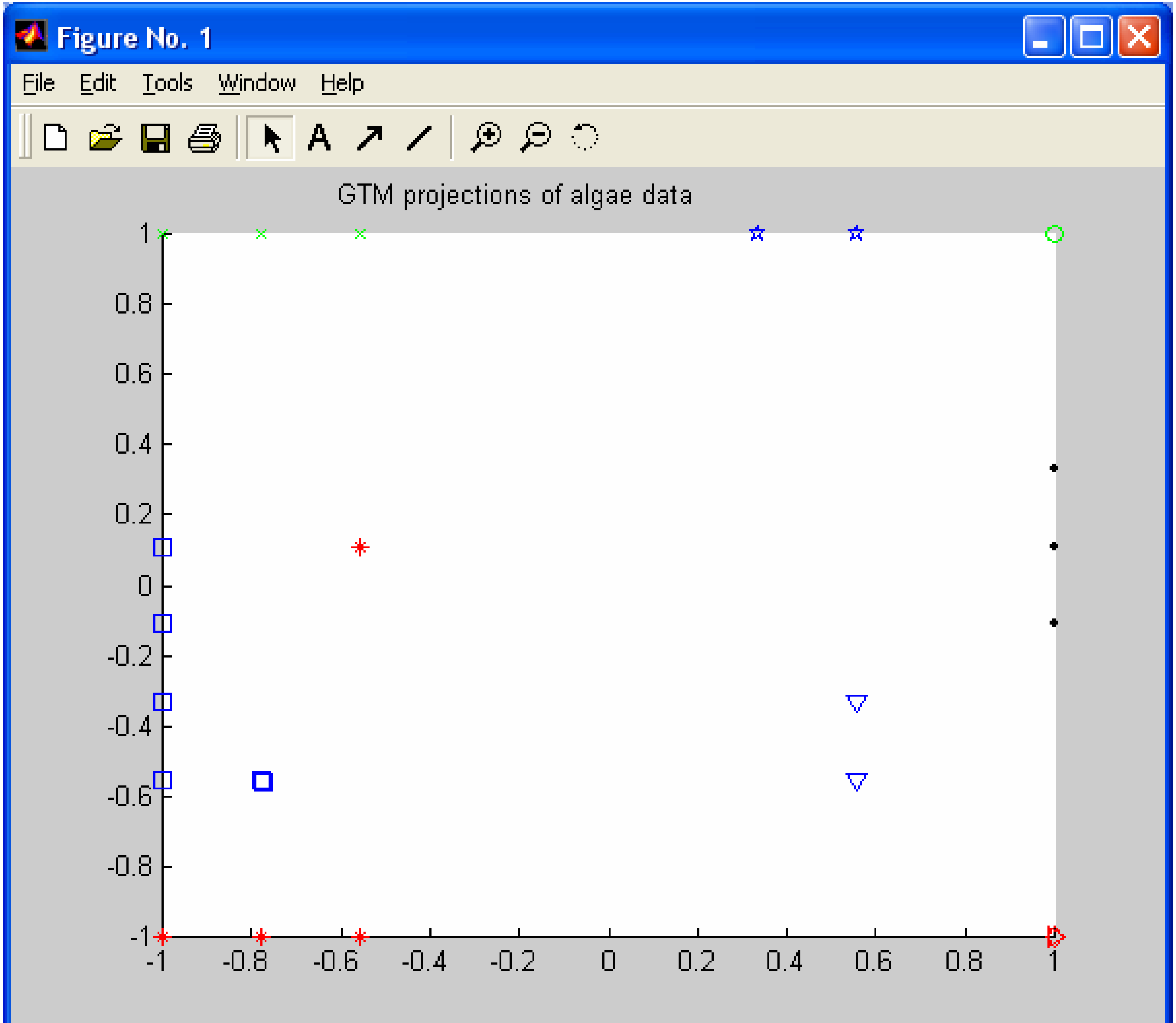


Figure No. 1
File Edit Tools Window Help
GTM projections of algae data
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

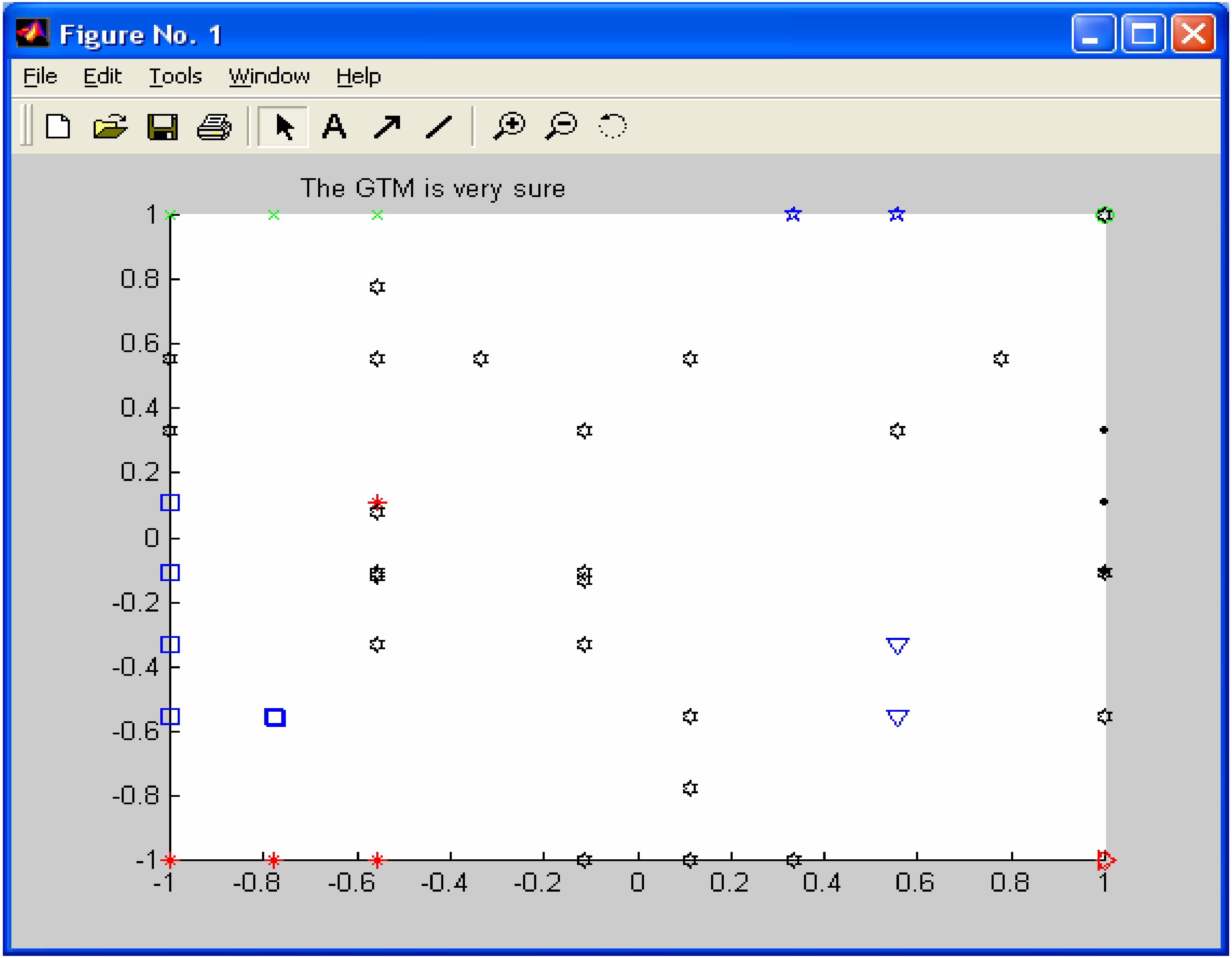


Figure No. 1
File Edit Tools Window Help
The GTM is very sure
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

Removing the Probabilistic Interpretation

$$J(x_n) = -\log(p(x_n)) \propto \sum_{k=1}^{K} \left\| x_n - m_k \right\|^2 r_{kn}$$

$$\Delta_n w_{md} \propto \frac{\partial J}{\partial w_{md}} = \eta \sum_{k=1}^{K} \phi_{km} (x_d^{(n)} - m_d^{(k)}) r_{kn}$$

$$m_d^{(k)} = \sum_{m=1}^{M} w_{md} \phi_{km}$$

Harmonic Averages

- Walk d km at 5 km/h, then d km at 10 km/h
- Total time = d/5 +d/10
- Average Speed = 2d/(d/5+d/10)= $\dfrac{2}{\frac{1}{5}+\frac{1}{10}}$

- Harmonic Average = $\dfrac{K}{\sum_{k=1}^{K} \frac{1}{a_k}}$

K-Harmonic Means beats K-Means and MoG using EM

- Perf = $$\sum_{i=1}^{N} \frac{K}{\sum_{k=1}^{K} \frac{1}{\left\| x_i - m_k \right\|^2}}$$

$$\frac{\partial Perf}{\partial m_k} = -K \sum_{i=1}^{N} \frac{4(x_i - m_k)}{d_{ik}^4 (\sum_{l=1}^{K} \frac{1}{d_{il}^2})^2}$$

$$m_k = \frac{\sum_{i=1}^{N} \frac{1}{d_{ik}^4 (\sum_{l=1}^{K} \frac{1}{d_{il}^2})^2} x_i}{\sum_{i=1}^{N} \frac{1}{d_{ik}^4 (\sum_{l=1}^{K} \frac{1}{d_{il}^2})^2}}$$

Growing Harmony Topology Preservation

- Initialise K to 2. Init W randomly.

1. Init K latent points and M basis functions.
2. Calculate $m_k = \varphi_k W$, $k=1,\ldots,K$.
   1. Calculate d_{ik}, $i=1,\ldots,N$, $k=1,\ldots,K$
   2. Re-calculate m_k, $k=1,\ldots,K$. (Harmonic alg.)
   3. If more, go back to 1.
3. Re-calculate $W=(\Phi^T\Phi+\gamma I)^{-1}\Phi^T Ш$
4. K= K + 1. If more, go back to 1.

Disadvantages-Advantages ?

- Don't have special rules for points for which no latent point takes responsibility.
- But must grow otherwise twists.
- Independent of initialisation ?
- Computational cost ?

Generalised K-Harmonic Means for Automatic Boosting

- Perf = $\sum_{i=1}^{N} \frac{K}{\sum_{k=1}^{K} \frac{1}{\left\| x_i - m_k \right\|^p}}$

$$\frac{\partial Perf}{\partial m_k} = -K \sum_{i=1}^{N} \frac{2p(x_i - m_k)}{d_{ik}^{p+1} (\sum_{l=1}^{K} \frac{1}{d_{il}^{p}})^2}$$

$$m_k = \frac{\sum_{i=1}^{N} \frac{1}{d_{ik}^{p+2} (\sum_{l=1}^{K} \frac{1}{d_{il}^{p}})^2} x_i}{\sum_{i=1}^{N} \frac{1}{d_{ik}^{p+2} (\sum_{l=1}^{K} \frac{1}{d_{il}^{p}})^2}}$$

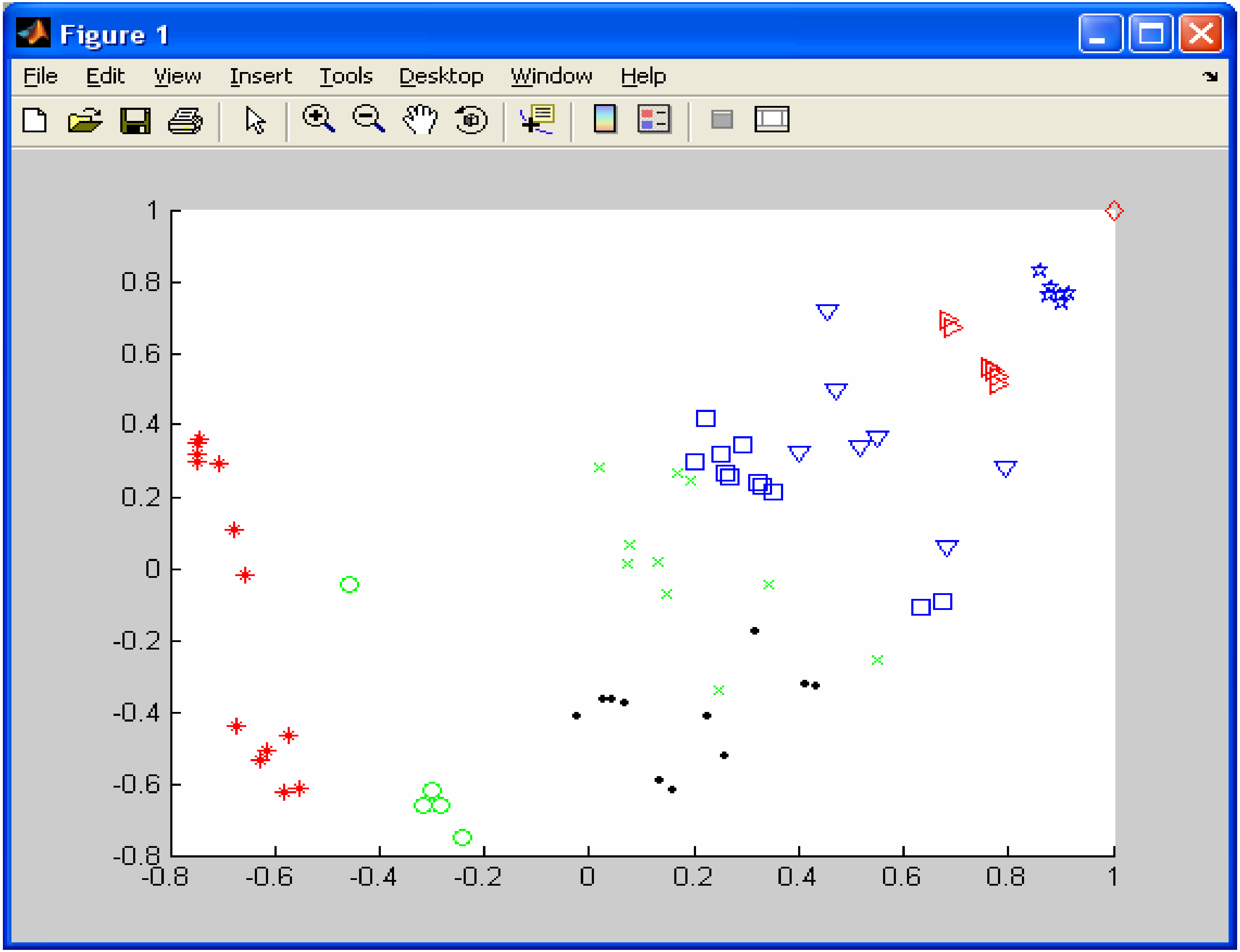


Figure 1
File Edit View Insert Tools Desktop Window Help
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

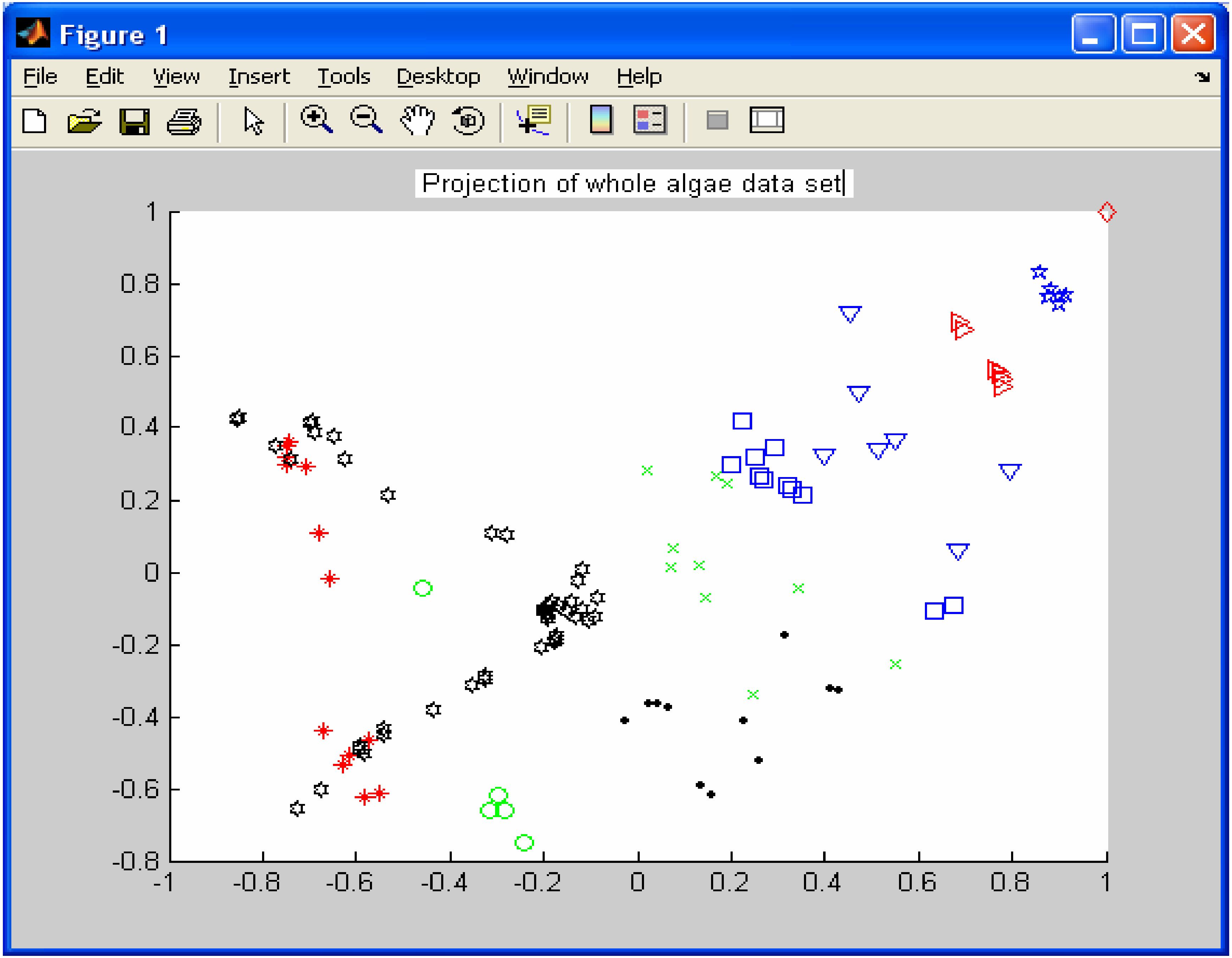
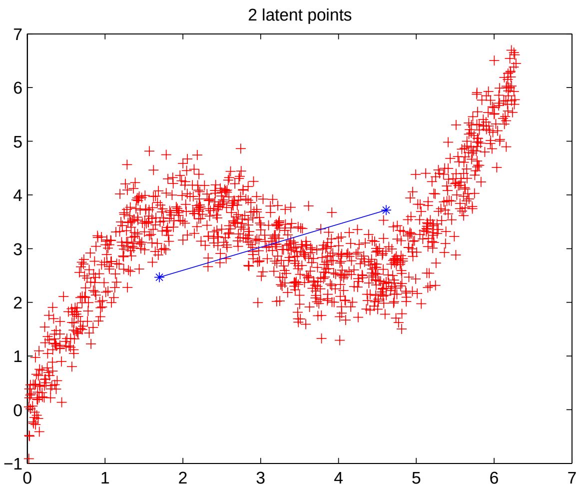


Figure 1
File Edit View Insert Tools Desktop Window Help
Projection of whole algae data set
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

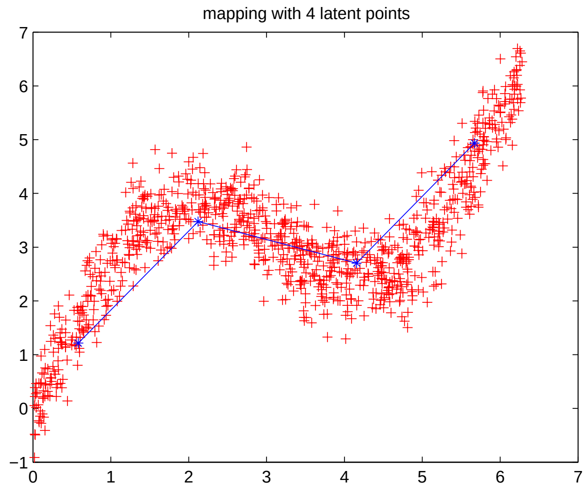
Two versions of HaToM

- D-HaToM (Data driven HaToM) :
  - W and m change only when adding a new latent point
  - Allows the data to influence more the clustering
- M-HaToM (Model driven HaToM) :
  - W and m change in every iteration
  - The data is continually constrained by the model

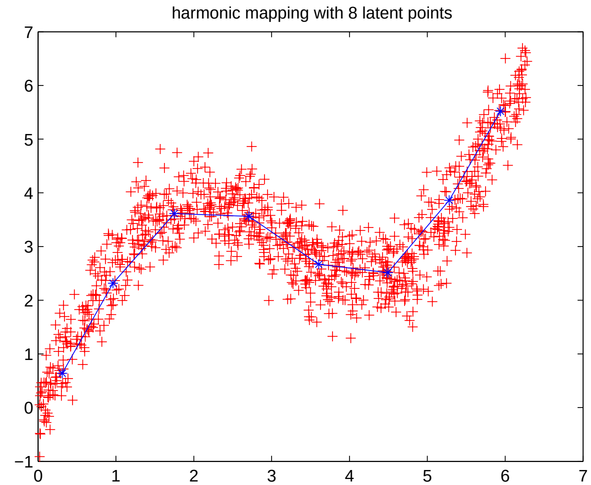
Simulations(1): 1D dataset



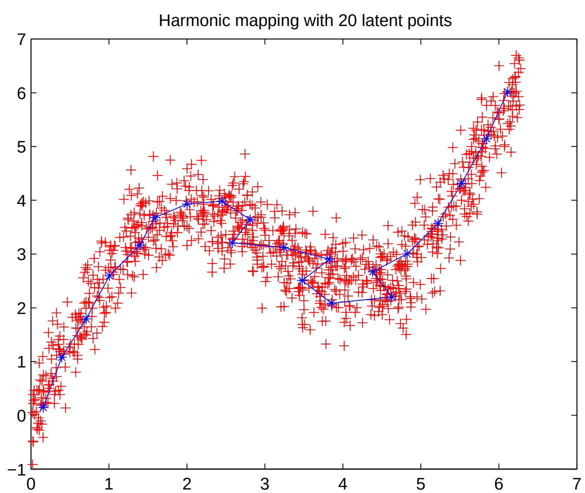
K=2



K=4

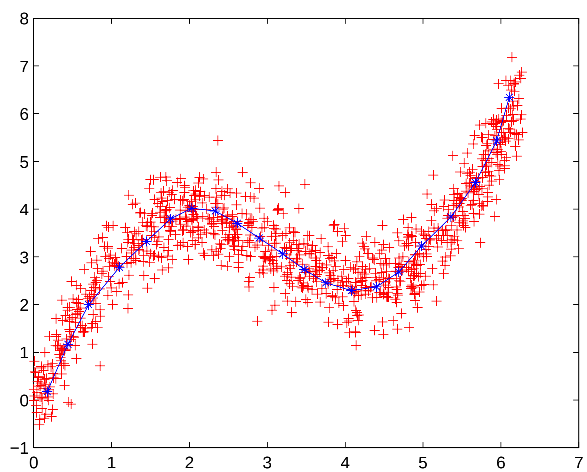


K=8



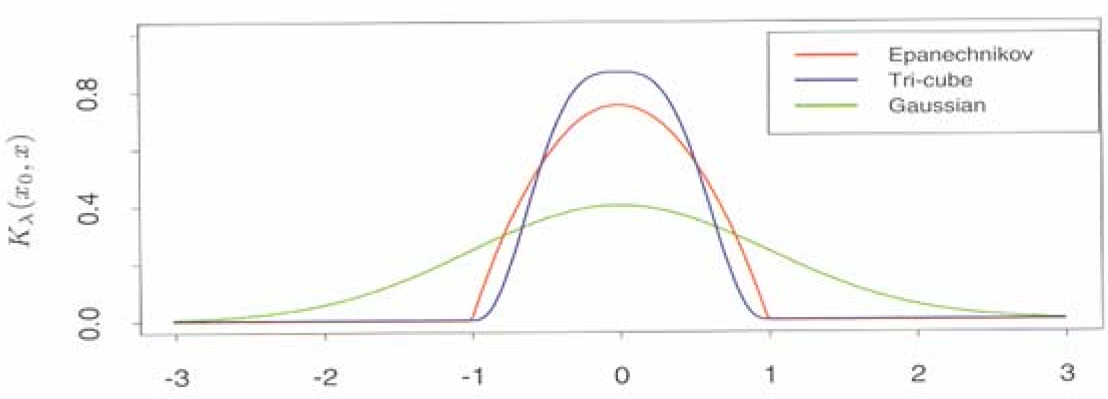
K=20

D-HaToM



M-HaToM

Kernels for the responsibilities



Epanechnikov

$$C_\lambda(k,n) = D\left(\frac{|\mathbf{x}_n - \mathbf{m}_k|}{\lambda}\right)$$

$$\text{where } D(t) = \begin{cases} \frac{3}{4}(1-t^2) & \text{if } |t| < 1 \\ 0 & \text{otherwise} \end{cases}$$

tri-cube function

$$D(t) = \begin{cases} (1-t^3)^3 & \text{if } |t| < 1 \\ 0 & \text{otherwise} \end{cases}$$

Performance Functions

- K-means

$$perf = \sum_{i=1}^{N} \min_{j=1}^{K} \left\| x_i - m_j \right\|^2$$

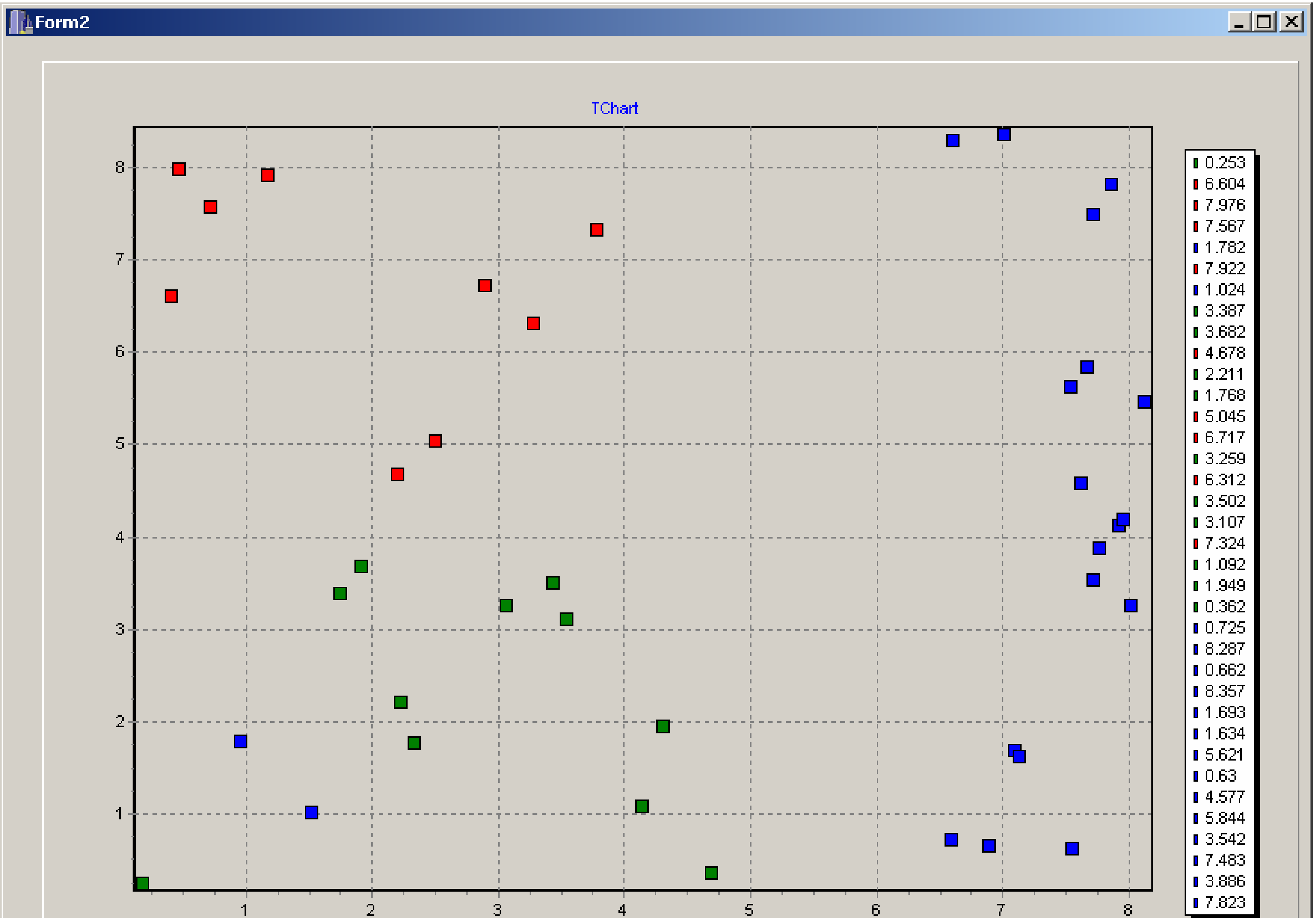
- Weighted K-means

$$perf = \sum_{i=1}^{N} \left[\sum_{j=1}^{K} \left\| x_i - m_j \right\| \right] \min_{l=1}^{K} \left\| x_i - m_l \right\|^2$$

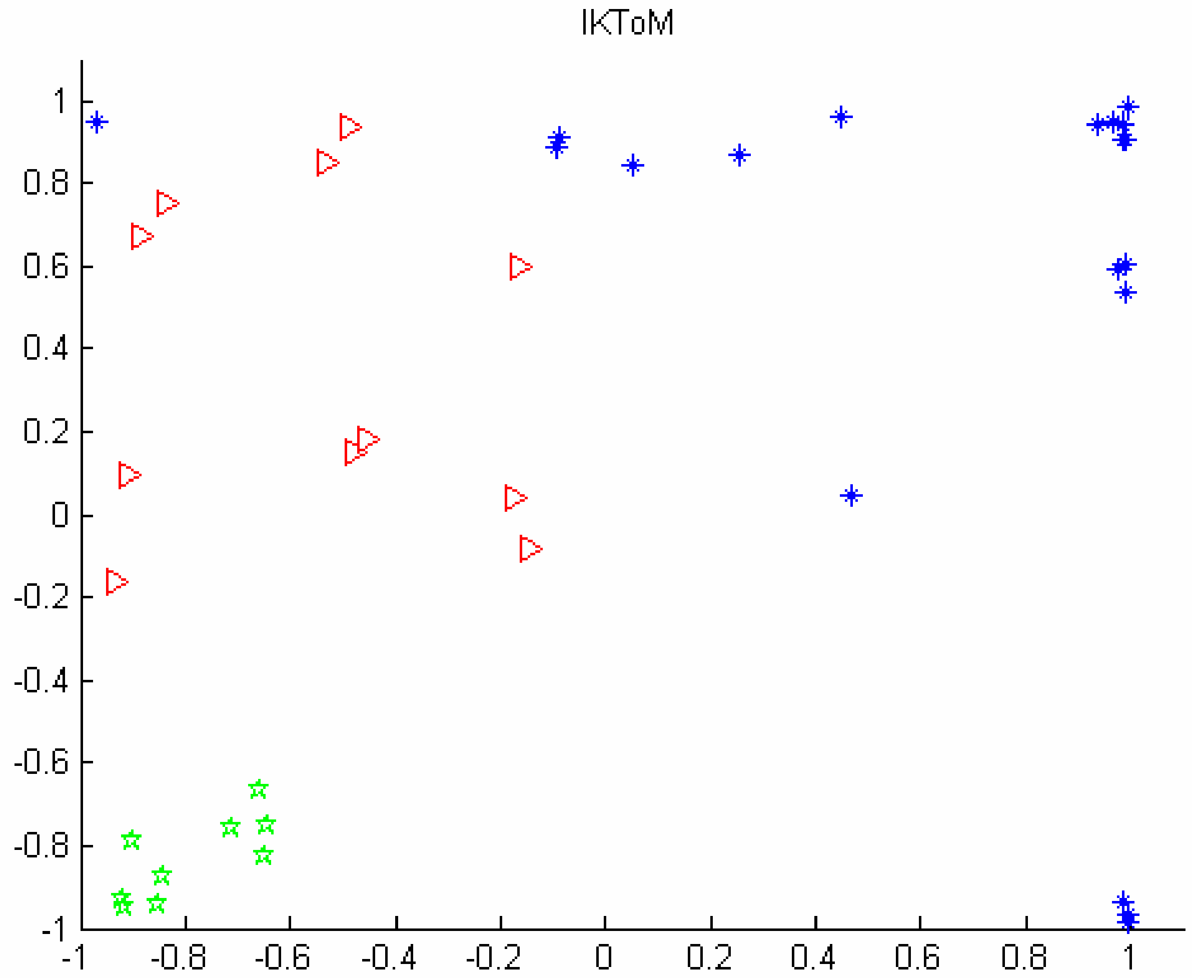
- Inverse weighted K-means

$$perf = \sum_{i=1}^{N} \left[\sum_{j=1}^{K} \frac{1}{\left\| x_i - m_j \right\|^p} \right] \min_{l=1}^{K} \left\| x_i - m_l \right\|^n$$

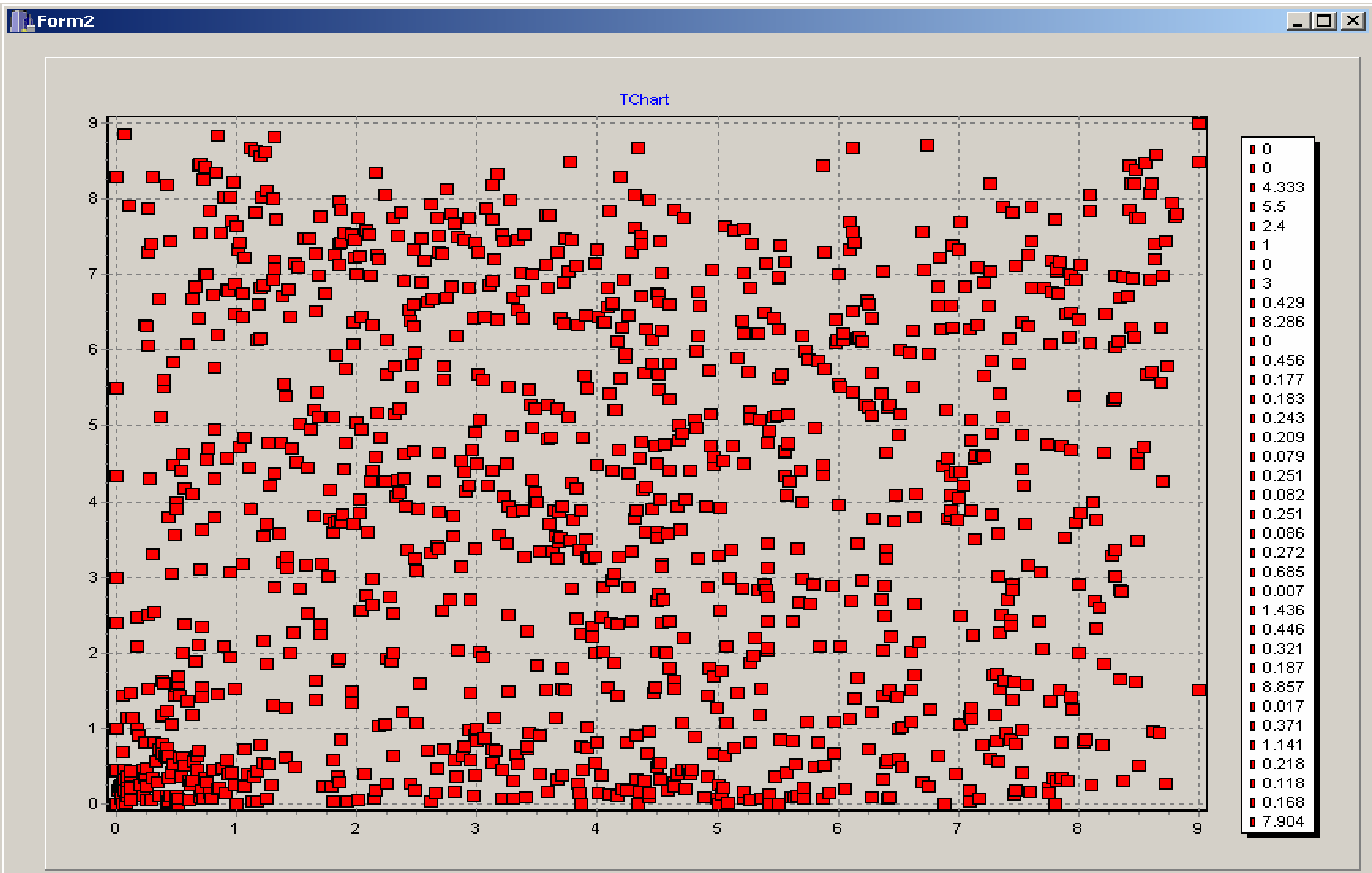
Cancer Data Set 2 - ToPoE



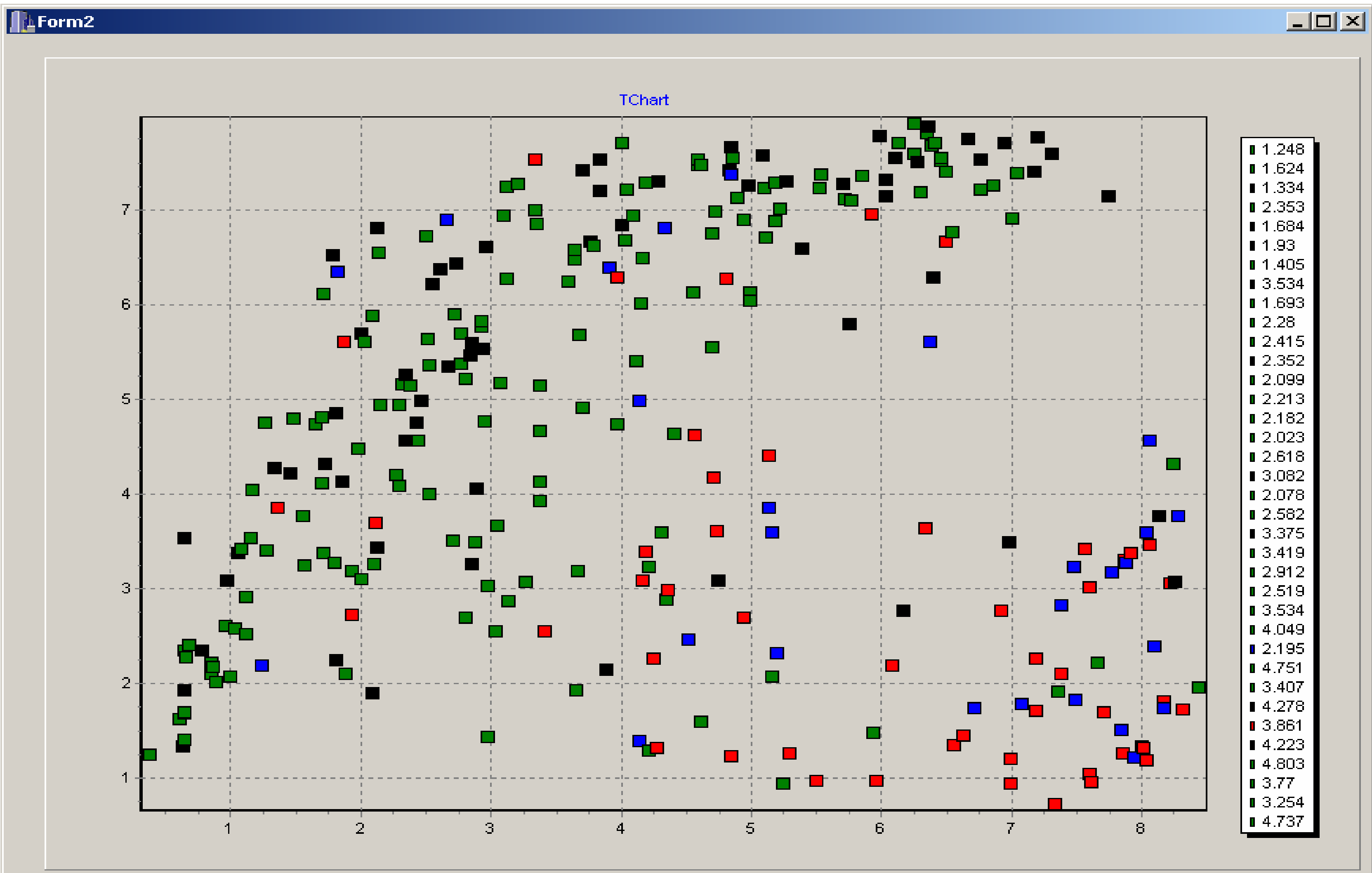
Cancer Data set 2 - IKTOM



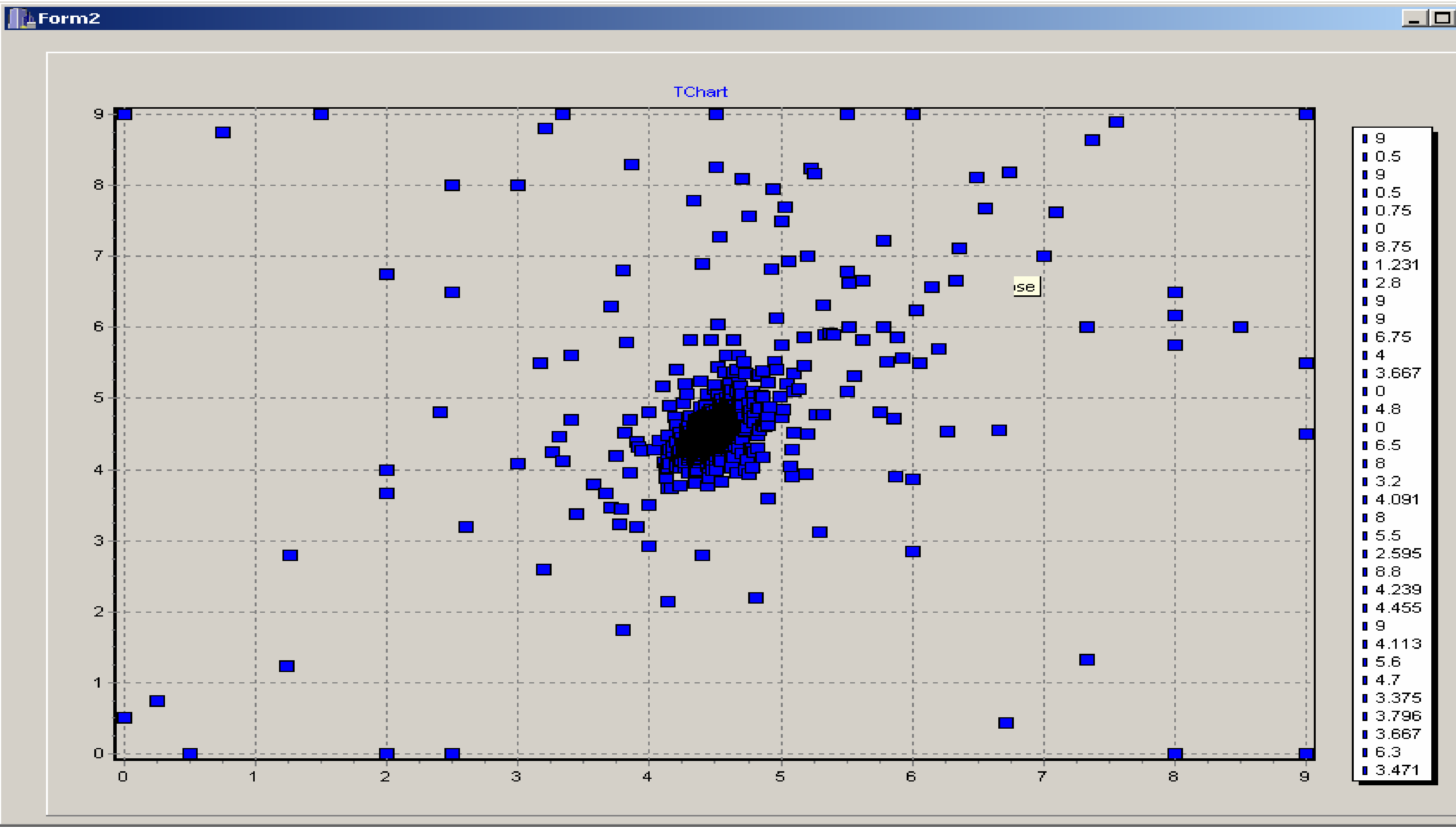
Cancer Data set 2 – ToPoE on Genes



Cancer Data set 1 – ToPoE A, ER+(green), A,ER- (red), B,ER+(blue), B,ER- (black)



Cancer Data set 1 - genes



Conclusion

- New forms of topographic mapping.
- Based on latent space concept but
  - free from probabilistic constraints.
- Product → Mixture of experts.
  - automatic setting of local variances.
- Two types based on K-harmonic Means
- One based on inverse weighted K-means.
- Very sensitive to data, or not, as required.