

# Clustering

Maxime CHAMBREUIL  
maxime.chambreuil@insa-rouen.fr

---

## Table des matières

<b>1</b>	<b>But du TP</b>	<b>1</b>
<b>2</b>	<b>Explication</b>	<b>1</b>
2.1	Les fonctions . . . . .	1
2.1.1	som_make . . . . .	1
2.1.2	som_show . . . . .	2
2.1.3	som_cplane . . . . .	2
2.2	Déroulement de l'algorithme . . . . .	2
2.3	Résultats . . . . .	3
<b>3</b>	<b>Conclusion</b>	<b>7</b>

---

## 1 But du TP

L'objectif du TP est de mettre en oeuvre des cartes de Kohonen, et de comparer cette méthode aux K-Means et à la Classification Hiérarchique Ascendante.

## 2 Explication

### 2.1 Les fonctions

#### 2.1.1 som\_make

```
function sM = som_make ( donnees , 'msize' , [10 10] , 'lattice' , 'rect' )
```

**En entrée :**

donnees : matrice des données d'apprentissage

msize : option permettant de spécifier la taille de la carte ( ici 10 par 10 )

lattice : option permettant de spécifier le treillis de la carte ( ici rectangulaire )

**En sortie :**

sM : structure de la carte de Kohonen

Cette fonction projete les données d'apprentissage sur la carte pour l'initialiser.

### 2.1.2 som\_show

*function som\_show ( sM )*

**En entrée :**

sM : structure de la carte de Kohonen

Cette fonction permet d'afficher chacune des variables séparément, avec différentes teintes selon la valeur de la variable pour chaque noeud de la carte.

### 2.1.3 som\_cplane

*function h = som\_cplane ( sM.topol.lattice , sM.topol.msize , sM.codebook./255 )*

**En entrée :**

sM.topol.lattice : treillis de la carte

sM.topol.msize : taille de la carte ( ici 10 par 10 )

sM.codebook./255 :

**En sortie :**

h : scalaire

Cette fonction permet de visualiser la carte de Kohonen avec les formes ( carrées ou hexagones ) et la grille spécifiée ( 10 par 10 ).

## 2.2 Déroulement de l'algorithme

Après avoir chargé les données, on fait d'abord un som\_make pour générer la carte, puis on l'affiche de différentes façons à l'aide de som\_show ou som\_cplane.

## 2.3 Résultats

Voici les résultats obtenus :

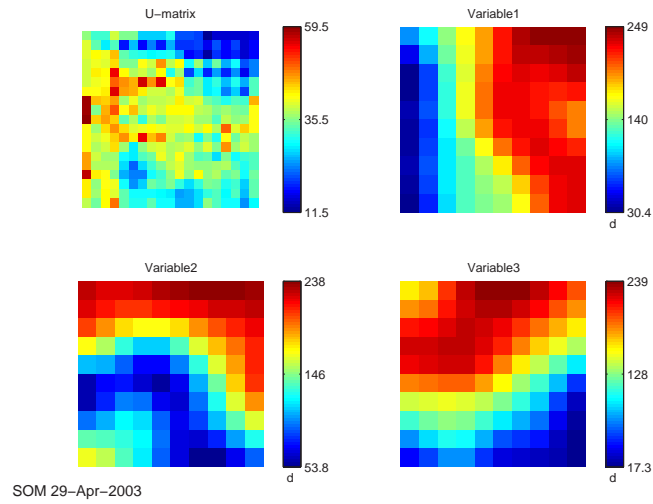


FIG. 1 – Som show avec tous les individus

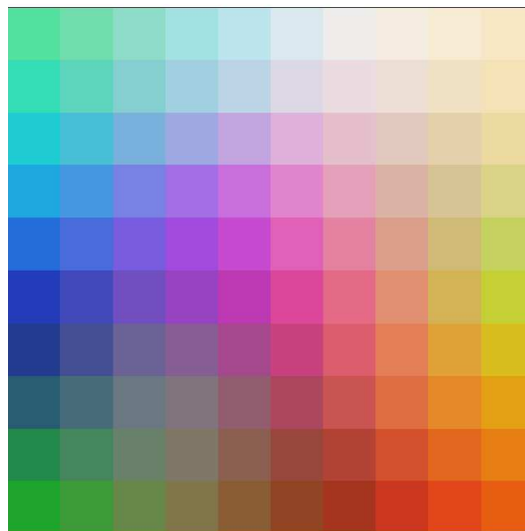


FIG. 2 – Som cplane avec tous les individus

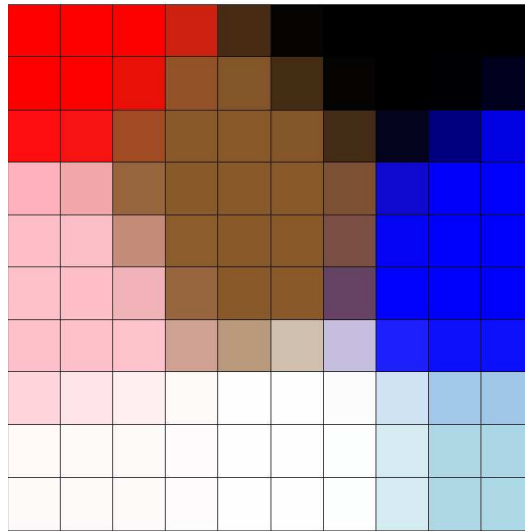


FIG. 3 – Som cplane avec 8 couleurs

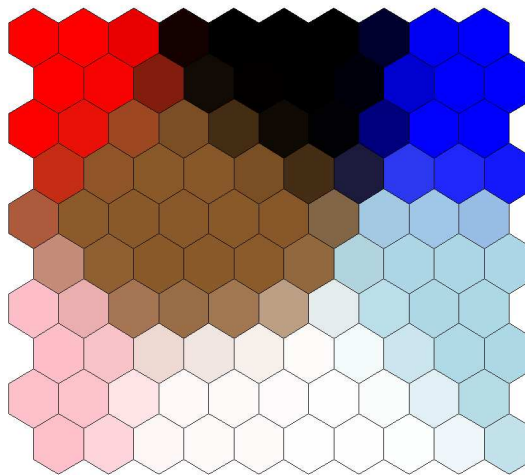


FIG. 4 – Som cplane avec 8 couleurs en hexagonale

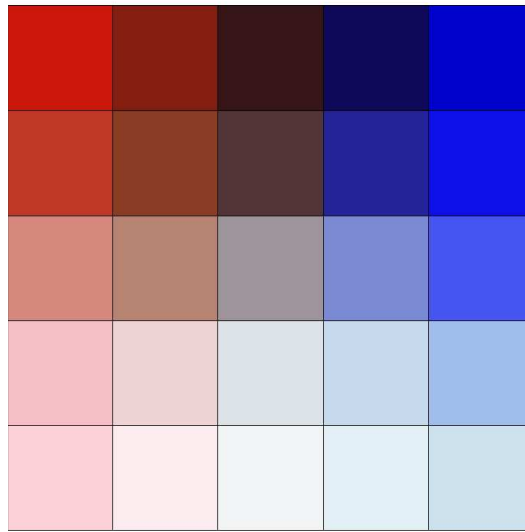


FIG. 5 – Som cplane avec 8 couleurs et une carte de 5 noeuds de coté

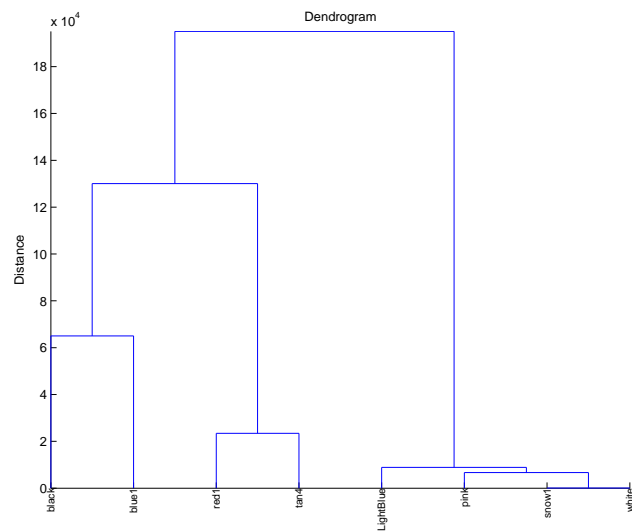


FIG. 6 – Dendrogramme avec 8 couleurs

couleurs.mat : Formes fortes. Distance Euclidienne. Nombre de boucles :50

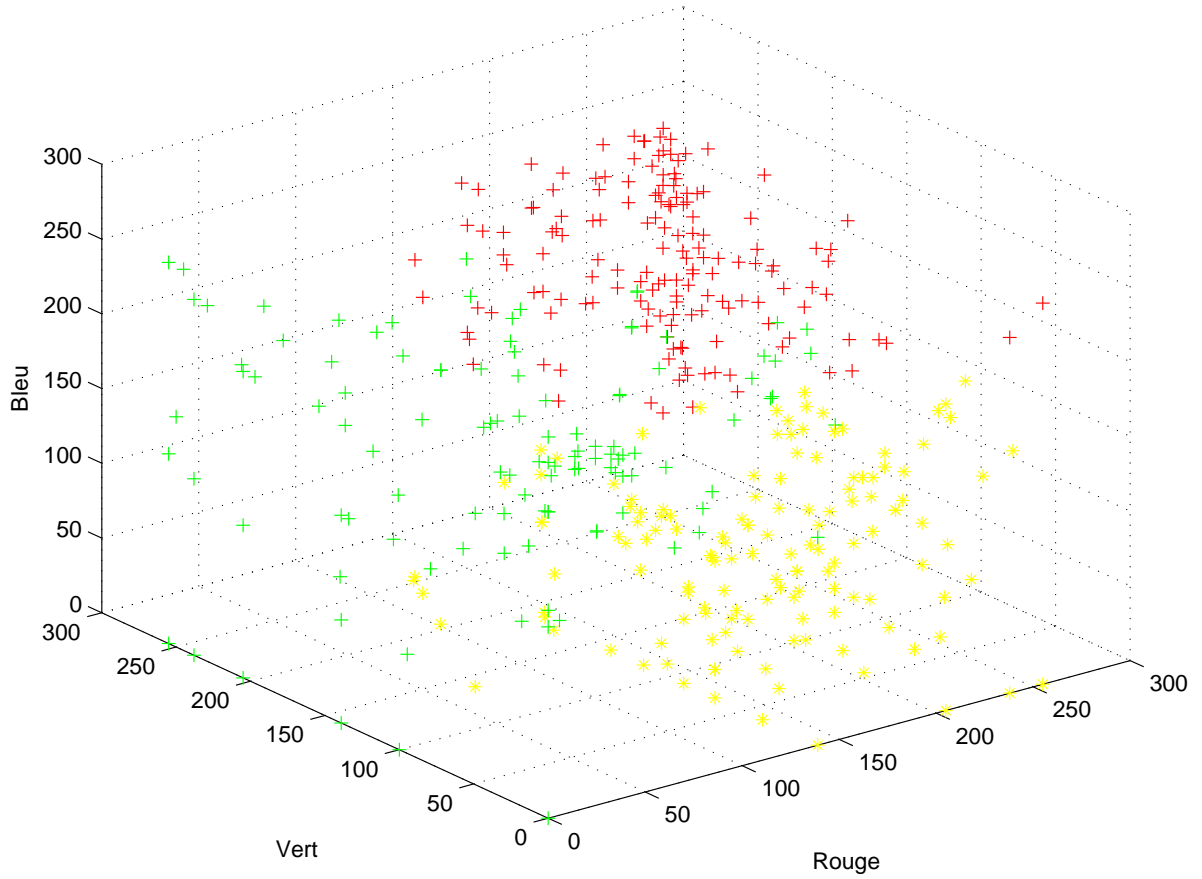


FIG. 7 – K-Means avec 3 centres et 50 itérations

couleurs.mat : Formes fortes. Distance Euclidienne. Nombre de boucles :20

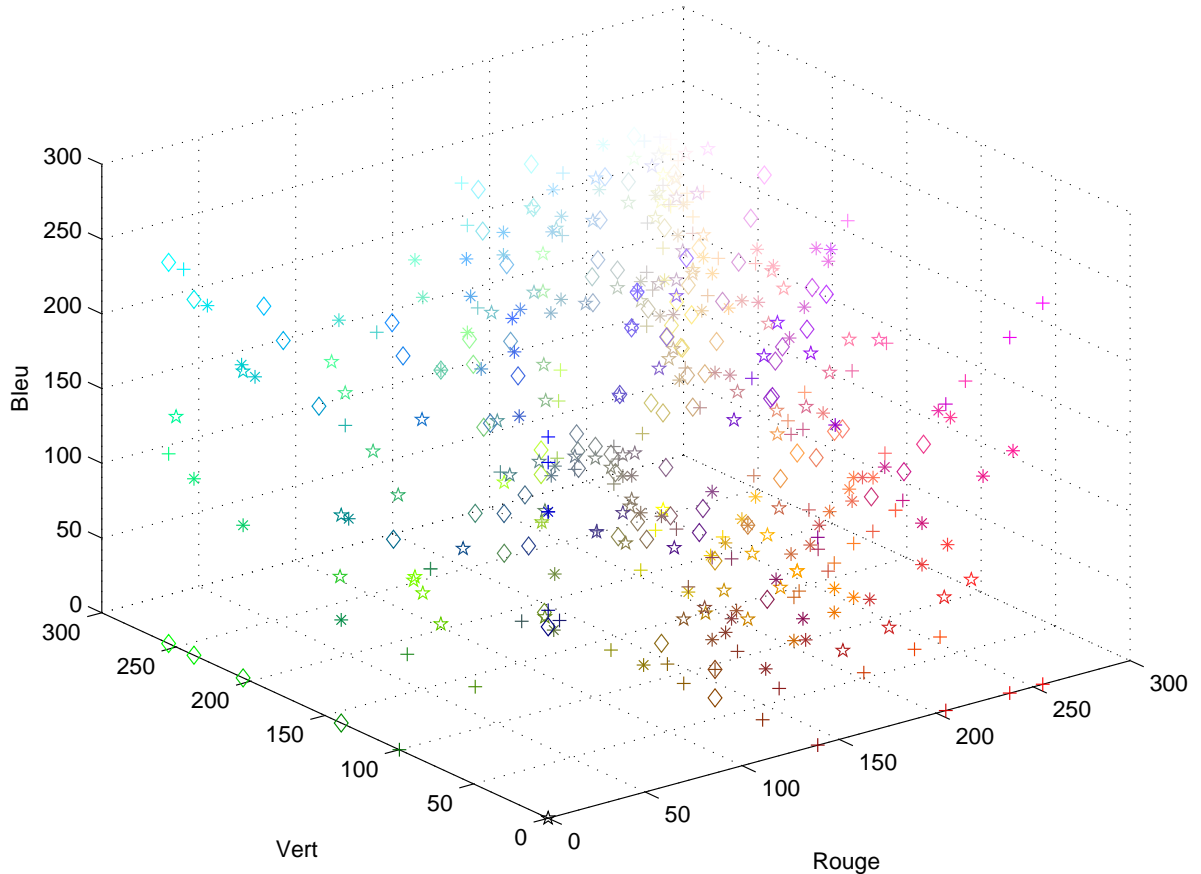


FIG. 8 – K-Means avec tous les points

### 3 Conclusion

J'ai pas réussi à trouver de meilleurs moyens pour afficher les résultats. J'aurais ainsi pu mieux les comparer.

On peut tout de même en conclure que le dendrogramme est infallible dans le regroupement des individus et permet de trouver le nombre de classe optimum (suivant les distances). Mais il est fonctionnel avec peu d'individu...

Concernant la carte de Kohonen, elle regroupe bien les individus similaires, mais la distinction du nombre de classe est très personnelle.

Quant aux K-Means, on peut trouver facilement le nombre de formes fortes et les individus qui les constituent, mais la méthode n'est pas très robuste à différents jeux de données ou de problèmes.