

# Exploration visuelle des données

Nicoleta ROGOVSCHI

[nicoleta.rogovschi@parisdescartes.fr](mailto:nicoleta.rogovschi@parisdescartes.fr)

M2-INFO

# Self Organizing Map (SOM)

# Outline

- Introduction et définitions des réseaux de neurones
- Introduction SOM
- Algorithme
- Exemple
- Conclusions

# Réseaux de neurones

- Un ensemble de neurones interconnectés/ des unités de traitement de l'information
- Une technique conçu pour modéliser la façon dont le cerveau effectue une tâche particulière
- Permet d'extraire le «pattern» d'information à partir d'ensembles de données qui contiennent des relations latentes
- Habilité de traiter des données bruitées

# L'apprentissage par réseaux de neurones

- Par apprentissage on arrive à extraire de l'information
- Cette information est stockées sur les liens entre les neurones

Réseau de neurones

- On l'appelle aussi poids



- A l'aide des réseaux de neurones on peut réaliser 2 types d'apprentissage:
  - Supervisé
  - Non supervisée

# L'auto-organisation

- Les cellules du cerveau s'auto-organisent en groupes, selon les informations reçues.
- Cette information est non seulement accueillie par une seule cellule nerveuse, mais influe également sur d'autres cellules dans son voisinage. Cette organisation se traduit par une sorte de carte, où les cellules nerveuses ayant des fonctions similaires sont disposées proches les unes par rapport aux autres.
- Le mécanisme des cartes SOM est aussi basé sur ce principe

# Le principe des SOM

- SOM produit un graphe de similarité des données de départ
- Cette technique convertit les relations non linéaires entre les données de haute dimension en relations géométriques simples

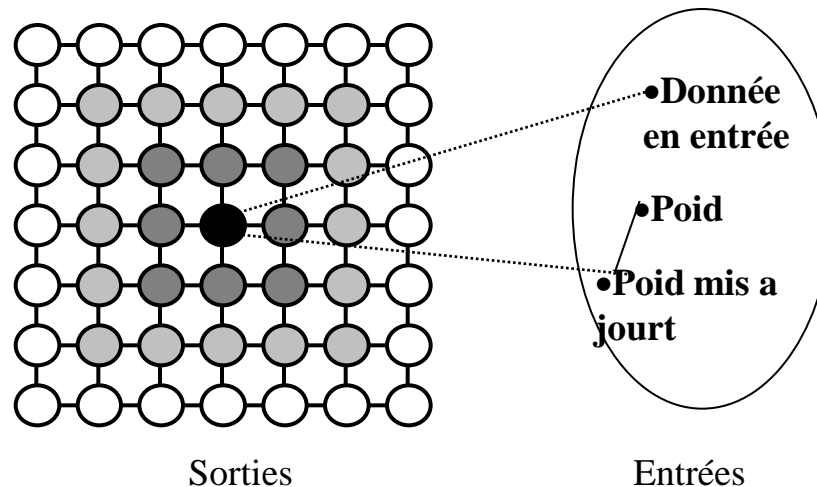


Illustration d'une carte SOM de taille 7x7

# SOM

- Introduite en 1984 par Teuvo Kohonen
- Quantification vectorielle + projection vectorielle
- Technique basée sur l'apprentissage non-supervisé
- Utilisée en classification et visualisation des données de grandes taille
- Utilisée dans de nombreux domaines

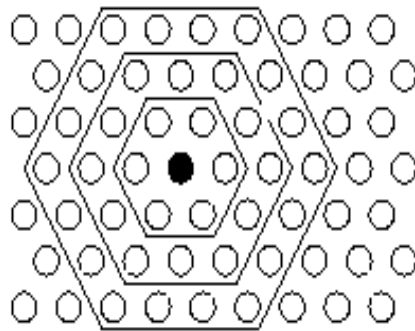


# Domaines d'applications

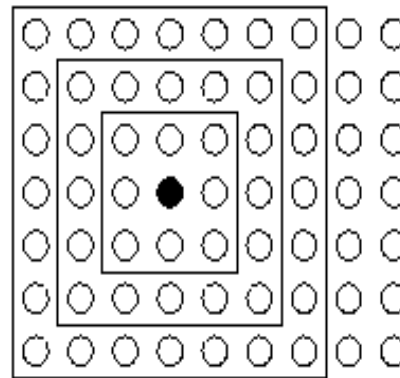
- Applications
  - Analyse exploratoire des données, clustering
  - Quantification, sélection des variables, détection des données aberrantes
  - Diagnostic, prédiction, données manquantes
- Domaines
  - Socio-economique
  - TextMining
  - Télécommunication
  - Processus industriels

# Architecture de SOM

- Un ensemble de neurones / unités des clusters
- Chaque neurone est associé à un vecteur prototype qui est pris à partir de l'ensemble de données d'entrée
- Les neurones de la carte peuvent être disposés sur une grille de forme rectangulaire ou hexagonale



Voisinage hexagonal

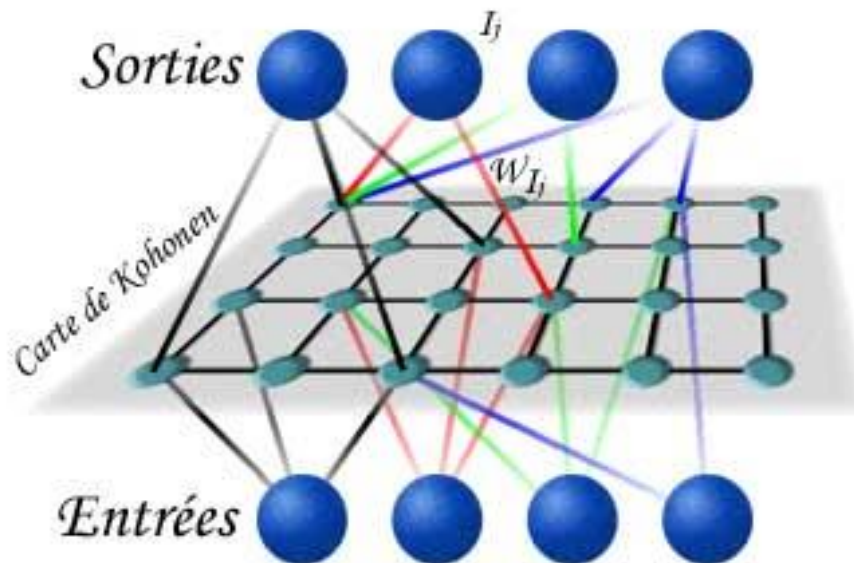


Voisinage rectangulaire

# SOM

- Fonction de coût:

$$C(w) = \sum_i \sum_k h_{k,s(x_i)} \|x_i - w_k\|^2$$



# Algorithme de SOM

## 1. Phase d'initialisation

- Définir la topologie de la carte
- Initialiser aléatoirement tous les prototypes pour chaque neurone.

## 2. Phase de compétition

- Présenter une donnée  $x_i$  choisie aléatoirement
- Déterminer le neurone gagnant selon la règle :

$$s(x_i) = \underset{1 \leq k \leq m}{\operatorname{Arg\,min}} \|x_i - w_k\|^2$$

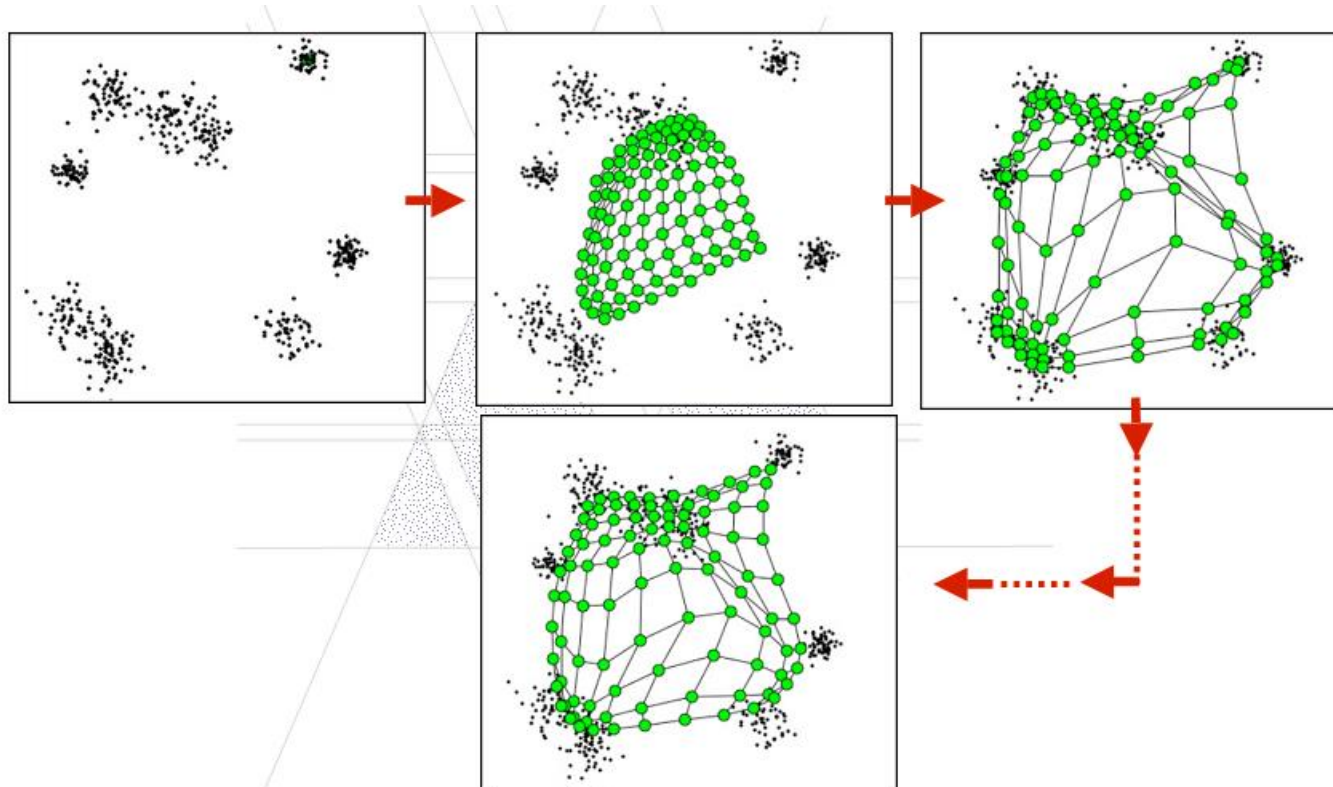
## 3. Phase d'adaptation

- Adapter les prototypes selon la règle

$$w_k(t+1) = w_k(t) + \varepsilon(t) h_{k,s(x_i)} (x_i - w_k(t))$$

- ## 4. Répéter les phases 2 et 3 jusqu'à ce que les mises à jours des prototypes soient négligeables

# Phase d'auto organisation



# Phase d'auto organisation

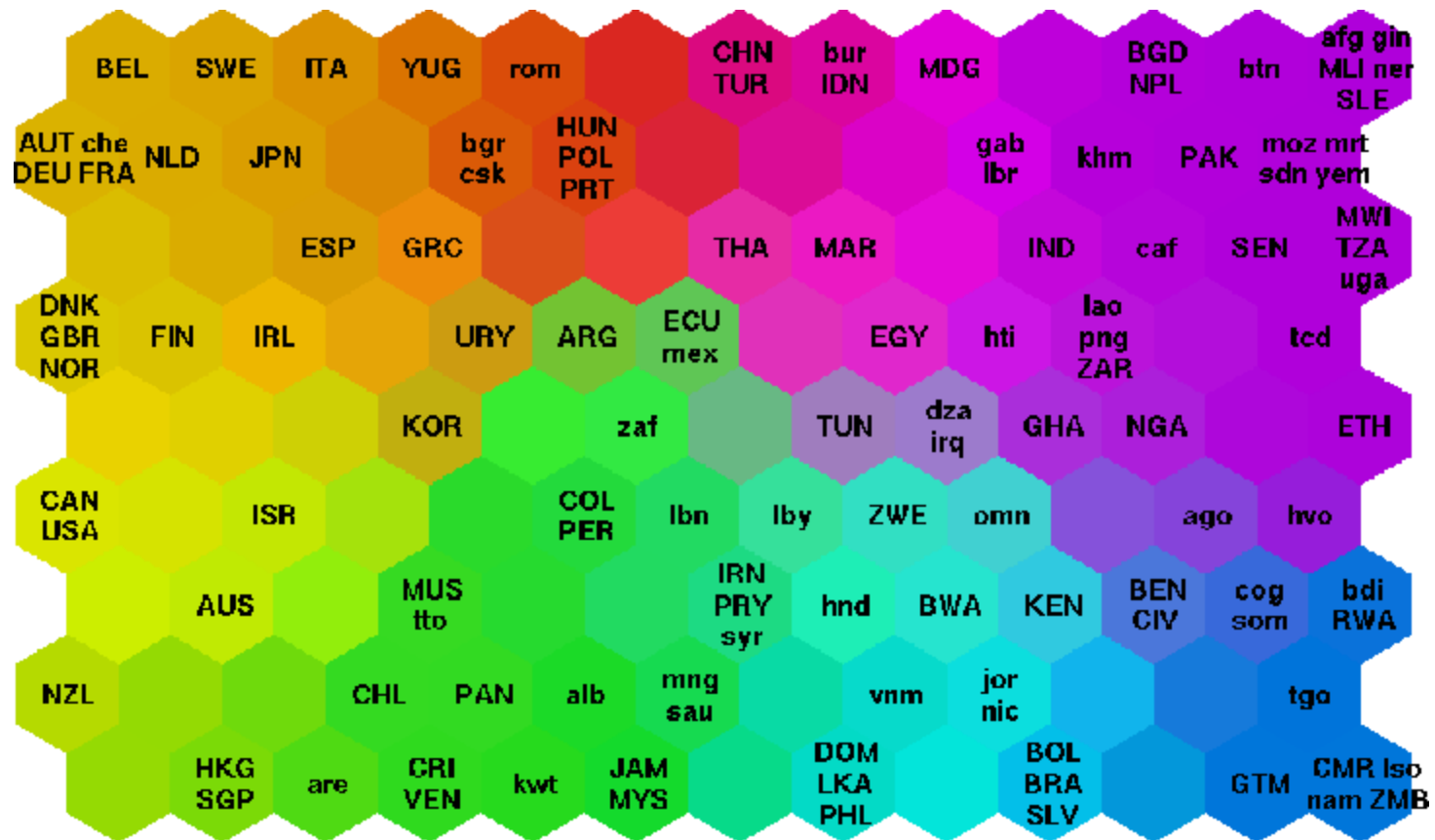
La carte s'organise en vérifiant les propriétés suivantes :

- Chaque neurone se spécialise dans une portion de l'espace d'entrée
- Des données similaires auront des projections proches sur la carte.

# Visualisation des données utilisant SOM

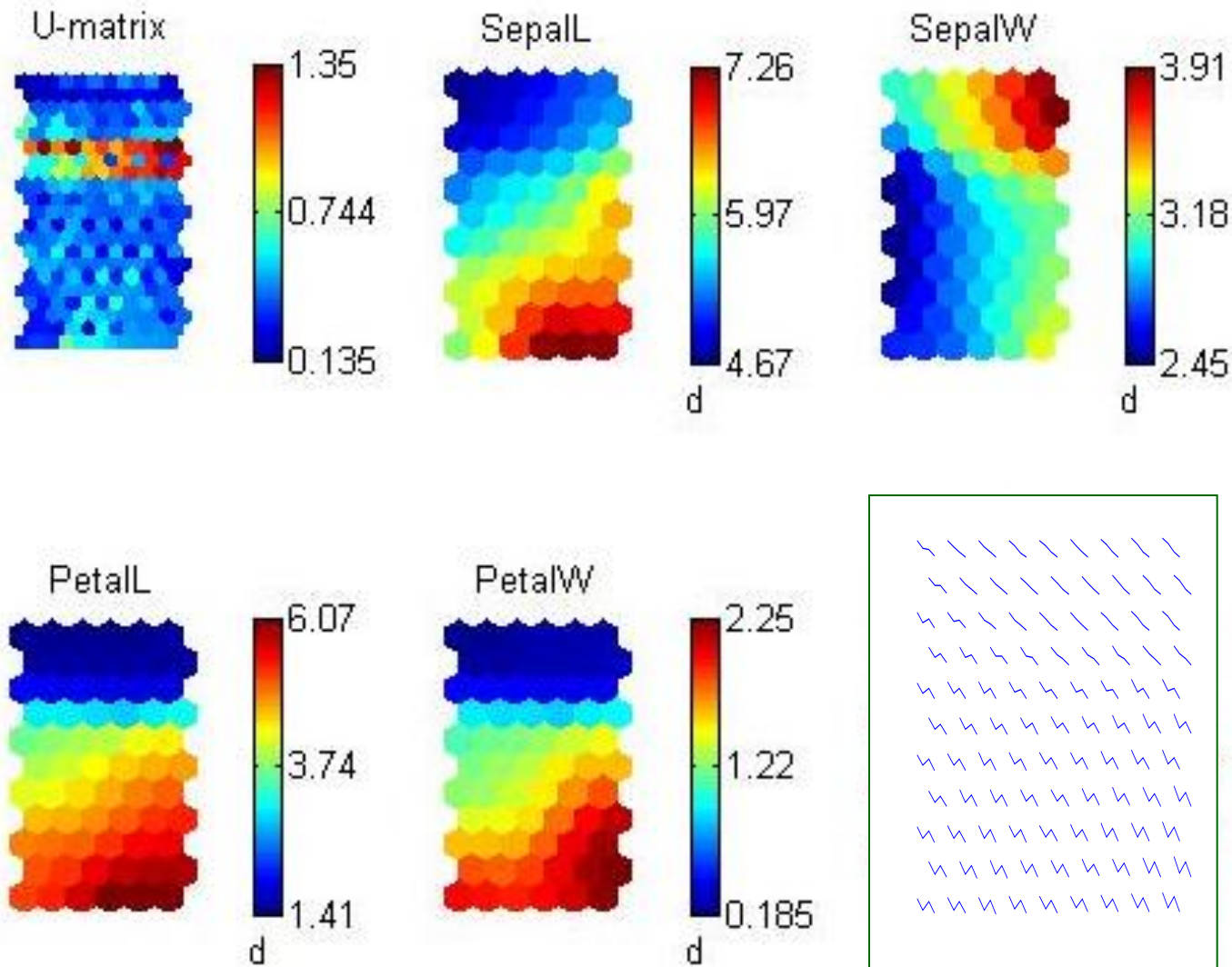
- Visualisation des clusters et de la forme des données (projections, U-matrices et d'autres matrices de distances)
- Visualisation des composantes/variables (scatter plots, plan des composantes)
- Visualisation des projections des données

## Exemple : Pays du monde

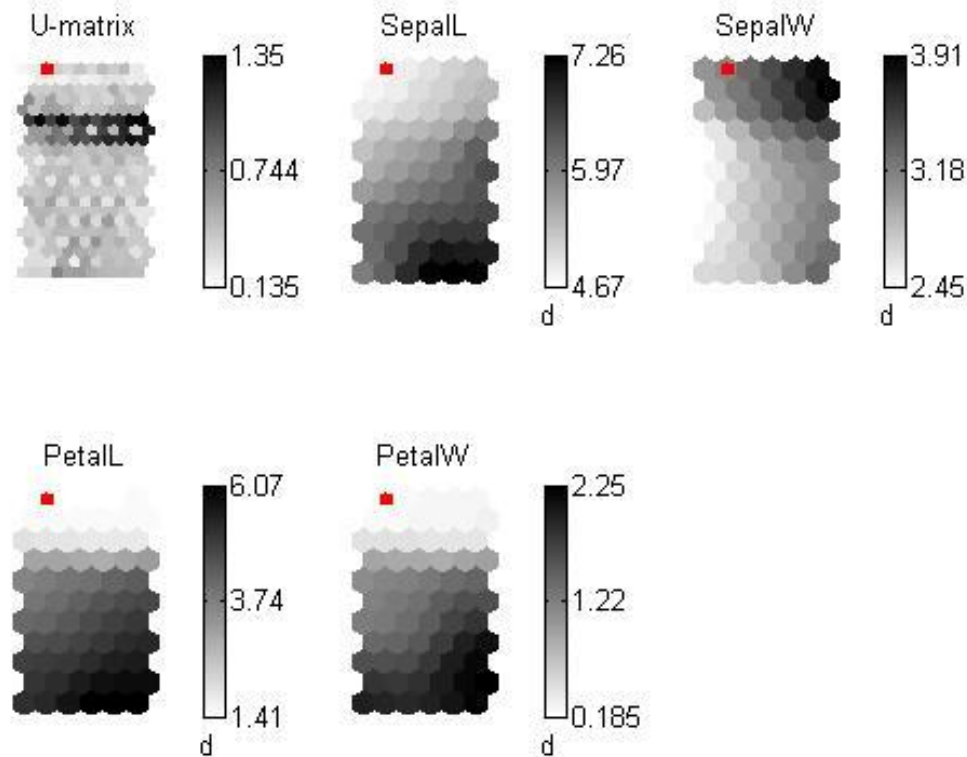




# Example: Iris

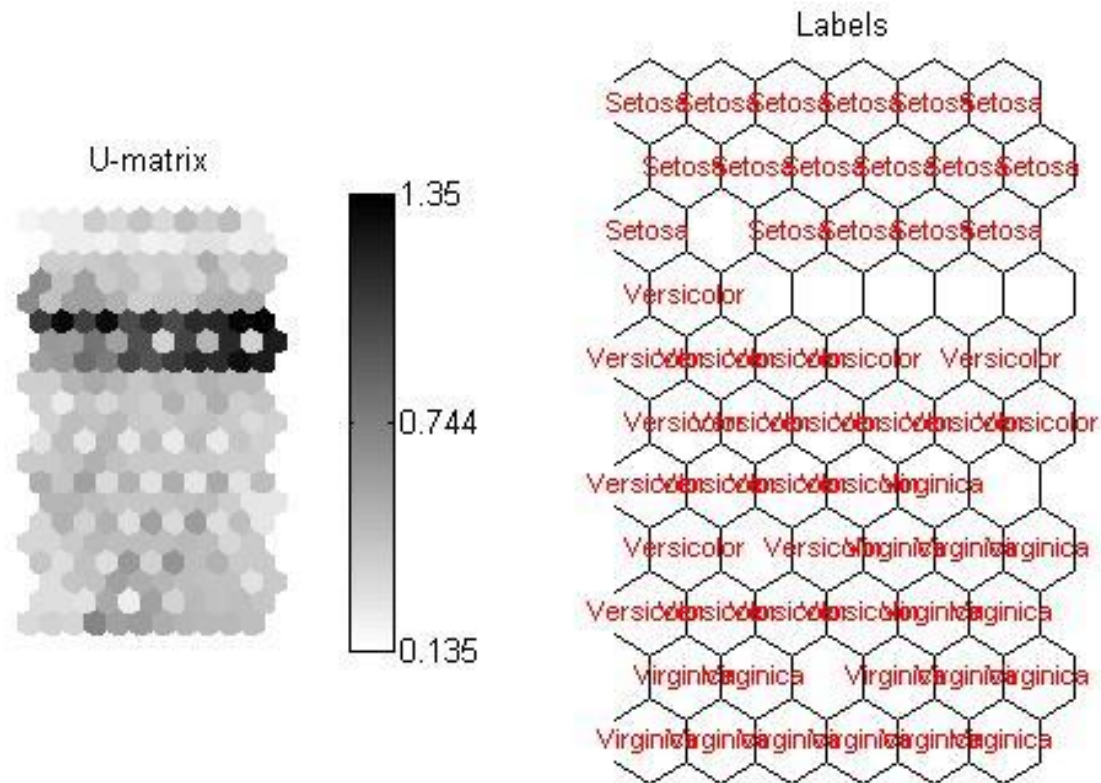


# Example: Iris

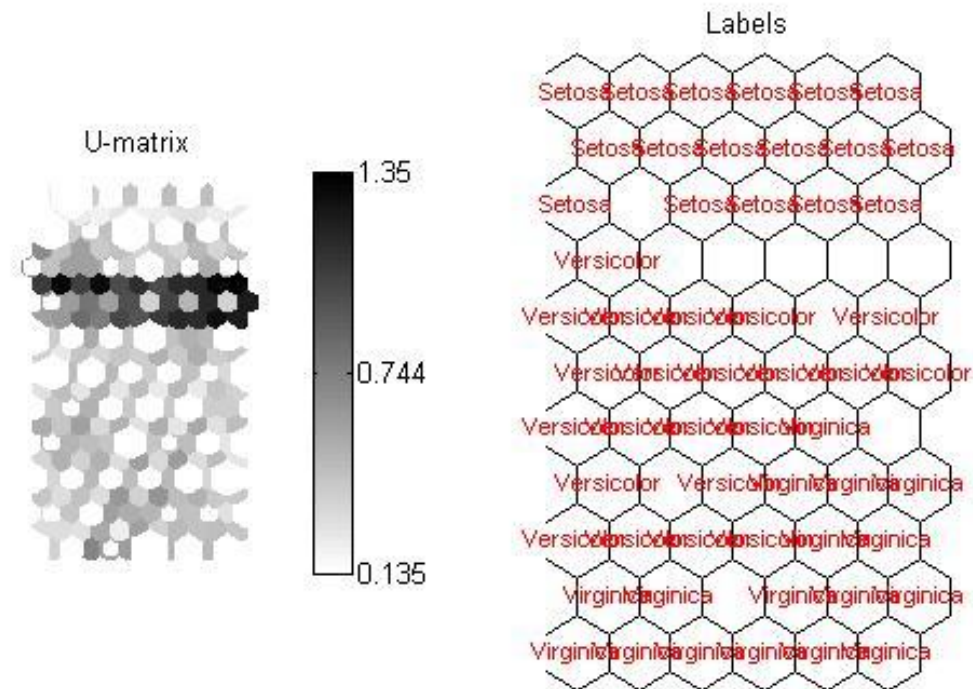


SOM 15-Nov-2016

# Example: Iris

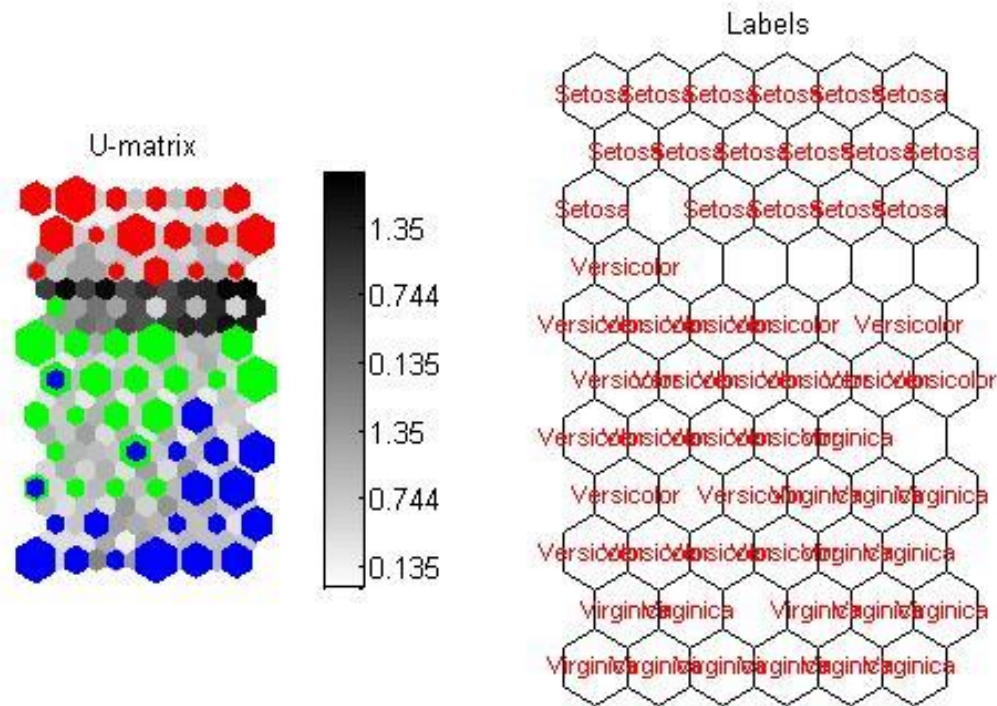


# Example: Iris

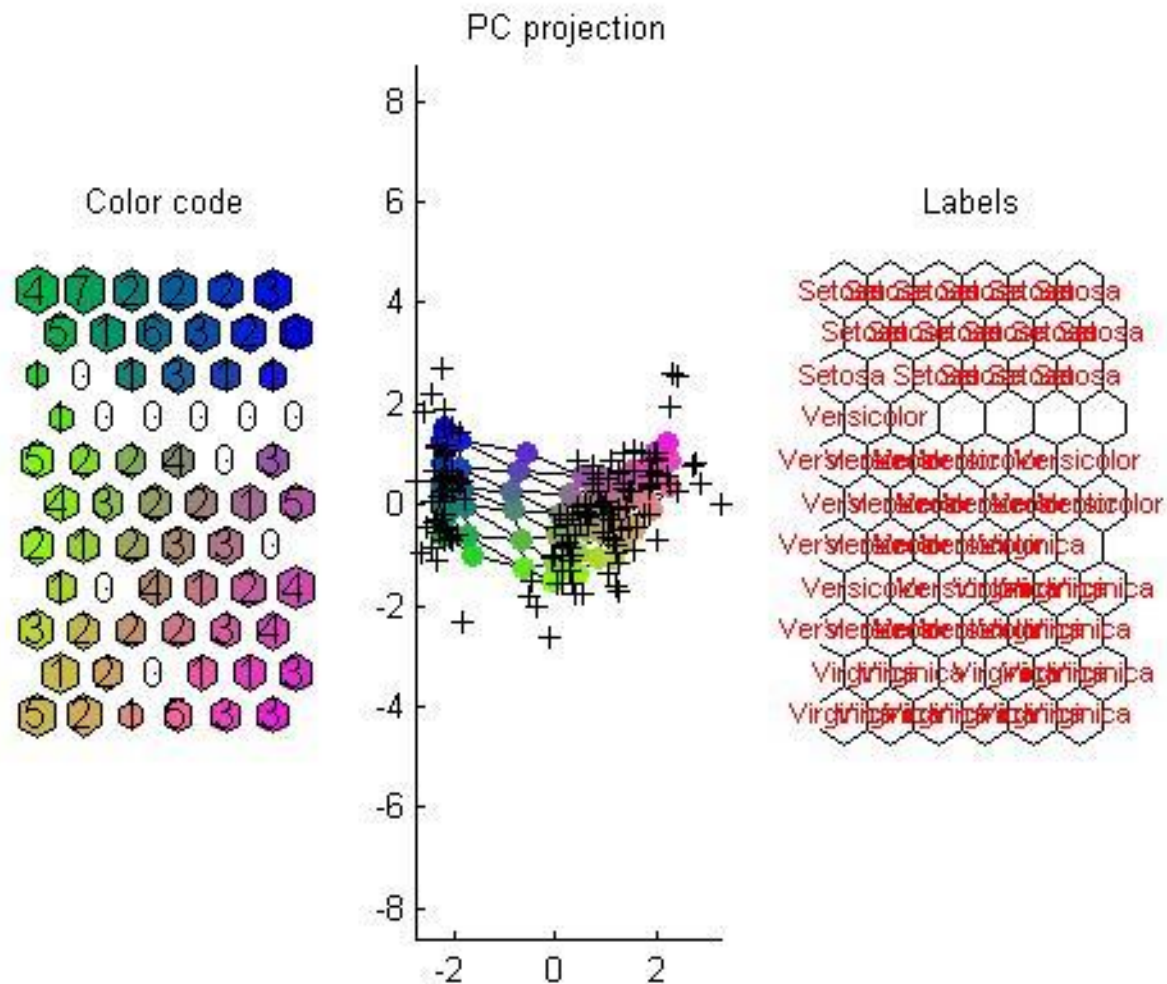


SOM 15-Nov-2016

# Example: Iris

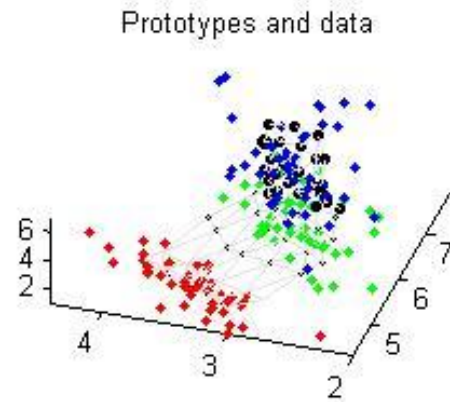
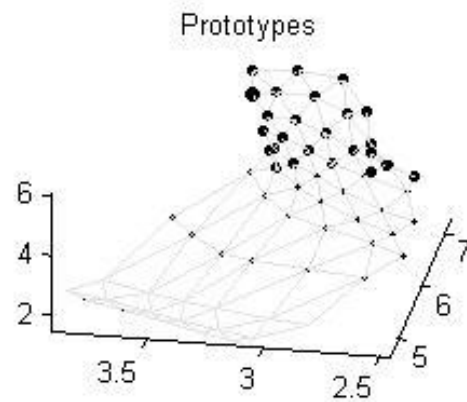
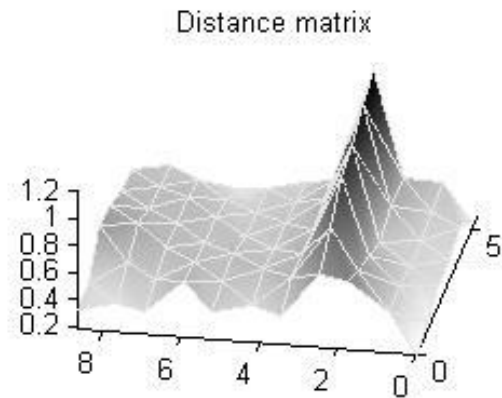
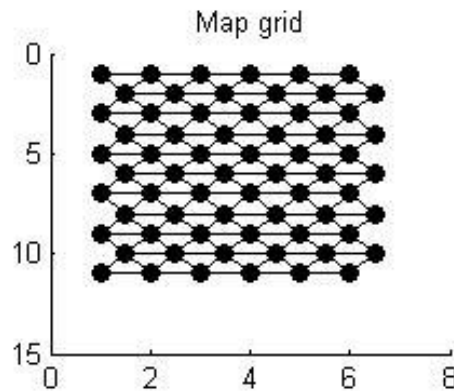


# Exemple: Iris

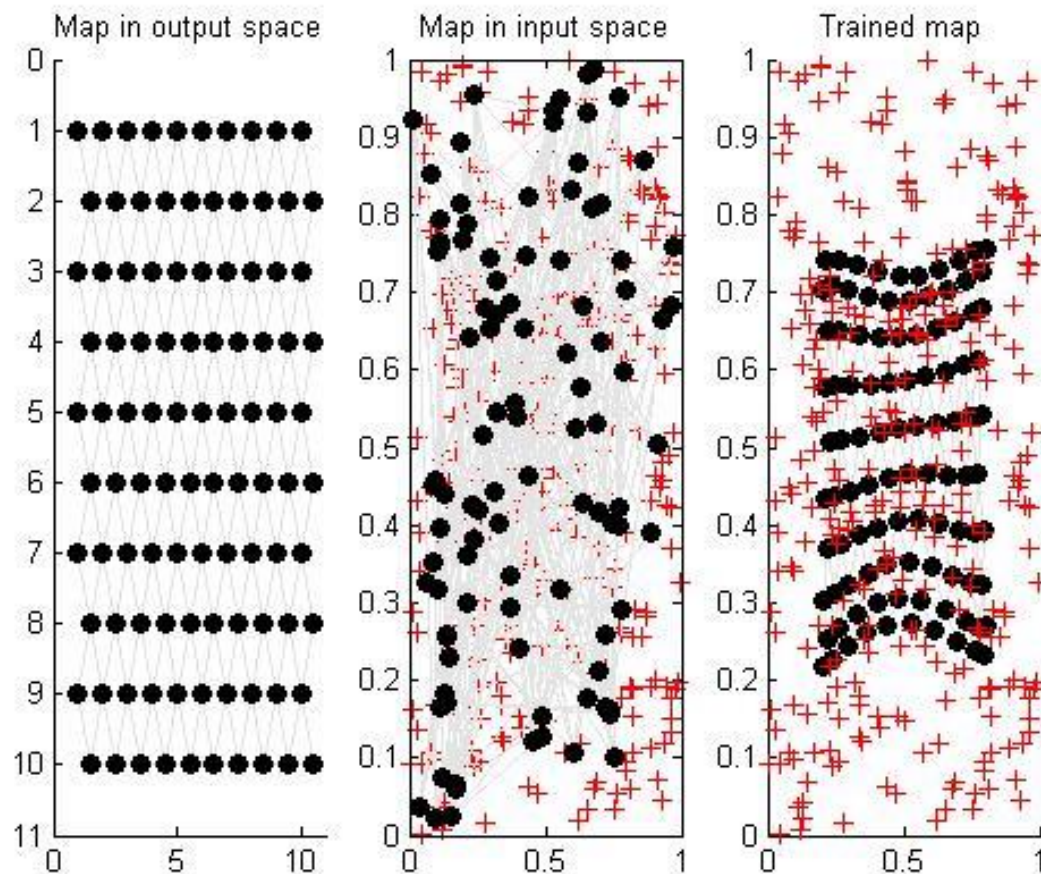




# Example: Iris

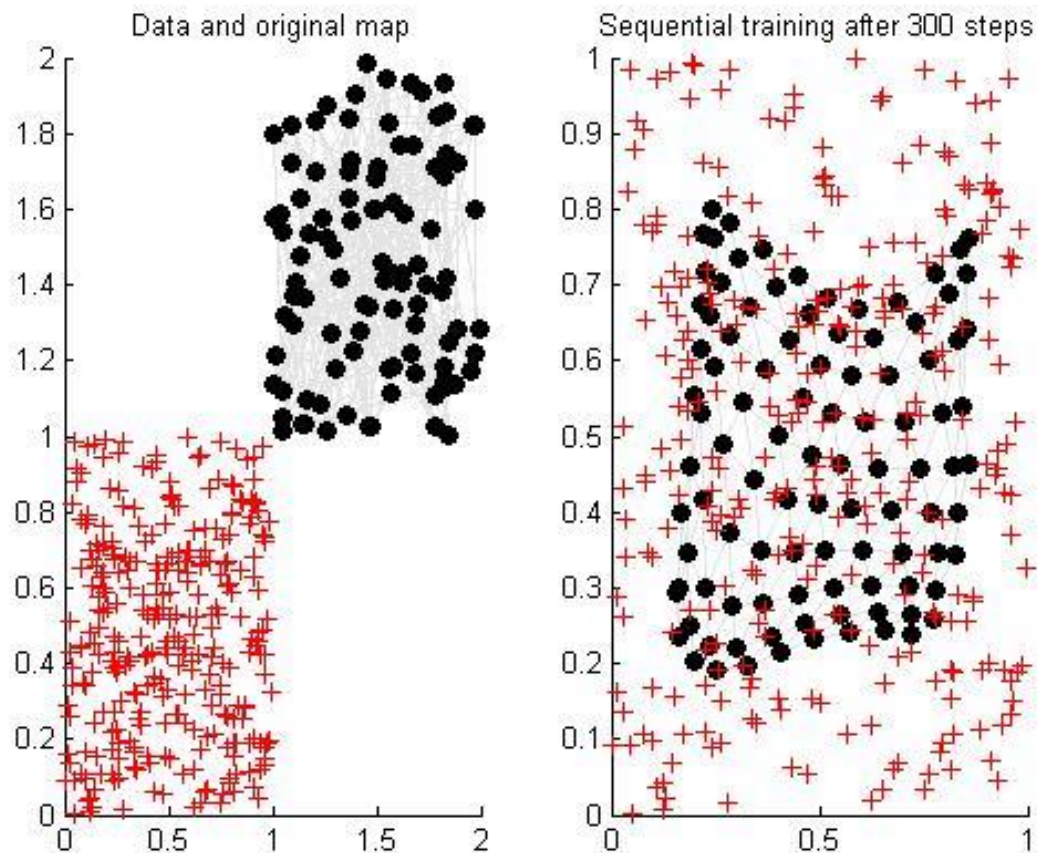


# Example

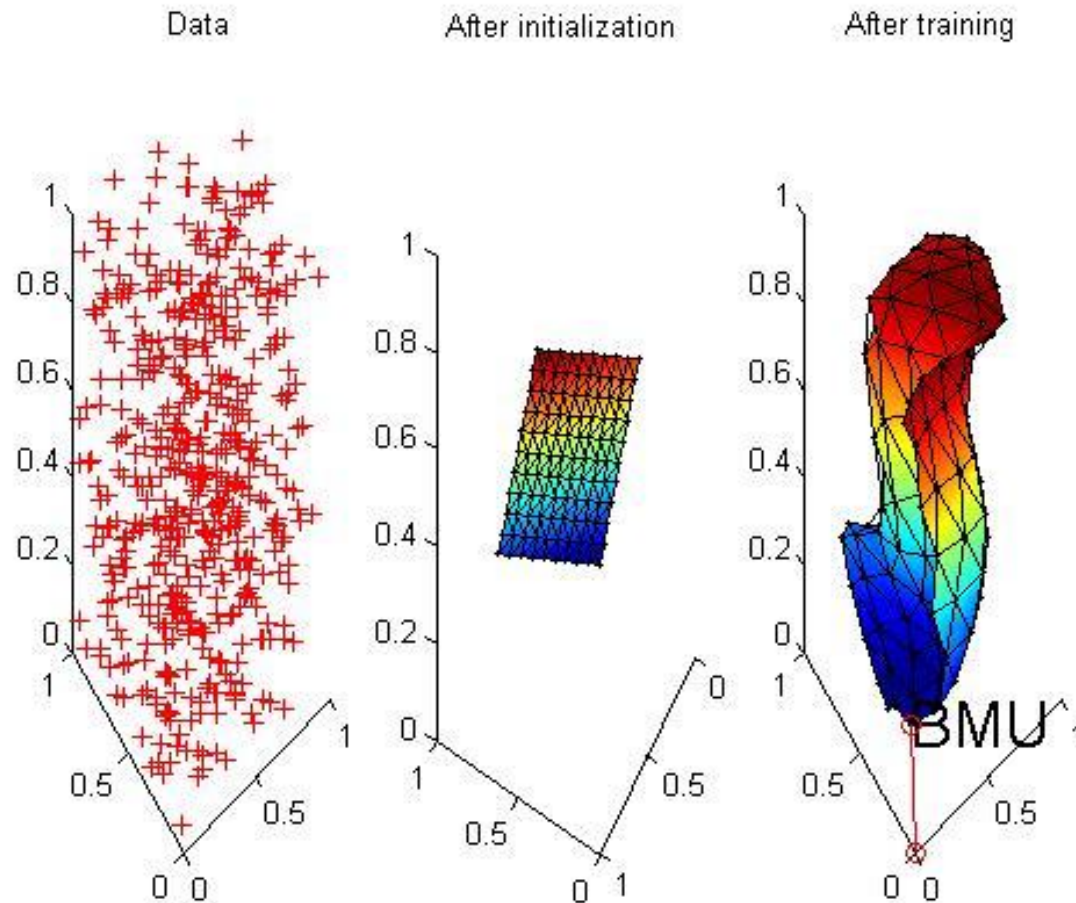




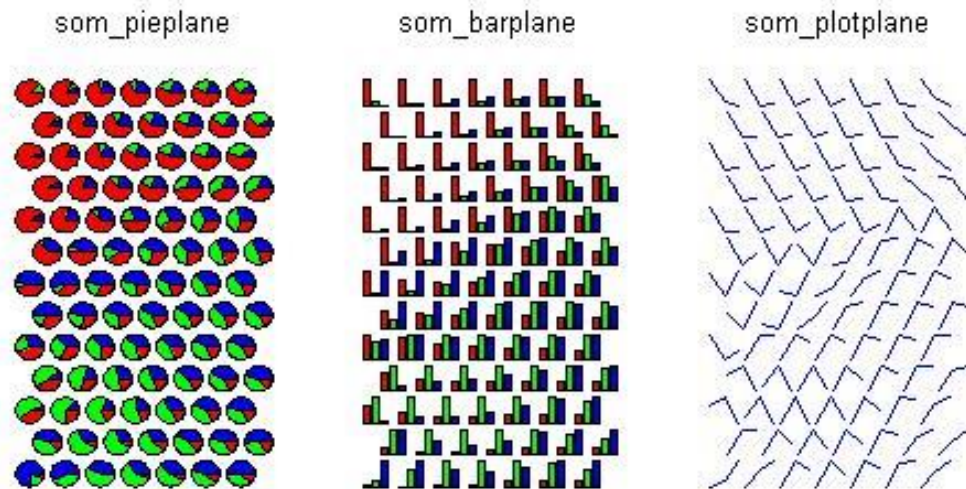
# Example



# Example

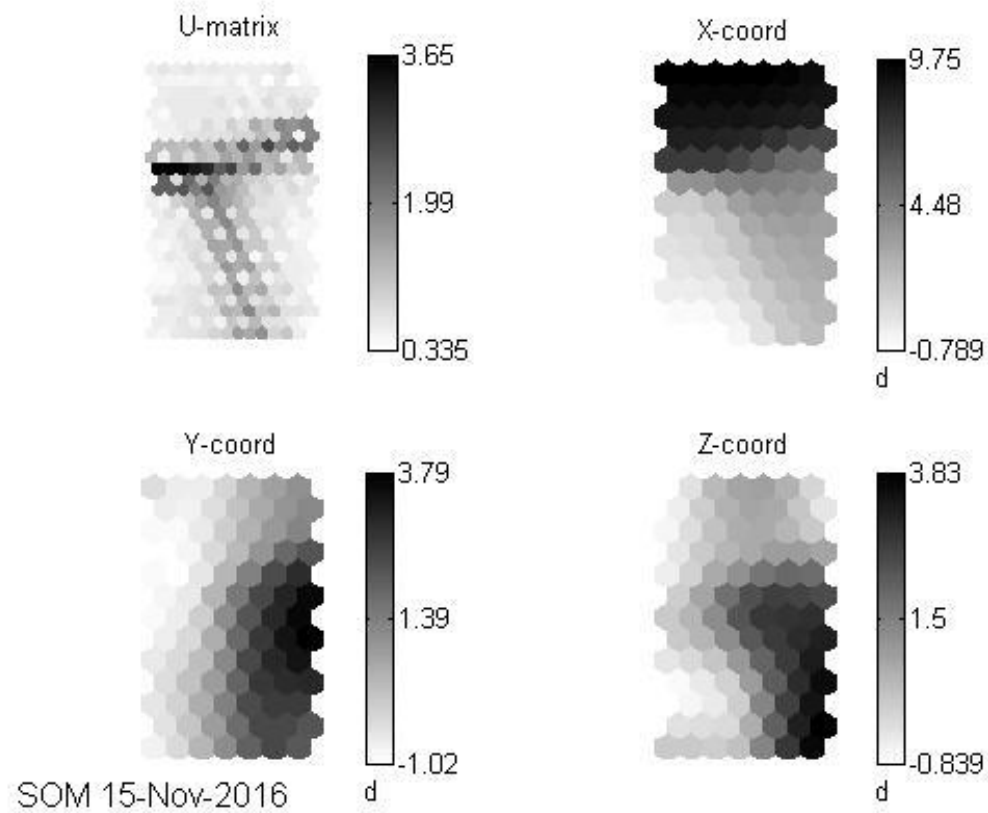


# Example



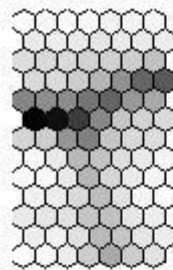
SOM 15-Nov-2016

# Example

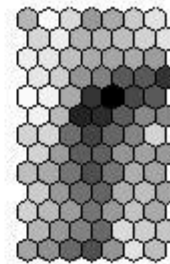


# Example

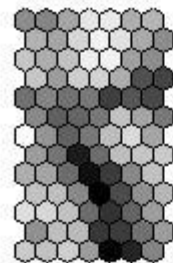
X-coord



Y-coord



Z-coord



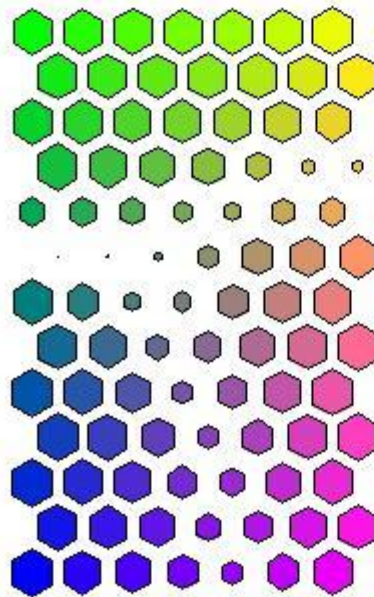
Relative importance



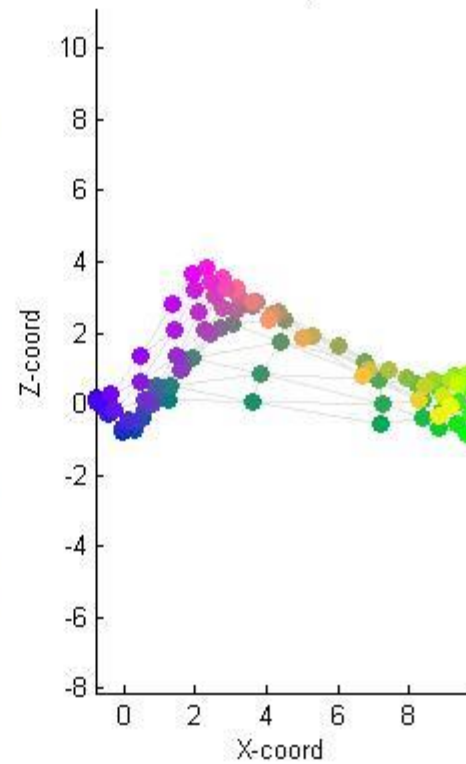
SOM 15-Nov-2016

# Example

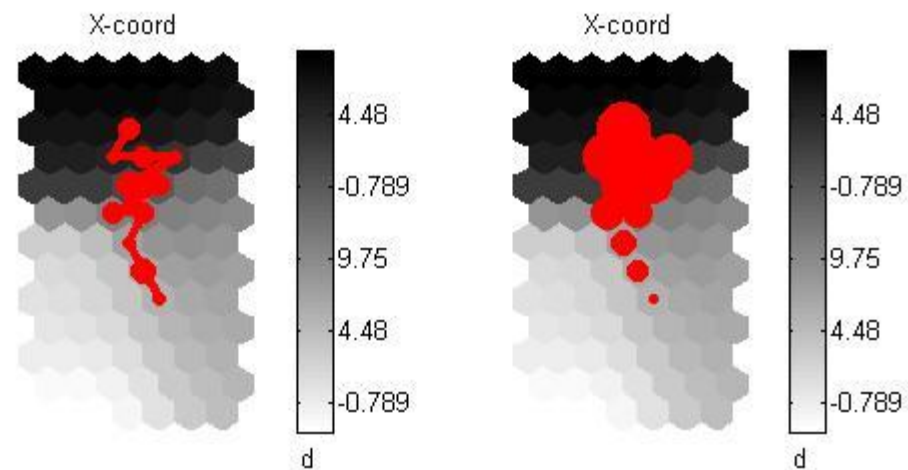
Color coding + distance matrix



Scatter plot



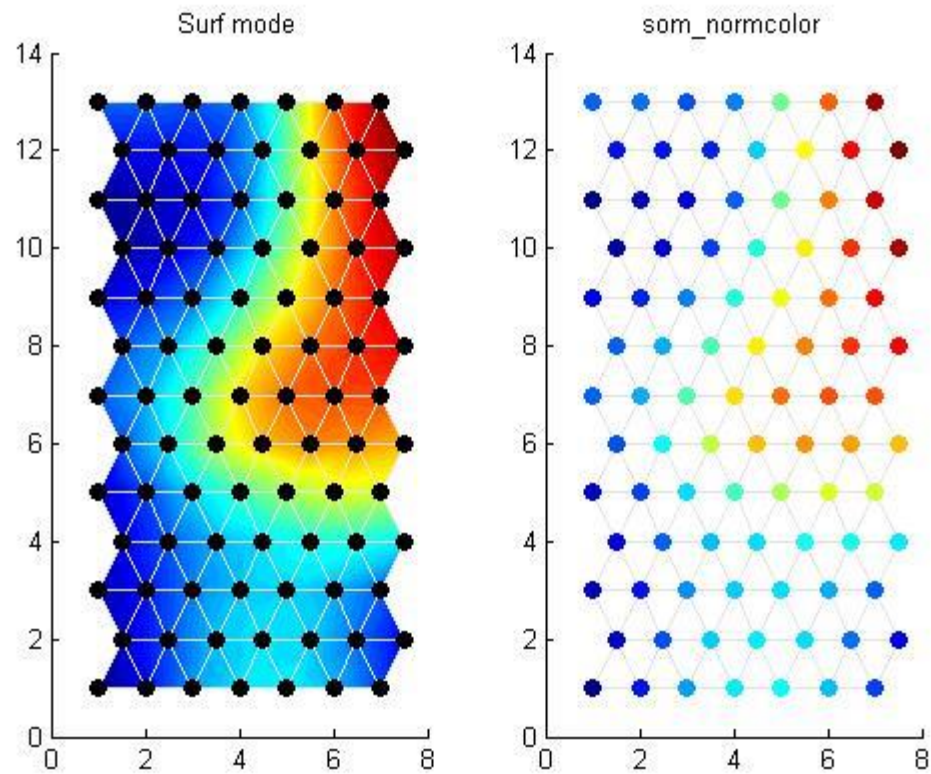
# Example



SOM 15-Nov-2016



# Example





# Example

