



Ministère de l'enseignement supérieur et de la recherche scientifique

Université des Sciences et de la Technologie Houari Boumediene

Faculté de Mathématiques

Département de Probabilités Statistiques

Mémoire de Magister en Probabilités Statistiques

Options : Analyse des Données

Présenté par: M^{me} Boulahbel Samira, épouse Bachari

Thème

Modélisation statistique par cartes topologiques de Kohonen

Soutenu le : 10 Février 2009

Membres du jury:

<i>M^r Aïder Mezyane,</i>	<i>Professeur,</i>	<i>USTHB</i>	<i>Président</i>
<i>M^r Djedour Mohamed,</i>	<i>Professeur,</i>	<i>USTHB</i>	<i>Directeur de thèse</i>
<i>M^{me} Djaballah Khadidja</i>	<i>Maître de conférences,</i>	<i>USTHB</i>	<i>Examinatrice</i>
<i>M^r Chaabane Djamel,</i>	<i>Maître de conférences,</i>	<i>USTHB</i>	<i>Examineur</i>
<i>M^r Rebbouh Amar,</i>	<i>Chargé de cours,</i>	<i>USTHB</i>	<i>Invité</i>

Remerciements

En premier lieu, je tiens à exprimer ma reconnaissance et ma gratitude *Monsieur le Professeur M^{ed} Djedour* pour m'avoir guidé durant la préparation de ce mémoire avec tant d'expertise.

Je désire également remercier Monsieur Aïder, Professeur au Département Recherche Opérationnelle, pour m'avoir fait l'honneur d'être président de mon jury.

Mes sincères remerciements vont également aux *membres du jury, Madame Djaballah, Messieurs Chaabane, et Rebbouh,* pour avoir accepté d'évaluer et de juger le présent travail.

Je tiens à témoigner ma reconnaissance à mon père, qui, par une lecture de la version préliminaire de mon manuscrit, a contribué à en améliorer la qualité et à mon époux, pour son soutien indéfectible.

Je tiens également à remercier le Directeur Général du CREAD, Mr Ferfara Yacine pour m'avoir alloué du temps afin de terminer mon mémoire au sein de son organisme.

Je remercie chaleureusement Leila et Houda, étudiantes à l'INI, pour leur précieuse aide et leur participation essentielle à l'élaboration de l'interface du logiciel.

Merci à Belarbi Yacine, Kahina et Lylia pour les discussions que nous avons pu avoir.

Enfin, il serait particulièrement ingrat de ne pas remercier mon ancien Directeur Mr Bouzar et mes anciens collègues : M^{me} Benmwafek, Keltoum, Samah, Samia, Mouloud et Mouhab pour l'accueil chaleureux et la sympathie dont j'ai bénéficié à mon arrivée à l'ANSEJ ; les moments passés ensemble resteront à jamais gravés dans ma mémoire.

A mes très chers parents

Khaled et Louaï

Houda et Bilel

*« Cherchons comme cherchent ceux qui doivent
trouver et trouvons comme trouvent ceux qui
doivent chercher encore, car celui qui est arrivé
au terme ne fait que commencer »*

Résumé :

Le travail présenté dans ce mémoire décrit d'une manière détaillée les cartes auto organisatrices de Kohonen, certaines de leurs applications et leurs principales variantes. Ces cartes appartiennent à la famille des réseaux à apprentissage non supervisé et plus précisément à la famille des réseaux à compétition. Le réseau est constitué de deux couches : une couche d'entrée (vecteurs d'entrée) et une couche de sortie (les neurones). Les neurones sont représentés par des vecteurs code (des poids).

L'algorithme qui régit ces cartes est itératif. Son initialisation consiste à associer à chaque classe un vecteur code choisi de manière aléatoire. A chaque étape de l'algorithme, une observation est choisie au hasard, elle est comparée à tous les vecteurs codes à l'aide d'une distance donnée a priori. Le vecteur code représentant la plus petite distance déterminera la classe gagnante. On rapproche alors de l'observation les codes de la classe gagnante et des classes voisines.

Deux applications de l'algorithme de Kohonen ont été faites en utilisant des données du PNUD relatives à 118 pays et les données des 48 wilayas d'Algérie en se servant des résultats du RGPH de l'année 1998.

Mots clés : Classification, Réseaux de Neurones, cartes auto organisatrices de Kohonen, préservation du voisinage.

Abstract

The present work relates in details Self Organising Maps of Kohonen, some of their applications and their principals variants.

Self Organising Maps of Kohonen belong to the family of unsupervised training networks with competitive layer. The SOM network has two layers of nodes, the input layer and the Kohonen layer (output layer). The output nodes are chosen randomly.

The corresponding algorithm is iterative. At first, we chose an observation from the input data. Then, this observation is compared to all nodes with a particular distance. The node who presents the smallest distance is the winner. The nodes of the winner's neighbourhood and the winner are getting closer to the observation.

We applied the algorithm to a PNUD's data set of 118 countries and to the RGPH's data set for the 48 wilayate of Algeria.

Key words: Clustering analysis, Neural Networks, Self Organising Maps, Kohonen, preserving neighbourhood.

Liste des figures :

Figure I.1: Représentations graphiques de la distance euclidienne et de la distance de Manhattan.....	6
Figure I.2 : Agglomération progressive de 5 points.....	8
Figure I. 3 : Dendrogrammes d'une CAH selon les différentes distances.....	9
Figure I. 4 : Coupure de l'arbre hiérarchique par deux droites D et D'.....	10
Figure I.5:Illustration de l'algorithme des K-means.....	11
Figure I.6: Représentation géométrique de la première composante principale.....	15
Figure I.7: Représentation simultanée de la 1 ^{ère} et 2 ^{ème} composantes principales.....	16
Figure I.8 : Différents types de fonctions d'activation.....	21
Figure I.9 : Comparaison entre les différentes étapes de l'apprentissage supervisé et non supervisé dans le cas d'un problème de classification.....	23
Figure I.10 : fonctionnement d'un PMC.....	26
Figure I.11 : Réseau de Hopfield.....	27
Figure I.12 : Représentation générale d'un réseau RBF.....	28
Figure II.1 : Architecture du réseau de Kohonen.....	34
Figure II.2 : Exemple de formes possibles des cartes de Kohonen.....	35
Figure II.3 : Représentation des vecteurs poids dans l'espace d'entrée V (à gauche), et sur la grille A.....	35
Figure II.4.1 : Fonction de voisinage (ou d'activation) « chapeau mexicain ».....	37
Figure II.4.2 : Activité en chapeau mexicain autour du neurone le plus excité	37
Figure II.5 : différents types de fonctions de voisinage.....	38
Figure II.6 : Schéma de l'algorithme d'apprentissage de Kohonen.....	42
Figure II.7 : Conservation de l'ordre topologique.....	45
Figure III.1: Calcul géométrique de l'erreur.....	55
Figure III.2: Diagramme de Voronoï et Triangulation de Delaunay	55
Figure III.3: Différentes méthodes de visualisations de SOM : (a) U-matrices, (b) Histogrammes de densité des données, (c) les plans, et (d) le mappage de Sammon.	57
Figure III.4 : Visualisation des contenus des classes d'une carte (10 x 10).....	58
Figure III.5: Camembert d'une variable qualitative Q possédant trois modalités.....	62
Figure III.6 : Classification des 118 pays selon la variable IDH.....	77
Figure III.7 : Macro classes selon la classification hiérarchique.....	78
Figure III.8: Représentation des 118 pays classifiés sur la carte du monde.....	79
Figure III.9 : Dendrogramme de la classification hiérarchique sur les vecteurs code de l'application IDH.....	80
Figure III.10 : Résultats de la classification des 118 pays selon les deux variables PIB et Espérance de vie.....	82
Figure III.11 : Macro classes selon la classification hiérarchique.....	83
Figure III.12 : Dendrogramme de la classification hiérarchique sur les vecteurs code de l'application PIB / espérance de vie à la naissance.....	85
Figure III.13 : Déploiement de la carte à travers les différentes itérations.....	86
Figure III.14 : Représentation simultanée des données et des vecteurs codes après convergence.....	87
Figure III.15 : Classification des 48 wilayas d'Algérie.....	88

Figure III.16: Classification hiérarchique des wilayas d'Algérie.....	89
Figure III.17 : classification des wilayas d'Algérie en 8 classes.....	89
Figure III.18 : Dendrogramme de la classification hiérarchique sur les vecteurs code de l'application les 48 wilayas d'Algérie.....	91

TABLE DES MATIERES

INTRODUCTION.....	1
CHAPITRE I : Méthodes de classification et Réseaux de Neurones.....	4
1. Première partie : Méthodes de classification	5
I.1) La classification	5
– Définition	5
– Étapes préalables à toute méthode de classification.....	5
I.2) Méthodes de classification	6
A. Classification hiérarchique.....	6
– Les méthodes hiérarchiques ascendantes.....	7
B. Méthodes itératives.....	10
– La méthode des K-means ou k-moyennes	11
– Classification par nuées dynamiques.....	12
C. Méthodes factorielles.....	13
– Analyse en Composantes Principales (ACP).....	14
– Analyse factorielle des correspondances (AFC).....	17
2. Deuxième partie: Fonctionnement général d'un réseau de neurones.....	19
II. 1) Historique des réseaux de neurones.....	19
II.2) Définition des réseaux de neurones et caractéristiques.....	20
II.3) Méthodes d'apprentissage.....	21
– Définition du processus d'apprentissage	21
– Apprentissage supervisé.....	22
– Apprentissage non supervisé.....	22
– Etapes d'apprentissage.....	22
II.4) Règles d'ajustement des poids synaptiques.....	24
– La règle de Hebb.....	24
– La règle de corrélation en cascade.....	24
– La règle de Widrow-Hoff.....	24
– La règle de rétro-propagation.....	24
II.5) Quelques réseaux de neurones célèbres.....	25
A) Le perceptron	25
B) Le perceptron multicouches	25

C) Les réseaux de Hopfield	26
D) les réseaux à fonction radiale de base (RBF).....	27
E) les réseaux ART.....	28

CHAPITRE II : « Les cartes auto organisatrices de Kohonen ».....30

Introduction30

1) Quelques éléments historiques.....31

2) Fondements de base des Cartes de Kohonen.....33

3) Fonctionnement des Cartes de Kohonen.....34

a. Architecture.....34

b. Notions de voisinage.....34

c. Règle d'apprentissage.....38

d. Algorithme d'apprentissage.....41

4) Propriétés de la carte SOM.....43

• *Approximation de l'espace d'entrée.....*44

• *Variations statistiques de l'espace d'entrée*44

• *Conservation de l'ordre topologique.....*44

5) Principaux avantages et limitations des cartes de Kohonen.....45

6) Les extensions de la Cartes de Kohonen.....46

7) Classification de Kohonen par rapport aux autres méthodes.....50

CHAPITRE III : « Application des cartes auto organisatrices de Kohonen »..50

Première partie : « Mise en application et visualisation de l'algorithme de Kohonen ».....51

Introduction51

1) Mise en application de l'algorithme de Kohonen52

A) Réglages

B) Calcul de l'erreur.....

i. *Approche géométrique.....*54

ii. *Approche discrète.....*55

iii. *Approche par ensembles ordonnés.....*56

2) Méthodes d'analyse et de visualisation de la carte SOM.....	56
• Les U-matrices	56
• Les histogrammes de densité des données.....	56
• Les plans.....	57
• Le mappage de Sammon.....	57
3) Traitement de données qualitatives par l'algorithme Kohonen	58
3.1 Base de données constituée uniquement de variables qualitatives.....	59
1) l'Algorithme KACM.....	59
2) Traitements des individus.....	59
a) KACM_ind.....	60
b) KDISJ.....	60
3.2) Base de données constituée de variables quantitatives et qualitatives.....	62
4) Champs d'applications des cartes de Kohonen.....	62
1. Application à la validation des données.....	64
2. Adaptation de l'algorithme de Kohonen à des données incomplètes.....	65
 Deuxième partie : « Application de l'algorithme de Kohonen à des données réelles ».....	69
Introduction.....	69
1. Bases de données.....	69
2. Programmation de l'algorithme de Kohonen.....	70
3. Classification des 118 pays.....	73
- Classification selon l'Indicateur IDH.....	76
- Classification à deux dimensions : PIB/ espérance de vie à la naissance.....	81
4. Classification des 48 wilayas d'Algérie.....	87
Conclusion.....	91
Bibliographie.....	94
Annexe : Table de Burt et tableau disjonctif complet.....	98

INTRODUCTION

Introduction

La classification des données est identifiée comme une des problématiques majeures en extraction des connaissances. A travers les différentes méthodes de la classification, on cherche à construire des représentations simplifiées de données brutes et volumineuses, pour mettre en évidence les relations, les dominantes et la structure interne du *nuage* des observations. Deux grands groupes de techniques classiques sont généralement distingués: les *méthodes factorielles* et les *méthodes de classification automatiques*.

Plus récemment, depuis les années 80, de nouvelles méthodes sont apparues, connues sous le nom de *méthodes neuronales*. Elles proviennent de travaux pluridisciplinaires où se sont retrouvés des biologistes, des physiciens, des informaticiens et, plus récemment encore, des mathématiciens et notamment des statisticiens.

Ces méthodes ont rencontré rapidement un certain succès en particulier à cause de leur caractère de « boîte noire » et d'outil à tout faire.

En effet, dès leurs débuts, les réseaux de neurones formels simples pouvaient réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique).

Aujourd'hui, et après beaucoup de recherches qui ont permis leur évolution, les réseaux neuronaux sont très utilisés pour la classification, l'approximation, la prédiction ou la segmentation des données. Ils sont ainsi employés pour traiter des problèmes difficilement abordables par des méthodes linéaires ou par des équations algébriques.

Selon le type d'apprentissage utilisé, les réseaux de neurones sont répartis en deux grandes catégories : les réseaux de neurones à apprentissage supervisé et ceux à apprentissage non supervisé.

Dans l'apprentissage supervisé, les différentes classes de données doivent être déterminées *à priori*, alors que l'apprentissage non supervisé ne nécessite aucune connaissance sur l'appartenance des données aux différentes catégories.

Bien que l'utilisation des réseaux de neurones à apprentissage supervisé soit assez répandue, grâce notamment au *Perceptron Multicouches (PMC)*, le recours à cette approche n'est pas toujours possible.

Parfois, les classes ne peuvent pas être déterminées au préalable. Seule l'analyse des données réelles fera émerger des régularités permettant de les classer.

Dans de tels cas, une approche de classification non supervisée devra alors être mise en œuvre. Les approches de ce type ne nécessitent pas d'étiquetage préalable des données

d'apprentissage : les données proches seront regroupées et seront séparées des données qui leur sont différentes. En analysant les groupes ainsi formés, il sera alors possible de déterminer *à posteriori* des classes de données.

Il existe plusieurs réseaux d'apprentissage non supervisé : on peut citer le réseau de Hopfield, Learning Vector Quantization (LVQ) et les cartes auto organisatrices de Kohonen qui sont l'objet de ce mémoire.

Ces cartes (de Kohonen) sont devenues ces dernières années un des réseaux de neurones les plus célèbres. Elles sont utilisées dans de nombreux domaines d'analyse : données socio-économiques, financières, prédiction de séries temporelles, etc.

En plus d'une représentation synthétique de l'information, elles permettent une visualisation des données, souvent en deux ou trois dimensions. Cette représentation simple est souvent avantageuse lors de l'analyse des données de grande dimension.

Notre intérêt pour cette méthode de classification tient à la puissance de traitement qu'elle offre et à l'étendue de son champ.

Dans ce travail, nous définirons les concepts de bases de ce réseau de neurones, ses propriétés, ses avantages et ses limites, et nous identifierons certaines de ses applications. Une programmation de l'algorithme de ce réseau sera effectuée pour pouvoir l'appliquer à différentes bases de données.

Le mémoire comprend trois chapitres :

Le premier chapitre traitera des généralités relatives aux méthodes de classification et réseaux de neurones. Il sera articulé autour de deux sections.

La première définira ce que sont la classification et les différentes étapes qui la précèdent. Les méthodes de classification les plus importantes (Classification hiérarchique, *méthodes itératives*, *méthodes factorielles*) seront brièvement présentées.

La deuxième section abordera le fonctionnement général d'un réseau de neurones : y seront définis les réseaux de neurones, les différents types d'apprentissage et les étapes d'apprentissage. Une présentation de quelques réseaux célèbres sera faite également (le perceptron, le perceptron multicouches, les réseaux de Hopfield, les réseaux à fonction radiale de base (RBF), les réseaux ART).

Dans le deuxième chapitre, seront arborées en détail les cartes de Kohonen.

Dans un premier temps, la notion d'auto organisation sera explicitée et quelques éléments historiques qui sont à la base de l'algorithme d'auto organisation seront rappelés.

Dans un deuxième temps, l'architecture, le fonctionnement et les propriétés des cartes de Kohonen, ainsi que leurs avantages, leurs principales limites et leurs extensions les plus connues seront analysés.

Enfin, une comparaison de la classification de Kohonen avec d'autres méthodes de classification sera effectuée.

Le troisième et dernier chapitre sera composé de deux sections. La première sera consacrée à la mise en application de l'algorithme, aux méthodes de visualisation et aux champs d'applications de la carte. Deux applications bien particulières du réseau de Kohonen seront également décrites (application à la validation des données et application sur une base de données incomplète).

Dans la deuxième section, une application des cartes de Kohonen à des données réelles sera faite. Elle intégrera la programmation de l'algorithme à l'aide du logiciel Matlab.

Les bases de données utilisées concerneront des indicateurs de développement humain de 118 pays et des variables socio économiques des 48 wilayas de l'Algérie. Pour l'indicateur de développement humain, on s'intéressera à deux cas : le cas à une seule dimension en considérant la variable IDH, et le cas à deux dimensions en prenant en compte l'espérance de vie à la naissance et le PIB.

Une conclusion résumera l'essentiel du travail réalisé. Une annexe complète le mémoire. On y trouve les définitions relatives au tableau disjonctif complet et à la table de Burt.

CHAPITRE I

METHODE DE CLASSIFICATION ET

RESEAUX DE NEURONES

Généralités : Méthodes de classification et réseaux de neurones »

Après toute collecte d'information, il est nécessaire de faire une exploration des données pour définir les ressemblances ou les différences existantes entre les individus ou les variables.

Diverses méthodes mathématiques permettent de trouver les critères de reconnaissance de classes. Parmi ces méthodes les méthodes hiérarchiques, itératives, factorielles, ainsi que les réseaux de neurones

Dans ce chapitre seront abordées, dans un premier temps, les méthodes statistiques multidimensionnelles qu'englobe la classification (méthodes hiérarchiques, itératives et factorielles). La deuxième partie du chapitre sera consacrée pour les méthodes s'appuyant sur l'intelligence artificielle « les réseaux de neurones ».

Première partie : les méthodes de classification

I.1) La classification :

– Définition :

Les méthodes de classification sont destinées à produire des groupements d'objets ou d'individus décrits par un certain nombre de variables ou de caractères. Elles mettent en jeu une formulation et des calculs algorithmiques, et produisent des classes ou des familles de classes. Pour ce faire, on doit passer par des étapes essentielles. [25]

– Étapes préalables à toute méthode de classification :

Avant d'utiliser une des méthodes de classification, le choix des variables décrivant nos individus et le choix de la mesure de ressemblance entre deux individus s'impose. Ces choix sont importants puisqu'ils conditionnent le résultat final.

a) Choix des variables :

Il faut choisir les variables les plus pertinentes et un nombre de variables raisonnable. Chaque variable choisie peut être utilisée sous sa forme d'origine ou en la transformant en une variable normale par exemple.

b) Choix de la mesure de ressemblance :

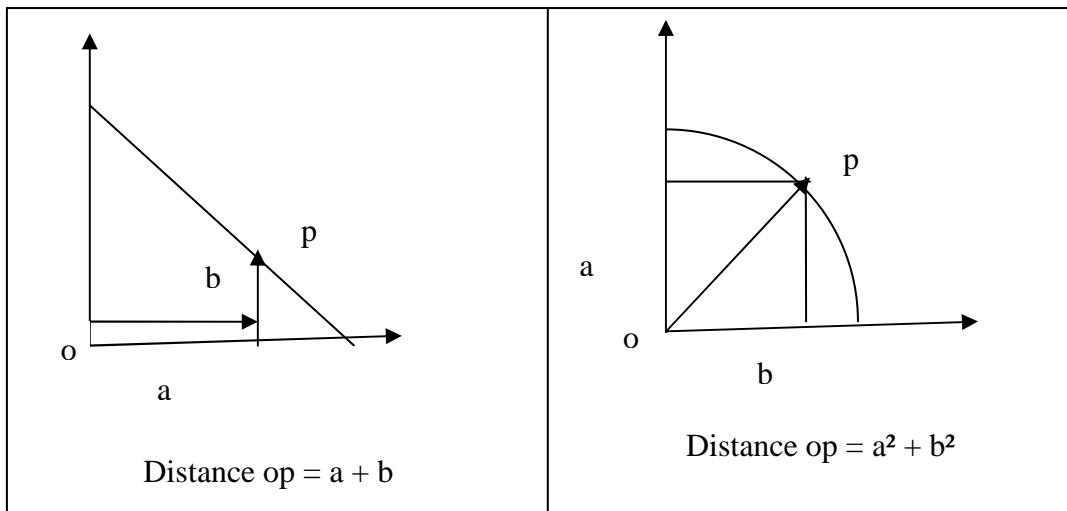
L'objectif de la classification étant de regrouper ensemble les individus «proches» les uns des autres, il faut définir cette notion de proximité. Il existe plusieurs mesures de la ressemblance entre deux individus. L'utilisation de l'une ou l'autre dépend notamment de la nature des données utilisées.

Pour des données numériques, on peut utiliser des distances comme :

a) Distance de Manhattan : $d_{ij} = \sum_{j=1}^p |x_{ij} - x_{i'j}|$;

b) Distance euclidienne : $d_{ij}^2 = \sum_{j=1}^p (x_{ij} - x_{i'j})^2$

Figure I.1 : Représentations graphiques de la distance euclidienne (à droite) et de la distance de Manhattan (à gauche).



c) Qualité d'une classification :

Pour réaliser une bonne classification à nombre de classes fixé n , on cherche à minimiser la somme des carrés intra-classes. On minimise donc :

$$SC_{\text{intra}}(I_1, I_2, \dots, I_N) = \sum_{i=1}^n \sum_{x_j \in \Gamma_i} \|x_j - \bar{\Gamma}_i\|^2$$

c) Où les $\bar{\Gamma}_i$ sont les centres de gravité des classes Γ_i . [25]

Les méthodes de classification sont très nombreuses et diverses. Les plus utilisées sont la Classification Hiérarchique, les méthodes itératives et les méthodes factorielles

I.2) Les méthodes de classification :

Cette section traitera des principales méthodes de classification et de leurs fondements de bases.

Le premier type de classification abordé est la classification hiérarchique.

A. Classification hiérarchique :

– **Le principe :**

L'objet des méthodes de classification hiérarchique est de fournir un ensemble de partitions plus ou moins fines obtenues par regroupements successifs de parties.

Pour cela on recherche les deux classes les plus proches, on les fusionne et on continue jusqu'à ce qu'il n'y ait qu'une seule classe. Cette *hiérarchie de partitions*, se présente sous la forme d'une structure arborescente que l'on appelle *dendrogramme*.

Il existe deux types de méthodes hiérarchiques : les méthodes hiérarchiques ascendantes et les méthodes hiérarchiques descendantes.

Le principe des méthodes ascendantes est de regrouper les différents éléments jusqu'à ce qu'il n'y ait plus qu'un seul groupe. Par contre, les méthodes hiérarchiques descendantes reposent sur le principe de diviser successivement des groupes d'individus en sous groupes.

Dans ce qui suit, nous allons nous intéresser seulement aux méthodes hiérarchiques ascendantes qui sont les plus utilisés.

- *Les méthodes hiérarchiques ascendantes :*

Le principe général de fonctionnement de ces algorithmes est de regrouper à chaque étape les individus (ou groupes d'individus) les plus «proches». Il existe plusieurs variantes de ces algorithmes. La différence réside dans la manière de calculer la proximité entre ces individus ou ces groupes d'individus.

L'algorithme fondamental de classification ascendante hiérarchique se déroule de la façon suivante :

Étape 1 : il y a n éléments à classer (qui sont les n individus);

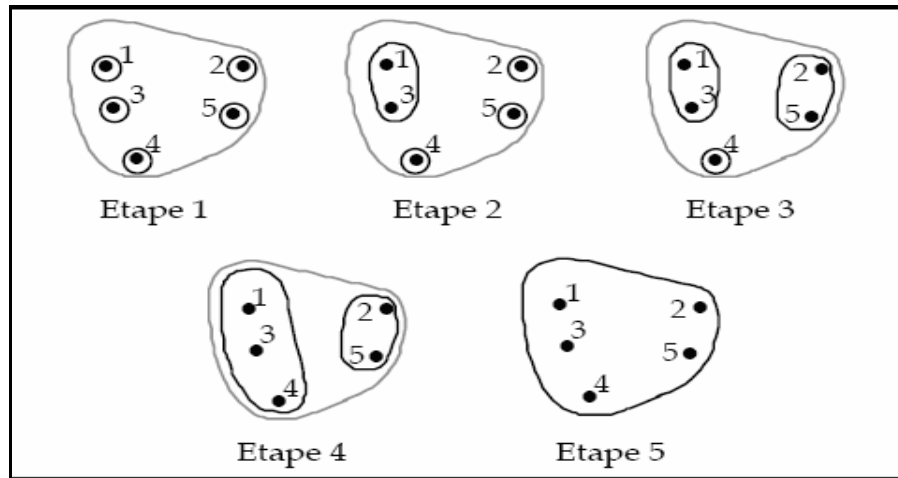
Étape 2 : on construit la matrice de distances entre les n éléments et on cherche les deux éléments les plus proches, que l'on agrège en un nouvel élément. On obtient une première partition à $n-1$ classes;

Étape 3 : on construit une nouvelle matrice des distances qui résultent de l'agrégation, en calculant les distances entre le nouvel élément et les éléments restants (les autres distances sont inchangées). On se trouve dans les mêmes conditions qu'à l'étape 1, mais avec seulement $(n-1)$ éléments à classer et en ayant choisi un critère d'agrégation. On cherche de nouveau les deux éléments les plus proches, que l'on agrège. On obtient une deuxième partition avec $n-2$ classes et qui englobe la première;

Étape m : on calcule les nouvelles distances, et l'on réitère le processus jusqu'à n'avoir plus qu'un seul élément regroupant tous les objets et qui constitue la dernière partition.

Cette procédure est illustrée dans la figure I.2. Les objets à classer sont représentés par cinq points :

Figure I.2 : Agglomération progressive de 5 points



– Distances entre éléments et entre groupes :

On suppose au départ que l'ensemble des individus à classer est muni d'une distance 1. On construit alors une première matrice de distances entre tous les individus.

Une fois constitué un groupe d'individus, il convient de se demander ensuite sur quelle base on peut calculer une distance entre un individu et un groupe et par la suite une distance entre deux groupes. Cette distance entre groupements pourra en général se calculer directement à partir des distances des différents éléments impliqués dans le regroupement.

Par exemple, si x , y , z sont trois objets, et si les objets x et y sont regroupés en un seul élément noté h , on peut définir la distance de ce groupement à z par la plus petite distance des divers éléments de h à z :

$$d(h, z) = \text{Min} \{d(x, z), d(y, z)\}$$

Cette distance s'appelle le *saut minimal* (*single linkage*) et constitue un critère d'agrégation. On peut également définir la distance du *saut maximal* (ou diamètre) en prenant la plus grande distance des divers éléments de h à z :

$$d(h, z) = \text{Max} \{d(x, z), d(y, z)\}$$

Une autre règle simple et fréquemment employée est celle de la *distance moyenne* ; pour deux objets x et y regroupés en h : [27]

$$d(h, z) = \{d(x, z) + d(y, z)\}/2$$

d) Dendrogrammes :

L'intérêt des dendrogrammes de la CAH est qu'il peut donner une idée du nombre de classes existant effectivement dans la population.

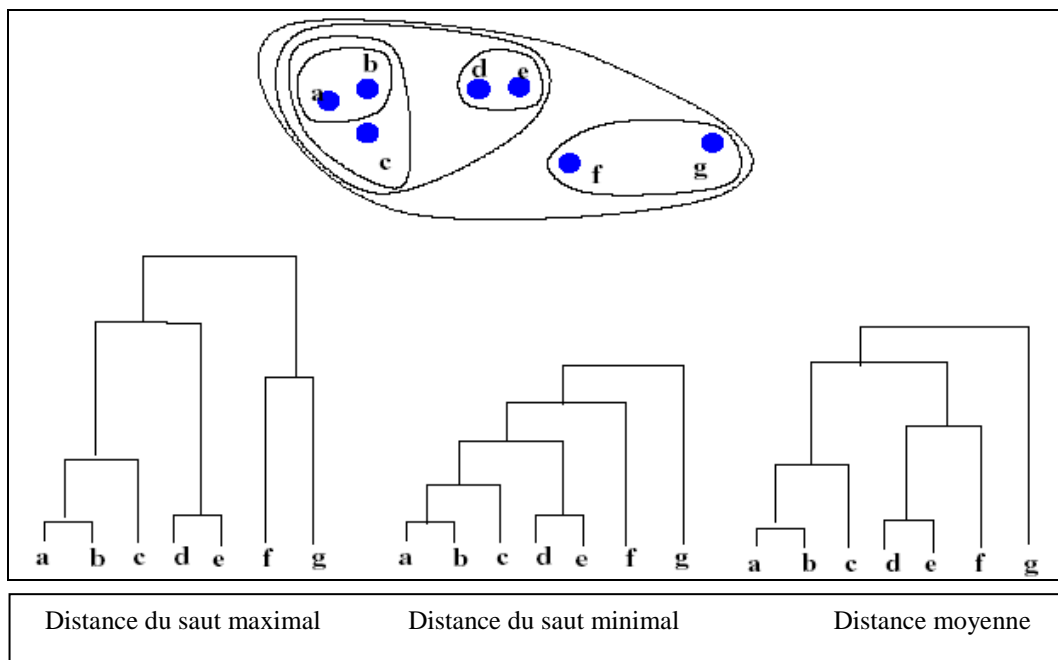
L'algorithme de la construction d'un dendrogramme consiste à ordonner les observations de telle sorte qu'il n'y ait aucun croisement entre les diverses branches du dendrogramme.

L'algorithme s'énonce comme suit :

- i. placer les observations selon un ordre quelconque de gauche à droite ;
- ii. s'il ne reste qu'un seul groupe, on termine.
- iii. prendre les groupes compris entre les groupes qui fusionnent et les déplacer rigidement à la droite de la dernière observation du groupe fusionné situé le plus à droite.
- iv. retourner à ii. [34]

La figure suivante présente les dendrogrammes relatifs aux trois mesures de distances définies plus haut :

Figure I. 3 : Dendrogrammes d'une CAH selon les différentes distances.



Connaissant les différentes manières d'agréger les éléments, il reste maintenant à choisir une partition.

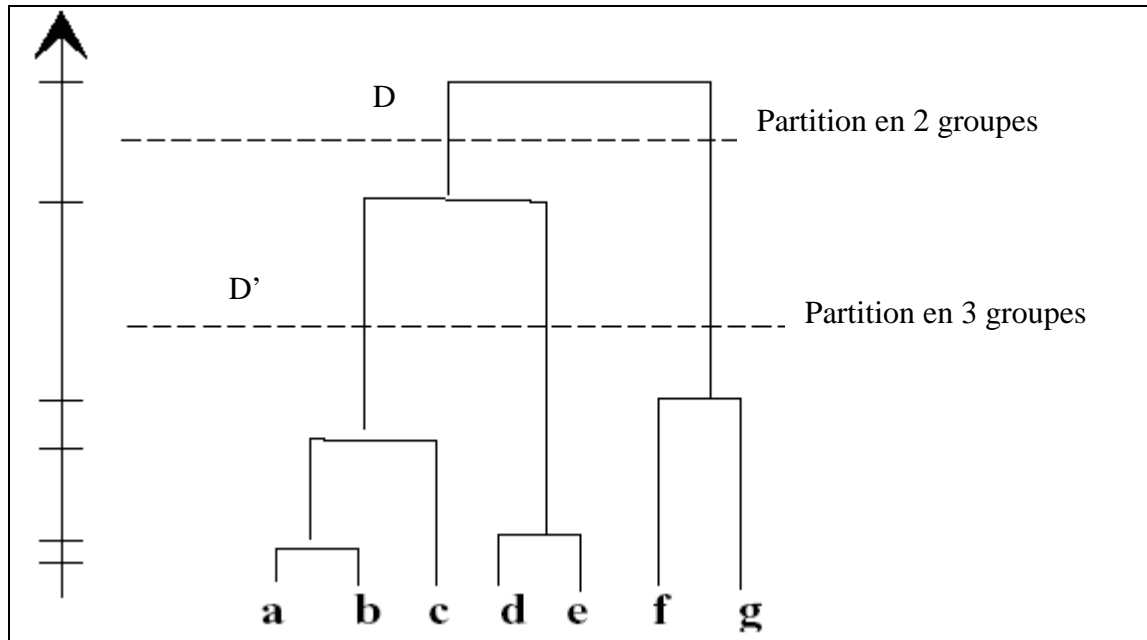
- *Choix du nombre de classes par coupure de l'arbre hiérarchique :*

Toute coupure de l'arbre par une droite horizontale donne une partition de l'ensemble de départ. Ainsi dans la figure suivante, la droite D coupe cet ensemble en deux classes

Partition en 2 classes

composées des éléments : {a, b, c, d, e} et {f, g}, et la droite D' en trois classes : {a, b, c}{d,e}{f, g}.

Figure I. 4 : Coupure de l'arbre hiérarchique par deux droites D et D'



Pour le choix du niveau de coupure, la référence [34] donne une revue détaillée des différentes manières de procéder.

L'un des désavantages des méthodes hiérarchiques est qu'il est impossible de réallouer un point qui aurait été mal classé à l'une des étapes de l'algorithme. De plus, les méthodes hiérarchiques ne sont pas adaptées aux vastes ensembles.

Ces méthodes ont aussi l'inconvénient de changer de partitions en changeant l'ordre de présentation des observations d'entrée. Les méthodes dites itératives permettent de pallier à cette lacune. Elles feront l'objet de la section suivante.

B. Méthodes itératives :

Ces méthodes utilisent des algorithmes au cours desquels les points sont affectés, ou réaffectés, à une classe donnée. La plupart du temps, le nombre de classes est fixé a priori.

L'idée générale consiste à choisir arbitrairement une partition initiale puis de déplacer les objets d'une classe à l'autre pour obtenir une meilleure partition. On cherche ainsi à optimiser la répartition des individus dans les différentes classes.

Parmi ces méthodes, on peut citer : la méthode des *K-means* ou *k-moyennes* et la *classification par nuées dynamiques*.

– **La méthode des K-means ou k-moyennes :**

La méthode des K-means fait partie de la famille d’algorithmes de la classification dit à « *centres mobiles* ». C’est une méthode de partition des données par minimisation de la variance intra-groupe.

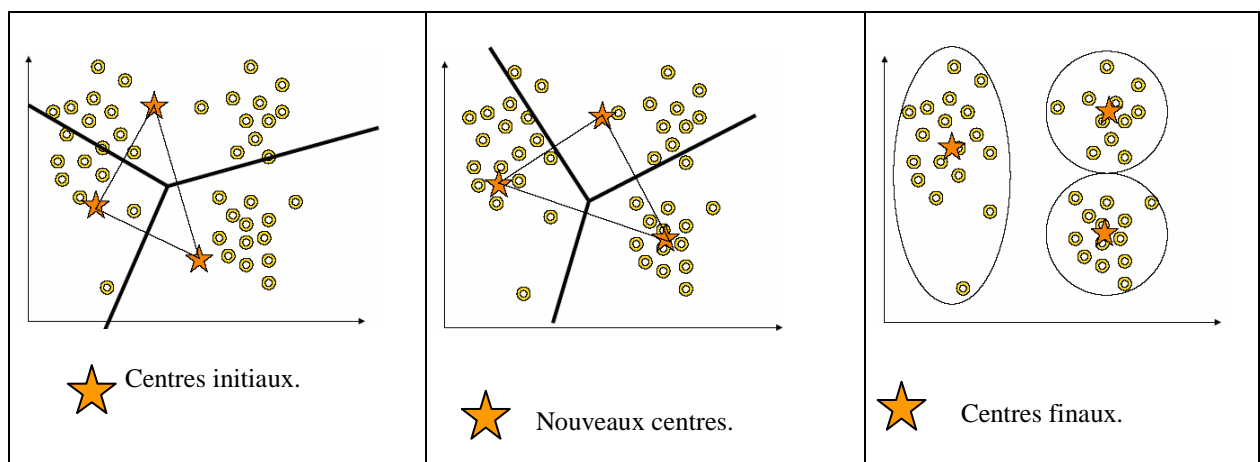
La répartition des données en classes est une tâche non supervisée qui nécessite seulement la précision du nombre de classes k qu’on veut obtenir. Cette procédure consiste en une succession d’étapes :

- 1) Tirage de k individus au hasard sans remise. Ces individus ainsi déterminés représentent k centres provisoires de classes. Ces k centres induisent, par affectation des individus au centre le plus proche, une partition $P^{(0)}$ de l’ensemble d’entrée en k classes ;
- 2) Calcul des centres de gravité des différentes classes. Ces derniers, constituent les k nouveaux centres de classes ;
- 3) Affectation des observations aux nouveaux centres de classes comme dans le point (1);

On arrête l’algorithme quand deux itérations successives fournissent la même partition.

Le point faible de cet algorithme est que le résultat dépend du point de départ choisi pour l’initialisation des centres de classes.

Figure 1. 5 : Illustration de l’algorithme des K-means.



La méthode des K-means est rapide, efficace et relativement extensible dans le traitement d’ensembles de taille importante. Cependant elle présente plusieurs inconvénients comme le fait qu’elle n’est applicable que dans le cas où la moyenne des objets est définie et elle

est incapable de traiter les données bruitées. Aussi, le nombre de clusters doit être spécifié a priori et les points isolés sont mal gérés. [51]

- Classification par nuées dynamiques:

La méthode des nuées dynamiques peut être considérée comme une généralisation de l'algorithme des k-moyennes. En effet, chaque classe n'est plus représentée par son barycentre, mais par un sous ensemble de classe, appelé **noyau**, constitué des individus les plus centraux.

Les deux fonctions de base sur lesquelles repose l'algorithme de cette méthode sont les suivantes :

a) La fonction de réallocation :

Elle affecte chaque individu du nuage des observations aux centres d'attractions que forment les noyaux.

Elle est définie par l'équation :

$$\pi(X, A_j) = \frac{1}{n_j} d(X, X')$$

où n_j est le nombre d'éléments du noyau

b) La fonction de recentrage :

Elle recalcule les nouveaux noyaux à partir des classes déjà formées. Elle est définie par l'équation :

$$v(A_j, P_j) = \frac{1}{N_j} \sum_{X \in P_j} \pi(X, A_j)$$

où N_j est le nombre d'éléments de la classe.

Déroulement de l'algorithme :

- e) Initialisation aléatoire des K premiers noyaux ;
 - f) Affectation: Calcul de la classe de chaque point du nuage ;
 - g) Mise à jour des centres (ou attributs des classes) ;
 - h) Calcul des nouveaux centres de chaque classe (barycentre) ;
- Test de convergence.

L'exécution de l'algorithme se termine lorsque le partitionnement n'évolue plus, c'est-à-dire lorsque le critère d'inertie intra-classe, défini par l'équation, converge.

$$I_w = \sum_{j=1}^K \sum_{X \in P_j} d^2(X, G_j)$$

Où G_j est le centre de gravité de la classe.

Le résultat change selon le choix des conditions initiales. Il faut donc exécuter plusieurs fois l'algorithme et comparer les résultats de manière à extraire les classes stables, c'est à dire à dégager ce qu'on appelle des formes fortes. [42]

Ces méthodes itératives de classification sont les principales, mais il en existe encore d'autres. On peut aussi utiliser des combinaisons de ces méthodes pour créer des algorithmes mixtes.

Un des problèmes de ces méthodes est le coût en temps de calcul nécessaire à leur exécution qui les rend difficilement utilisables pour de très grands jeux de données.

Les méthodes précédentes avaient pour objet de rassembler l'ensemble des individus en formant des groupes homogènes.

Dans ce qui suit, nous verrons une autre manière de classer : réduire le nombre de variables en les résumant par un petit nombre de composantes synthétiques en utilisant les méthodes dites *factorielles*.

C. Méthodes factorielles ;

Les méthodes factorielles sont apparues au début des années 30 et ont été développées dans les années 60 par l'exploitation des aspects géométriques et des présentations graphiques. Dans la mesure où ce sont des méthodes descriptives, elles ne s'appuient pas sur un modèle probabiliste, mais dépendent d'un modèle géométrique.

Ces méthodes visent à résumer de manière synthétique de vastes ensembles de valeurs numériques par le biais d'un nombre plus restreint de variables artificielles nouvelles, les *facteurs*. Chaque facteur représente un groupement de traits inter-corrélés. Les facteurs sont généralement représentés sous forme de visualisations graphiques où les objets à décrire deviennent des points sur un axe ou dans un plan. Sur ces graphiques, on voit, au sens propre du terme, des regroupements, des oppositions, des tendances directionnelles, impossible à discerner sur un grand tableau de nombres même après un examen prolongé.

Les méthodes d'analyse factorielle les plus connues sont *l'Analyse en Composantes Principales* (ACP) qui permet de projeter des données quantitatives et *l'Analyse des*

Correspondances qui permet d'analyser les relations entre les différentes modalités de variables qualitatives croisées (Analyse Factorielle des Correspondances, AFC).[40]

– **L'analyse en Composantes Principales (ACP) :**

Chercher à réduire la dimensionnalité d'un ensemble de données est historiquement associé à l'Analyse en Composantes Principales.

L'Analyse en Composantes Principales (ACP) consiste à chercher un sous-espace de dimension inférieure à celle de l'espace initial où l'on pourra projeter les données en perdant un minimum d'information.

Supposant qu'on ait besoin de q variables pour prendre en compte toute la variabilité d'un système. L'application de l'ACP permettra de décrire le système à l'aide de $q < p$ composantes en expliquant le maximum de cette variabilité.

Ceci permet :

- une réduction des données à q nouveaux descripteurs ;
- une visualisation des données à 2 ou 3 dimensions (si $q=2$ ou 3) ;
- une interprétation des données : liaisons inter-variables. [28]

– **Recherche des composantes principales :**

Soient $C_1, C_2, \dots, C_k, \dots, C_q$ nos q composantes. Chaque composante C_k est une nouvelle variable qui est définie par une combinaison linéaire des variables d'origines X_1, \dots, X_p :

$$C_k = a_{1k} X_1 + a_{2k} X_2 + \dots + a_{pk} X_p$$

Les C_k doivent être non corrélées 2 à 2, de variance maximale et d'importance décroissante.

○ *Première composante principale :*

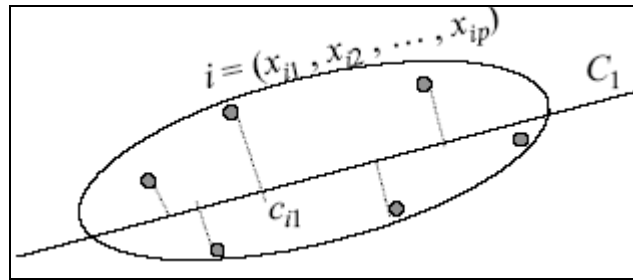
C_1 représente la 1^{ère} composante principale et doit être de variance maximale.

Géométriquement : C_1 détermine une nouvelle direction dans le nuage de points qui suit l'axe d'allongement (étirement) maximal du nuage.

c_{i1} = coordonnée du point i sur l'axe C_1 et c 'est la projection de xi sur C_1

$$c_{i1} = \sum_{j=1}^p a_{1j} x_{ij}$$

Figure I.6: Représentation géométrique de la première composante principale.



Comme C_1 doit être de variance maximale, les projections c_{i1} sont les plus dispersées possible.

La droite C_1 est la droite passant par le centre de gravité réalisant le meilleur ajustement possible du nuage de points, c-à-d celle qui conserve au mieux la distance entre les points (après projection). Et donc C_1 est la droite de projection assurant une distorsion minimale.

Comme la variance totale du système est égale à la somme des valeurs propres, chercher la projection de variance maximale C_1 revient donc à projeter les données sur le vecteur propre ayant la valeur propre la plus élevée de la matrice des variances-covariances de la matrice de données.

Cette valeur propre mesure la part de la variance maximale expliquée par le premier axe factoriel.

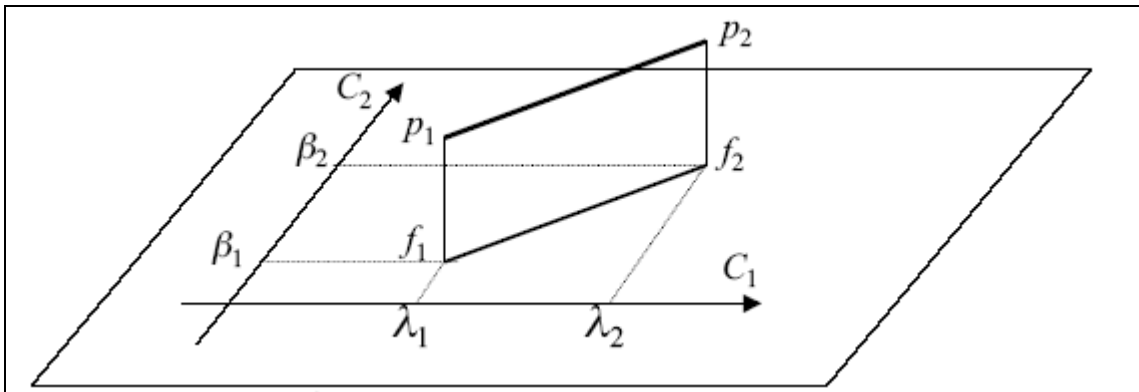
○ *Deuxième composante principale :*

C_2 représente la 2^{ème} composante principale. Elle doit être orthogonale à C_1 et doit avoir la 2^{ème} plus grande variance.

Géométriquement :

C_2 détermine une droite perpendiculaire à C_1 (au point \mathbf{g}), suivant un axe (perpendiculaire au 1^{er}) d'allongement maximum.

Les deux composantes C_1 et C_2 déterminent le plan principal : le meilleur plan de projection (de distorsion minimum).

Figure I.7: Représentation simultanée de la 1^{ère} et 2^{ème} composantes principales.

- C_1 est telle que la moyenne des $d^2(\lambda_i, \lambda_{i'})$ max.
- C_2 est \perp à C_1 telle que la moyenne des $d^2(\beta_i, \beta_{i'})$ max.
- C_1 et C_2 déterminent le plan tel que $d^2(\mathbf{f}_i, \mathbf{f}_{i'})$ soit maximum.

La troisième composante principale C_3 , représente la droite \perp à C_1 et C_2 (par \mathbf{g}) et possède la troisième plus grande variance. Ainsi de suite pour les autres composantes principales de degrés plus élevé.

Chaque valeur propre mesure la part de la variance expliquée par l'axe factoriel y correspondant.

Les k premières composantes principales sont choisies tels que la somme de leurs valeurs propres représente plus de 60 % de la variance totale. Soit λ_i la valeur propre de la composante i et il existe n composantes principales. $\mathbf{k} < n$ est choisi, tel que :

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \geq 0.6$$

Malheureusement, l'ACP n'effectue qu'une projection linéaire des entrées, ce qui la rend inapplicable lorsque les dépendances sont non-linéaires.

De nombreuses méthodes ont été proposées afin de pallier cette limitation ; deux voies principales se sont dégagées : la première basée sur des méthodes algébriques, repose sur l'idée de considérer que l'ensemble de données est localement linéaire, la seconde basée sur des méthodes neuronales et notamment les cartes auto-organisatrices de Kohonen. [35]

L'ACP a permis de réduire la dimension de la base de données. Si par contre, on cherchait à regrouper les individus de la base en groupes et que ces individus soient décrits

par deux variables qualitatives. Une des manières adéquates pour le faire serait d'utiliser l'analyse factorielle des correspondances.

– **Analyse factorielle des correspondances (AFC) :**

L'analyse factorielle des correspondances (AFC), ou analyse des correspondances simples, est une méthode exploratoire d'analyse des tableaux de contingence.

Soient deux variables nominales X et Y, comportant respectivement I et J modalités. On a observé les valeurs de ces variables sur une population et on dispose d'un tableau de contingence à I lignes et J colonnes.

Les valeurs de ce tableau seront notées n_{ij} (nombre d'individus partageant la modalité i de la première variable ligne et la modalité j de la seconde variable colonne). L'effectif total sera noté n.

On définit successivement

- la table F des fréquences relatives, aux entrées :

$$f_{ij} = \frac{n_{ij}}{n} \text{ avec } n = \sum_{i,j} n_{i,j}$$

- Les marges aux entrées :

$$f_{i..} = \sum_j f_{ij} \text{ et } f_{.j} = \sum_i f_{ij}$$

- La table P_R des I profils lignes de somme 1, aux entrées :

$$P_{ij}^R = \frac{f_{ij}}{f_{i..}}$$

- La table P_C des J profils colonnes de somme 1, aux entrées :

$$P_{ij}^C = \frac{f_{ij}}{f_{.j}}$$

Ces profils sont en fait des distributions de probabilités conditionnelles, il est d'usage de leur appliquer la distance du chi-deux, définie comme suit :

- Pour les lignes :

$$X^2(i, i') = \sum_j \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i..}} - \frac{f_{i'j}}{f_{i'..}} \right)^2 = \sum_j \left(\frac{f_{ij}}{\sqrt{f_{.j} f_{i..}}} - \frac{f_{i'j}}{\sqrt{f_{.j} f_{i'..}}} \right)^2$$

- Pour les colonnes :

$$X^2(j, j') = \sum_i \frac{1}{f_{i..}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{i'j'}}{f_{.j'}} \right)^2 = \sum_i \left(\frac{f_{ij}}{\sqrt{f_{i..} f_{.j}}} - \frac{f_{i'j'}}{\sqrt{f_{i..} f_{.j'}}} \right)^2$$

Chaque ligne i est pondérée par $f_{i\cdot}$ et chaque colonne est pondérée par $f_{\cdot j}$.

Comme en ACP, on s'intéresse alors aux directions de "plus grande dispersion" de chacun de ces nuages de points. Mais, pour mesurer la "distance" entre deux individus, la *métrique du χ^2* est utilisée au lieu de la distance habituelle (distance *euclidienne*).

Pour le calcul de l'inertie de ces deux ensembles de profils :

- Inertie totale des profils lignes :

$$\sum_i f_{i\cdot} X^2(i, i')$$

- Inertie totale des profils colonnes :

$$\sum_j f_{\cdot j} X^2(j, j')$$

On remarque que les inerties totales ligne et colonnes sont égales.

L'inertie totale sera notée I :

$$I = \sum_{i,j} \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}} = \sum_{i,j} \frac{f_{ij}^2}{f_{i\cdot} f_{\cdot j}} - 1$$

Chaque ligne du tableau des fréquences lignes peut être vue comme la liste des coordonnées d'un point dans un espace à J dimensions. On obtient ainsi le nuage des individus-lignes. On définit de même le nuage des individus-colonnes à partir du tableau des fréquences colonnes.

Le coefficient $f_{i\cdot} f_{\cdot j}$ représente le "poids théorique" de chaque cellule dans le tableau. La somme de ces coefficients vaut 1.

Le choix du nombre d'axes factoriels à conserver se fait comme dans le cas de l'ACP.

La première valeur propre est systématiquement égale à 1 et les autres valeurs propres sont des nombres positifs inférieurs à 1 et leur somme est égale à χ^2 .

Par contre, contrairement à l'ACP, les graphiques factoriels pourront être construits en faisant figurer sur un même graphique les individus lignes et les individus colonnes. [28]

Dans cette première section nous avons abordé les principales méthodes de classification. Comme la classification est un domaine riche du connexionnisme ; il est donc indispensable de décrire le fonctionnement général d'un réseau de neurones ainsi que les principaux types de réseaux.

Deuxième partie : Réseaux de neurones

II.1) Historique des réseaux de neurones :

Disponibles aujourd'hui au sein de la plupart des grands logiciels de statistique, les méthodes neuronales sont apparues au début des années 40 avec la définition du neurone formel de W. Mc Culloch et W. Pitts. Ceux sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique).

En 1949, D. Hebb présente une règle d'apprentissage qui sera plus tard la source d'inspiration de nombreux modèles de réseaux.

Le modèle du perceptron ne sera développé qu'en 1958 par F. Rosenblatt. C'est le premier système artificiel, inspiré du système visuel, capable d'apprendre par expérience. F. Rosenblatt construit le premier neuro ordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance de formes.

Dans la même période, Le modèle de L'Adaline (ADaptive LINar Element) a été présenté par B. Widrow et Hoff. Dans sa structure, ce modèle ressemble au Perceptron, cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétropropagation de gradient très utilisé aujourd'hui avec les Perceptrons multicouches.

En 1969, M. Minsky et S. Papert publient un ouvrage qui met en exergue les limitations théoriques du perceptron. Limitations alors connues, notamment concernant l'impossibilité de traiter par ce modèle des problèmes non linéaires. Ils étendent implicitement ces limitations à tous modèles de réseaux de neurones artificiels. Leur objectif est atteint, il y a abandon financier des recherches dans le domaine surtout aux U.S.A.

Ce n'est qu'en 1982 et grâce à J. J. Hopfield que le renouveau d'intérêt pour les réseaux de neurones artificiels a eu lieu. Alors que les auteurs s'acharnent jusqu'alors à proposer une structure et une loi d'apprentissage, puis à étudier les propriétés émergentes ; J. J. Hopfield fixe préalablement le comportement à atteindre pour son modèle et construit à partir de là, la structure et la loi d'apprentissage correspondant au résultat escompté.

Dans la même année, T. Kohonen présente ses travaux sur les mémoires associatives et donne son nom à un réseau de neurones à apprentissage non supervisé basé sur l'auto organisation « les cartes de Kohonen ».

Les méthodes neuronales ont rencontré rapidement un certain succès en particulier à cause de leur caractère d'outil à tout faire, ayant de très nombreux domaines

d'applications. En effet, les réseaux de neurones sont utilisés pour la prévision et la modélisation, la compression des données, la reconnaissance des formes et surtout dans la classification. [41]

II.2) Définition des réseaux de neurones et caractéristiques:

Un réseau de neurones est un système statistique de résolution d'un système d'équations non linéaires. Il s'inspire métaphoriquement d'un vrai réseau de neurone dans la mesure où des noeuds sont affectés d'un poids, sont répartis sous forme de couches et transfèrent un signal d'une couche à une autre avec une couche d'entrée (données d'entrée) et une couche de sortie (résultat final). [42]

Il est constitué d'un ensemble de processeurs élémentaires de traitement, *les neurones*, qui sont connectés les uns aux autres au moyen des connexions qui les relient.

On appelle « *état* » du neurone la valeur de sortie du neurone. Un neurone calcule son état à partir des informations qui lui parviennent de l'extérieur du réseau ou bien de l'état des neurones auxquels il est connecté.

On peut alors caractériser un réseau de neurones par : [3]

- _ Un ensemble de connexions, appelées les connexions synaptiques qui permettent d'accepter les signaux d'entrée et qui sont définies par des poids synaptiques w_{nj} déterminant l'effet du signal par le neurone n sur le neurone j ;
- _ Une fonction de combinaison ou un additionneur qui réalise la somme pondérée net_j . La somme pondérée est égale à :

$$net_j = \sum_n w_{nj} x_n \quad \text{Où } x_n \text{ est l'entrée.}$$

Cette somme d'activation pondérée converge vers le neurone j ;

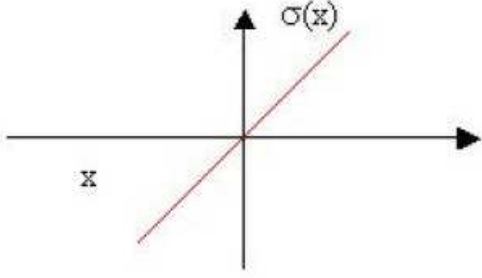
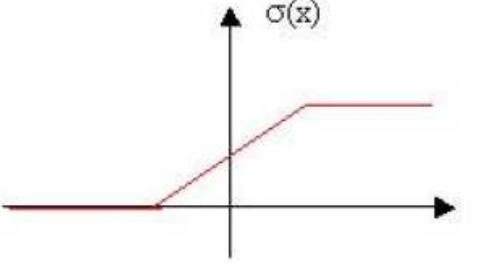
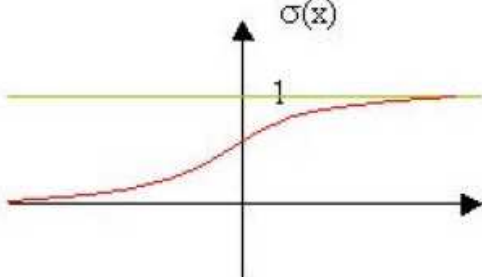
- _ Une fonction d'activation qui anime le neurone en déterminant son activation ;
- _ Une activation o_j , équivalente de la sortie du neurone. Elle est égale à:

$$o_j = \Phi\left(\sum_{n=1}^n w_{nj} x_n + \Theta_j\right)$$

Où Θ_j est le seuil du neurone j . Le seuil est une sorte de « poids local ».

Différents types de fonctions d'activation peuvent être utilisés : [20]

Figure I.8 : différents types de fonctions d'activation.

<p>Dans ce modèle, la fonction d'activation σ est la fonction identité. S'il a l'avantage d'être simple, il n'a que peu de rapport avec la réalité (le signal de sortie est non borné, linéaire par rapport aux signaux d'entrée, ce qui ne correspond pas du tout au fonctionnement des neurones biologiques).</p>	
<p>Le modèle linéaire par morceaux est plus réaliste en faisant apparaître des phénomènes de saturation.</p>	
<p>Le modèle sigmoïde est le modèle le plus utilisé dans la pratique</p>	

Pour déterminer les poids synaptiques du réseau, on a recours à des méthodes dites d'apprentissage.

II.3) Méthodes d'apprentissage :

■ Définition du processus d'apprentissage :

Le processus d'apprentissage consiste à faire évoluer les poids des connexions des neurones, en fonction des échantillons de référence présentés au réseau, jusqu'à l'obtention d'un état d'équilibre pour le réseau.

On distingue deux types d'apprentissage :

- *L'apprentissage supervisé :*

Dans ce type d'apprentissage, le réseau s'adapte par la comparaison entre le résultat qu'il a calculé, en fonction des entrées fournies, et la réponse attendue en sortie.

Ainsi, le réseau va se modifier jusqu'à ce qu'il trouve la bonne sortie, c'est-à-dire celle attendue, correspondant à une entrée donnée.

Après l'apprentissage, le réseau est testé en lui donnant seulement les valeurs d'entrée, et en regardant si le résultat obtenu est proche du résultat désiré.

Le perceptron multicouches est un exemple d'apprentissage supervisé.

- *L'apprentissage non supervisé :*

Soit un ensemble de données représenté par des vecteurs de grande dimension, l'apprentissage consiste à les regrouper selon des critères de ressemblance qui sont inconnus a priori. Dans ce type d'apprentissage, la catégorie de chaque forme fournie à l'entrée pour apprendre le réseau n'est pas connue, il évolue librement jusqu'à ce qu'il se stabilise.

Les cartes auto-organisatrices sont une technique à apprentissage non supervisé permettant de retrouver l'information "prégnante" dans un ensemble de données de haute dimensionnalité.

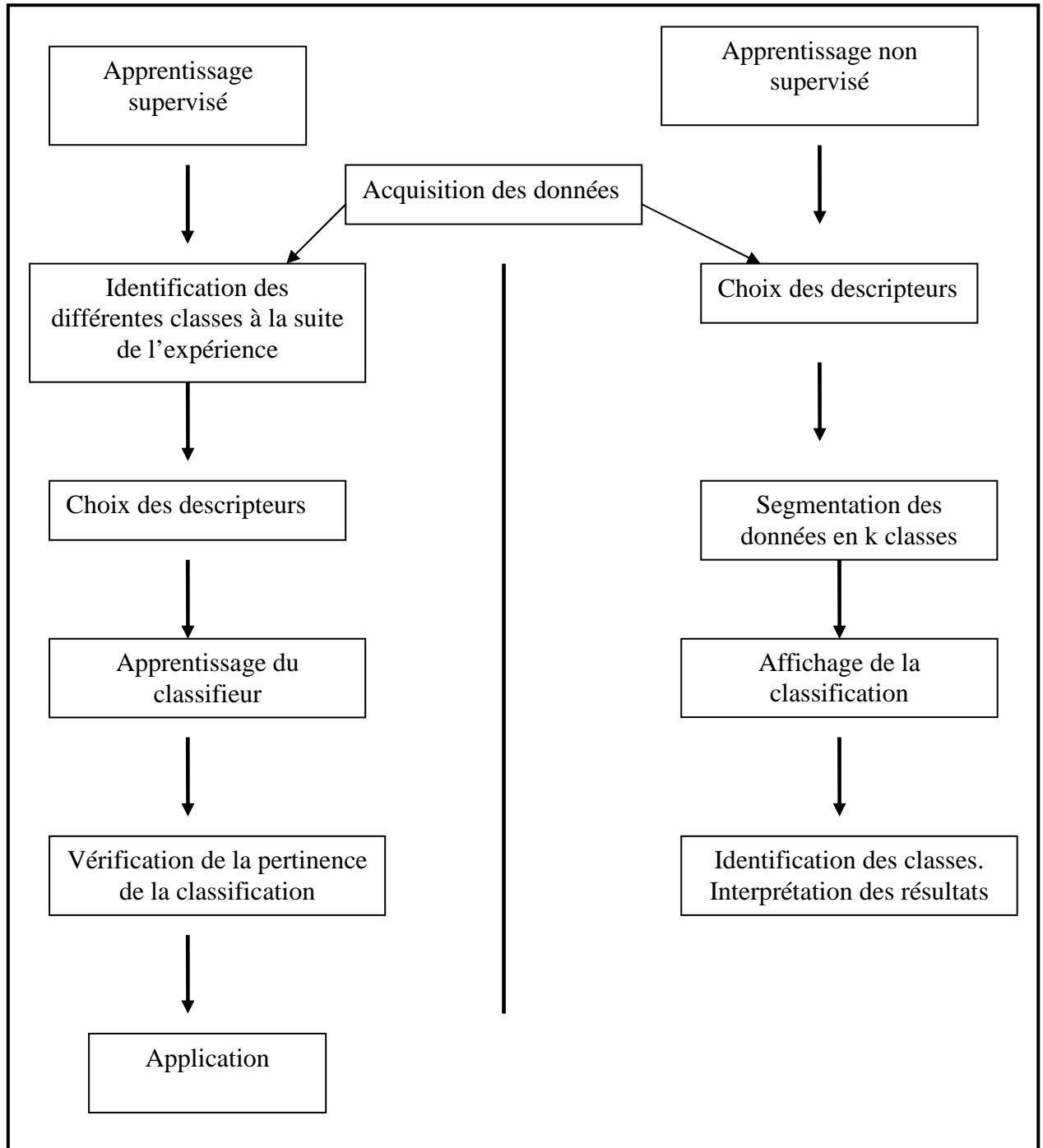
• *Etapas d'apprentissage :*

la phase d'apprentissage des différents modèles de réseaux de neurones suit le plus souvent le schéma suivant :

- Initialisation de tous les poids à des valeurs aléatoires (valeurs généralement comprises entre 0 et 1) pour tous les neurones du réseau ;
- Présentation d'un vecteur d'entrée provenant d'un ensemble d'échantillons choisis spécifiquement pour la phase d'apprentissage ;
- Calcul des valeurs de sortie correspondantes ;
- Modifications des poids en fonction de la différence entre les valeurs obtenues et celles désirées ;
- Définition d'un critère d'arrêt.

La différence d'un modèle de réseau à un autre porte généralement sur le 4^{ème} point. [25]

Figure I.9 : Comparaison entre les différentes étapes de l'apprentissage supervisé et non supervisé dans le cas d'un problème de classification



Les deux types d'apprentissage ont tous deux pour but l'ajustement des poids de connexions entre les neurones, en s'accordant de certaines *règles*.

II.4) Règles d'ajustement des poids synaptiques :

Les règles les plus répandues sont :

_ La règle de Hebb :

Cette règle permet de modifier la valeur des poids synaptiques en fonction de l'activité des unités qui les relient. Le principe de cette règle est le suivant :

- un vecteur est présenté au réseau ;
- les neurones calculent leurs sorties ;
- les sorties sont propagées et on itère jusqu'à la convergence.

Si deux unités s'activent en même temps, la connexion qui les lie est renforcée (c'est une connexion excitatrice) sinon elle est affaiblie (c'est une connexion inhibitrice) ;

_ La règle de corrélation en cascade :

C'est une technique d'apprentissage qui ajoute progressivement des neurones cachés au réseau jusqu'à ce que l'effet bénéfique de ces nouveaux neurones ne soit plus perceptible. Cette règle suit les deux étapes suivantes :

1. Le système est entraîné par un apprentissage classique qui s'effectue premièrement dans un petit réseau ne possédant pas de couche cachée.
2. Un petit groupe de neurones supplémentaires est entraîné de manière à diminuer l'erreur résiduelle du réseau en modifiant les poids de ces neurones. Le neurone qui réussit le mieux est ensuite retenu, et intégré au réseau. Puis l'étape 1 est relancée, pour permettre au réseau de s'adapter à la nouvelle ressource.

- La règle de Widrow-Hoff :

Si la sortie d'une cellule de décision est égale à la sortie désirée, le poids de la connexion entre ce neurone et le neurone d'association qui lui est connecté n'est pas modifié. Dans le cas contraire le poids est modifié proportionnellement à la différence entre la sortie obtenue et la sortie désirée :

$$w \leftarrow w + k |d - s|$$

s est la sortie obtenue, d la sortie désirée et k une constante positive. [24]

_ La règle de rétro-propagation :

Cette règle est une généralisation de la règle de Widrow-Hoff pour des fonctions d'activation non linéaire et pour des réseaux multicouches. Elle s'utilise pour ajuster les poids de la couche d'entrée à la couche cachée.

Les poids dans le réseau de neurones sont au préalable initialisés avec des valeurs aléatoires.

Un ensemble de données est considéré afin de servir à l'apprentissage. Chaque échantillon possède ses valeurs cibles qui sont celles que le réseau de neurones doit à terme prédire lorsque le même échantillon est présenté.

Après avoir vu le fonctionnement général d'un réseau de neurone, dans ce qui suit seront définis les principaux modèles de réseaux de neurones : le perceptron, le perceptron multicouches, les réseaux de Hopfield, les réseaux à fonction radiale de base et les réseaux ART. [14]

II.5) Quelques réseaux de neurones célèbres :

A) *Le perceptron :*

C'est un des premiers réseaux de neurones, conçu en 1958 par Rosenblatt. Il est linéaire et monocouche. Il est inspiré du système visuel. Les sorties des neurones ne peuvent prendre que deux états (-1 et 1 ou 0 et 1).

La règle de modification des poids utilisée est la règle de Widrow-Hoff.

Une des limites du perceptron classique, son inaptitude de simuler la fonction ou exclusif (xor). [52]

B) *Le perceptron multicouches (PMC) :*

Le perceptron multicouches est une amélioration du perceptron comprenant une ou plusieurs couches intermédiaires dites couches cachées. Ces couches ont une utilité fondamentale pour le réseau de neurones mais n'ont pas de contact direct avec l'extérieur. Chaque neurone n'est relié qu'aux neurones des couches directement précédente et suivante, mais à tous les neurones de ces couches.

Les PMC utilisent, pour modifier leurs poids, un algorithme de rétropropagation du gradient.

La modification des poids est propagée de la couche de sortie jusqu'à la couche d'entrée et se fait selon la relation suivante :

$$W(t+1) = W(t) - \lambda(t) \text{grad}_w(E_p(t))$$

Où $E_p = (Y_p - D_p)^2$ est l'erreur calculée sur un seul exemple tiré au hasard à l'instant t ; $\lambda(t)$ est le pas de gradient (fixe, décroissant, ou adaptif) Y_j la sortie calculée et D_j la sortie désirée).

Le PMC agit comme un séparateur non linéaire et peut être utilisé pour la classification, le traitement d'image ou l'aide à la décision. [33]

On ne peut parler du perceptron multi couches sans citer le théorème d'approximation universelle.

Le théorème d'approximation universelle :

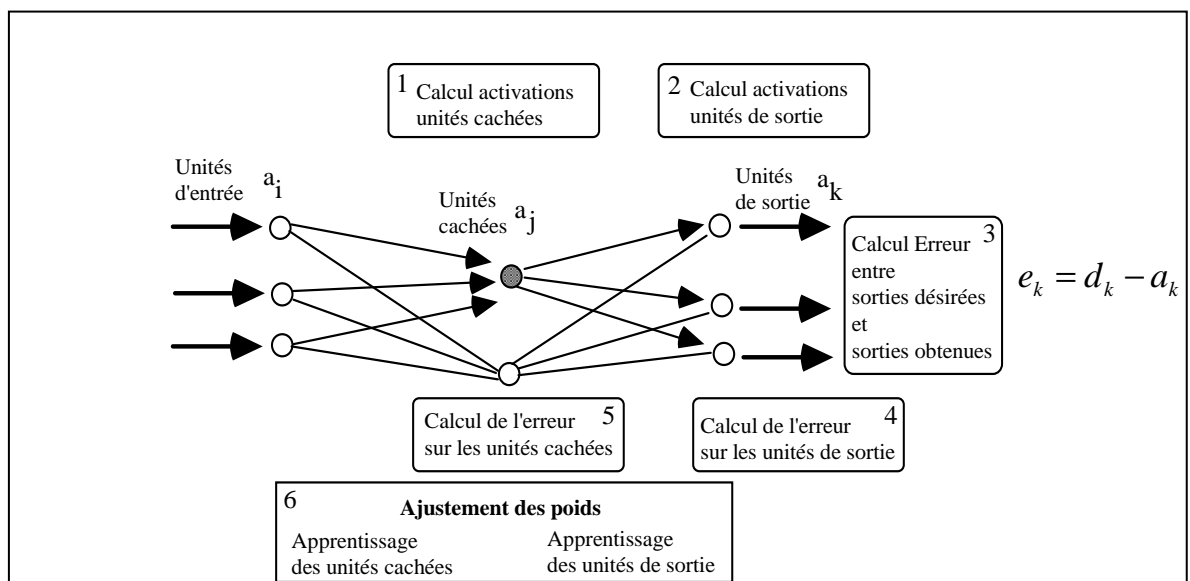
– Cybenko (1989) :

Pour toute fonction F continue et définie sur un ensemble borné et pour tout ε , il existe un réseau à une couche cachée de neurones sigmoïdes qui approxime F à ε près.

– Sussman (1992) :

Les réseaux à une couche cachée forment une famille d'approximateurs parcimonieux à nombre égal de paramètres qui approxime correctement plus de fonctions que le feraient des polynômes. [15]

Figure I.10 : fonctionnement d'un PMC.



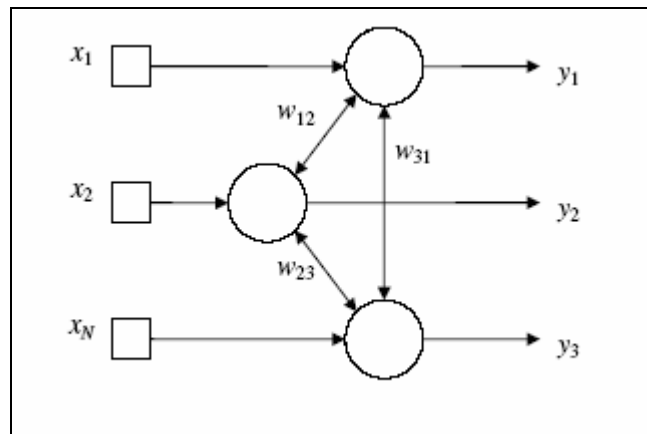
C) Les réseaux de Hopfield :

Le réseau de Hopfield est un réseau avec sorties binaires. Tous les neurones du réseau sont interconnectés avec des poids symétriques (le poids de connexion du neurone N_i au neurone N_j est égal au poids du neurone N_j au neurone N_i). Les poids sont choisis par l'utilisateur. Les états et les poids des neurones permettent de définir l'"énergie" du réseau. C'est cette énergie que le réseau tente de minimiser pour trouver une solution. L'apprentissage de ces réseaux se fait selon la loi de Hebb.

Ce réseau peut apprendre à mémoriser les exemples présentés en entrée sous la forme d'états stables. En phase d'exploitation, les stimuli présentés en entrée évolueront dans le réseau vers l'état stable le plus ressemblant.

Une application typique de ce type de réseau est le problème du voyageur de commerce, où les poids représentent d'une part les distances entre les villes et d'autre part les contraintes qui assurent la visite de chaque ville une seule fois. [37]

Figure I.11 : réseau de Hopfield



D) Les réseaux de fonctions à base radiale (RBF) :

Un réseau RBF est composé d'une couche d'entrée, d'une seule couche cachée et d'une couche de sortie. Les unités en entrée d'un réseau RBF sont connectées avec toutes les unités de l'unique couche cachée.

Les réseaux RBF utilisent une approche de classification par noyaux. Un noyau est une fonction définie dans l'espace des entrées, qui délimite une région restreinte appelée zone d'activité. L'ensemble de l'espace d'entrée est recouvert par les différentes zones d'activités des noyaux, qui se recouvrent. Le problème à résoudre est divisé en un ensemble de sous problèmes, chacun d'eux étant résolu dans l'un des noyaux.

Afin de permettre cette approche de classification par noyaux, les fonctions d'activation (appelées aussi fonctions de base) sont radiales. En effet, leur sorties ne dépendent que de la distance entre l'entrée et un point particulier de l'espace d'entrée : le centre de la fonction. [19]

Les fonctions de base produisent une réponse différente de zéro seulement lorsque l'entrée se situe dans une petite région bien localisée de l'espace des variables. Bien que plusieurs modèles de fonctions de base existent, le plus courant est de type Gaussien:

$$y_{1,i} = \exp \left[-\frac{(x - w_{1,i})^T (x - w_{1,i})}{2\sigma_i^2} \right]$$

X est le vecteur d'entrée du réseau, $y_{1,i}$ est la sortie du neurone i de la première couche, W et σ_i^2 sont respectivement le vecteur de poids synaptiques et le paramètre de normalisation de ce neurone (W correspond ici aux coordonnées du centre de la Gaussienne).

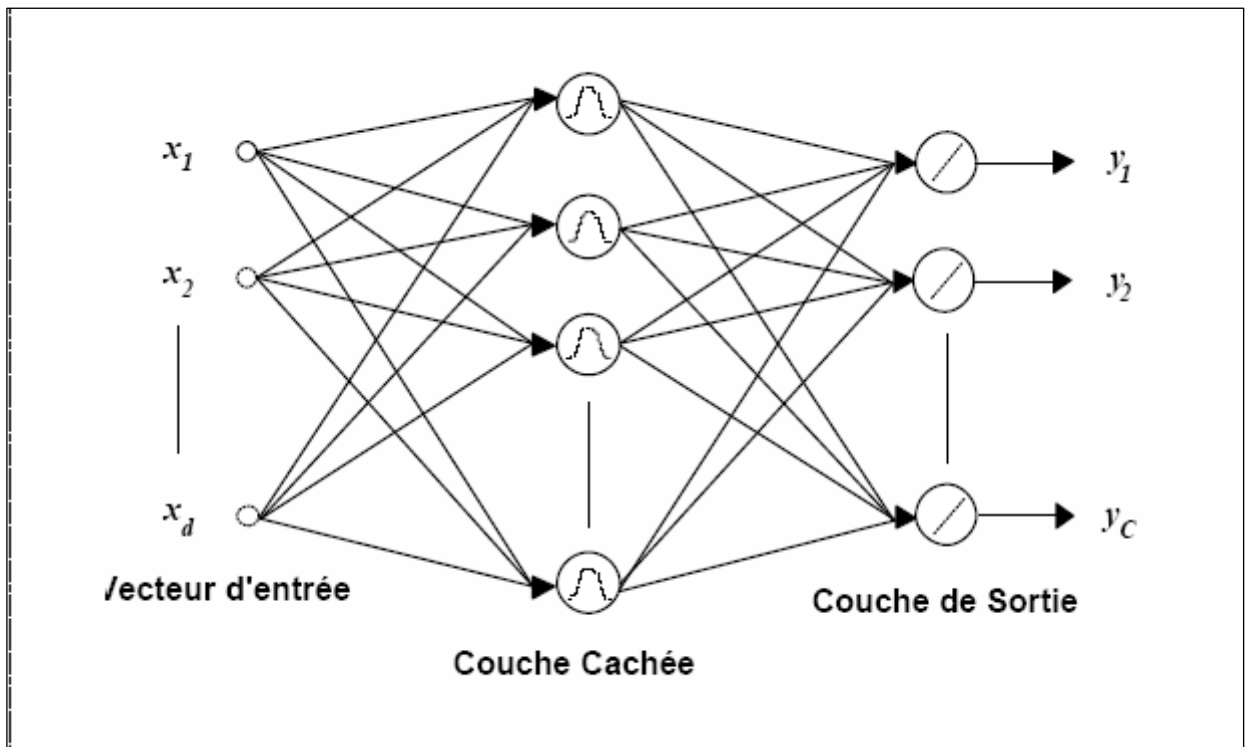
La sortie d'un neurone de la seconde couche est simplement donnée par :

$$y_{2,i} = W_{2,j}^T Y_1$$

Où Y_1 est le vecteur des sorties des neurones de la première couche.

Plus le vecteur d'entrée est proche du centre d'une Gaussienne, plus la sortie du neurone de la première couche qui lui correspond est élevée. [37]

Figure I.11 : représentation générale d'un réseau RBF.



E) Les réseaux ART (Adaptive Resonance Theory) :

Les réseaux de neurones ART appartiennent aux réseaux de neurones à apprentissage non supervisé de type compétitif.

Ces réseaux comprennent deux couches de neurones F1 et F2 :

- F1 qui détecte les caractéristiques des modèles étudiés.
- F2 qui classe ces modèles en catégories.

L'apprentissage de ces réseaux fait appel simultanément aux deux aspects de la mémoire :

1. la mémoire à long terme (LTM) qui gère les interactions entre les couches F1 et F2 et s'enrichit pendant la phase d'apprentissage.
2. la mémoire à court terme (STM) qui décrit l'état interne des couches F1 et F2 et se modifie à chaque cycle du système.

Contrairement aux autres types de réseaux de neurones, les réseaux ART réalisent simultanément la phase d'apprentissage et la phase d'exécution.

Le réseau en phase de test, s'adapte à des entrées inconnues en construisant de nouvelles classes (ajout de neurones) tout en dégradant au minimum les informations déjà mémorisées.

Les poids des interconnexions codent les prototypes des classes. Les valeurs de ces poids ne sont adaptées que s'il y a une approximation suffisante entre l'entrée du réseau et le prototype déjà connu par le réseau. Alors il existe une résonance. Sinon, l'entrée du système est très éloignée du prototype, on fonde alors une nouvelle classe. [6]

Après avoir vu les concepts généraux liés aux méthodes de classification et aux réseaux de neurones, le chapitre suivant traitera des cartes auto organisatrices de Kohonen qui sont un réseau de neurones utilisés généralement pour faire des classifications.

CHAPITRE II
LES CARTES AUTO ORGANISATRICES
DE KOHONEN

Introduction

L'auto-organisation constitue l'un des thèmes majeurs du paradigme connexionniste. L'algorithme de Kohonen est un des algorithmes d'apprentissage les plus célèbres. On l'appelle aussi algorithme d'auto-organisation, ou carte de Kohonen, ou plus brièvement SOM (Self-Organizing Map). Il représente le prototype de l'apprentissage non supervisé dans le domaine connexionniste.

Les cartes de Kohonen représentent un procédé d'auto-organisation qui cherche à projeter des données de grande dimension dans un espace discret et régulier de dimension moindre (en général 1 ou 2D). En fin d'apprentissage, la carte reproduit les corrélations présentes dans les données d'entrée. Grâce au procédé d'auto-organisation, non seulement la topologie qui lie les données initiales est conservée au niveau des réponses proposées par la carte, mais aussi l'algorithme fait correspondre à des entrées voisines des unités identiques ou voisines.

La présentation de ces cartes ainsi que les fondements sur lesquels repose l'algorithme qui les régit font l'objet de ce chapitre.

Dans un premier temps, une présentation de la notion d'auto organisation et une revue de quelques éléments historiques qui sont à la base de l'algorithme d'auto organisation seront faites.

Par la suite, l'architecture, le fonctionnement et les propriétés des cartes de Kohonen seront explicités ainsi que ses avantages, ses principales limites et ses extensions les plus connues.

Enfin, une comparaison de la classification de Kohonen par rapport aux autres méthodes sera réalisée.

1) *Quelques éléments historiques :*

Même si la notion d'auto-organisation a vu le jour bien avant grâce aux recherches des biologistes, elle ne s'est étendue qu'au milieu du 20^{ème} siècle, à la croisée des mathématiques et de l'informatique.

En effet, des recherches en biologie ont montré que des mécanismes d'auto-organisation semblaient être impliqués dans l'homéostasie, l'établissement de la polarisation cellulaire, le système nerveux et dans la différenciation cellulaire.

En 1973, le chercheur Van Der Malsburg, inspiré par l'organisation corticale des vertébrés, développa dans son étude sur le cortex visuel, un modèle d'apprentissage par compétition. Il mit en évidence que le modèle du cortex visuel ne se base aucunement sur des prédéterminations génétiques, mais sur des phénomènes auto-organisés¹. L'expérience a été réalisée sur des cellules corticales visuelles de chatons de cinq semaines.

Cette étude a montré que:

1. Le traitement des perceptions visuelles extérieures se décompose en un codage du signal par une série de neurones récepteurs suivi d'un traitement de l'information à l'intérieur du système nerveux.

2. L'organisation du cortex visuel s'effectue de sorte que des régions proches s'activent pour des stimuli visuels proches. De la même manière, pour le signal auditif, des fréquences proches activent des régions proches du cortex auditif.

3. Cette organisation est non supervisée, c'est la nature du signal extérieur qui structure le cortex. Il s'agit d'extraire des caractéristiques à partir des signaux reçus.

D'autres applications de l'auto-organisation suivirent, telles que les mouvements de foule, l'évolution des systèmes de villes, l'adaptation des écosystèmes, les structures cognitives, où l'auto-organisation rejoint la notion d'apprentissage.

Toutes ces applications soulignèrent l'importance des rétroactions, des non-linéarités, et du caractère ouvert et hors d'équilibre des systèmes pour qu'apparaissent des formes stables et reproductibles sans prescription extérieure.

L'auto-organisation a ainsi apporté l'idée que des formes stables et statistiquement reproductibles peuvent découler d'un équilibre dynamique, mettant en jeu des règles locales et stochastiques.

En 1976, Stephan Grossberg énonce le principe général du modèle de résonance adaptative (ART). Le modèle ART offre une solution originale à l'apprentissage autonome et non

¹ Phénomène d'adaptation du cerveau aux signaux qu'il reçoit

supervisé de catégories. La seule grande lacune du modèle ART est son incapacité à conserver la topologie.

Ce n'est qu'en 1982 que le chercheur finlandais Tuevo Kohonen proposa un algorithme modélisant le phénomène biologique appelé rétinopathie et reposant sur le concept de l'auto-organisation : il s'agit de l'établissement de relations structurées entre la rétine et le cortex, qui préservent la topologie de la rétine [7].

Cet algorithme produit une carte d'organisation topologique désignée communément par « carte SOM » (Self Organising Map).

Il publia en 1984, la première version de son ouvrage : *Self Organization and associative memory* qui est régulièrement remis à jour. En 1995, il a édité un nouvel ouvrage *Self Organizing Maps* présentant sa méthode et ses dernières évolutions, le tout complété par plus d'un millier de références bibliographiques.

Dès 1988 plusieurs études ont commencé à voir le jour pour comprendre le fonctionnement de ce modèle. De nombreuses améliorations relatives à l'architecture du réseau, au réglage des paramètres, à l'étude théorique du fonctionnement et à la représentation des sorties ont été apportées dans les années 90.

Le modèle d'auto-organisation de Kohonen se base sur le principe que plus un st_i^2 multi est fréquent, plus sa représentation sera développée sur la carte.

Les propriétés de l'algorithme SOM permettent conjointement d'analyser des données qui ont une structure non linéaire et de réaliser des représentations graphiques qui mettent en évidence certaines de leurs caractéristiques.

Ce processus ne dépend que des entrées et ne nécessite pas l'intervention d'un superviseur. L'algorithme fait correspondre les éléments de l'espace d'entrée avec des unités ordonnées sur la carte de sortie.

L'application la plus courante de l'algorithme SOM est la classification de l'espace d'entrée, où l'on définit une notion de voisinage entre les classes qui n'est pas prise en compte par les méthodes de classification classiques.

Les cartes SOM possèdent des applications variées, allant de la biologie à la linguistique, en passant par la physique et les sciences sociales [22].

² Une observation de l'espace d'entrée.

2) Fondements de base des Cartes de Kohonen:

L'approche par cartes topologiques de Kohonen repose sur le calcul de distances entre les vecteurs d'entrée et des vecteurs de référence selon un *apprentissage dit par compétition*. L'apprentissage par compétition est une méthode de détection des similarités.

L'architecture des réseaux à apprentissage par compétition possède une couche d'entrée et une couche de compétition. Une forme est présentée à l'entrée du réseau et est projetée sur chacun des neurones de la couche compétitive. Chaque neurone de sortie est connecté aux neurones de la couche d'entrée et aux autres cellules de sortie (connexion inhibitrice) ou à elle même (excitatrice). La sortie dépend alors de la compétition entre les connexions inhibitrices et excitatrices. A chaque neurone, un vecteur poids, de la même dimension que les vecteurs d'entrée, est attribué. [3]

Le principe de la méthode consiste à modifier le vecteur poids d'une cellule seulement si celle-ci est activée.

Il existe deux possibilités pour déterminer le nombre de cellules actives :

- Une seule cellule est active à chaque présentation. Cette règle est utilisée dans la plupart des réseaux en mode non supervisé ;
- Plusieurs cellules sont actives en même temps, on parle alors de représentation distribuée. L'analyse de cette règle est très complexe et elle a été peu étudiée.

Dans la suite du mémoire, la première règle sera retenue où, pour chaque entrée x , une seule cellule est activée. Cette cellule est sélectionnée comme étant celle dont le vecteur poids w_b est le plus proche de l'entrée x :

$$d^V(x, w_b) = \min d^V(x, w_i) \text{ avec } i = 1, \dots, N$$

Où $d^V(x, w_i)$ est une mesure de la distance entre les vecteurs x et w_i .

La modification des vecteurs prototypes se fait selon la règle d'apprentissage dite du *winner take all* (le gagnant prend tout) et dont la forme la plus simple est : [43]

$$\Delta w_i = \begin{cases} 0 & \text{si } i \neq b \\ \alpha(t)(x - w_b) & \text{sin on} \end{cases} \quad (1)$$

Où $\Delta w_i = w_i(t+1) - w_i(t)$ et où $\alpha(t)$ représente une fonction permettant la convergence en probabilité des w_i vers le centre de gravité de chaque classe. Une présentation de cette fonction ainsi que ses propriétés sera faites dans la section suivante.

3) *Fonctionnement des Cartes de Kohonen :*

D'une manière générale, un réseau de neurones peut être décrit par 3 caractéristiques : une architecture, une règle de transmission et une règle d'apprentissage.

a. Architecture des cartes de Kohonen :

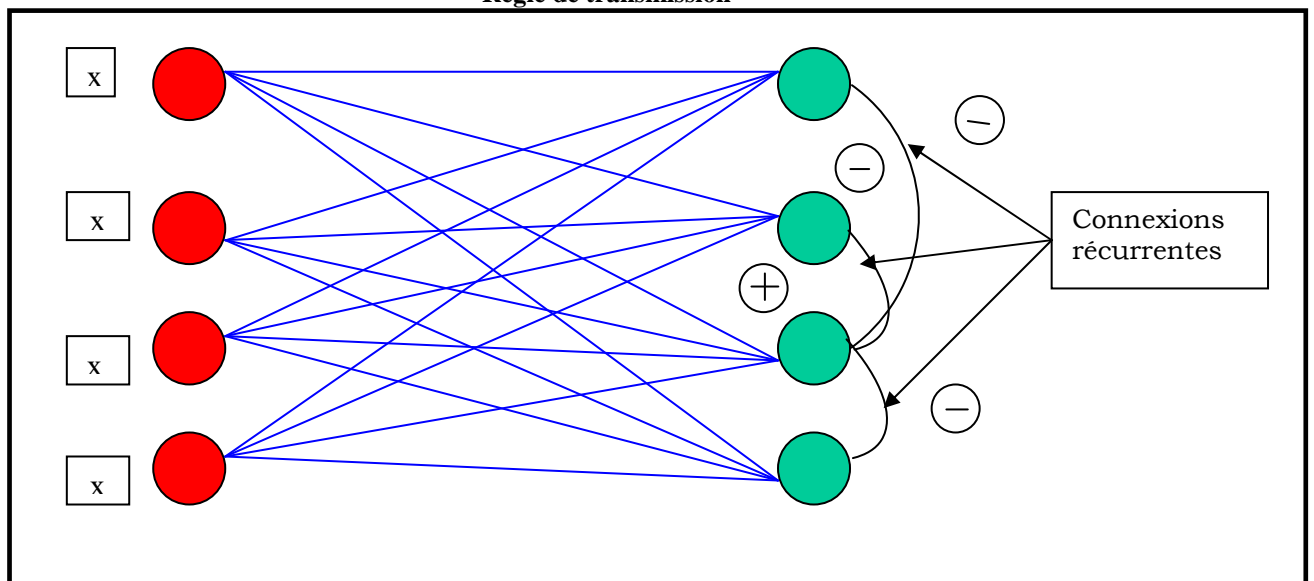
La carte de Kohonen est un réseau de neurones à deux couches dont l'architecture peut se définir de la manière suivante :

- Une couche d'entrée, servant à la présentation des données, et prenant la valeur du vecteur d'observation ;
- Une couche d'adaptation, ou couche de sortie, comprenant N neurones de sortie. (Figure 3.1).

Les p neurones d'entrée relient chacun des N neurones de sortie. Le réseau comporte donc $p \times N$ connexions au total.

Chaque neurone de sortie possède des connexions latérales récurrentes, qui viendront modifier la relation entre chacun des neurones d'entrée et de sortie. [21]

Figure II.1 : Architecture du réseau de Kohonen montrant la relation entre les vecteurs d'entrée x , les poids de connexions w_i , les connexions récurrentes d'excitation (+) et d'inhibition (-) et les sorties y .
Règle de transmission



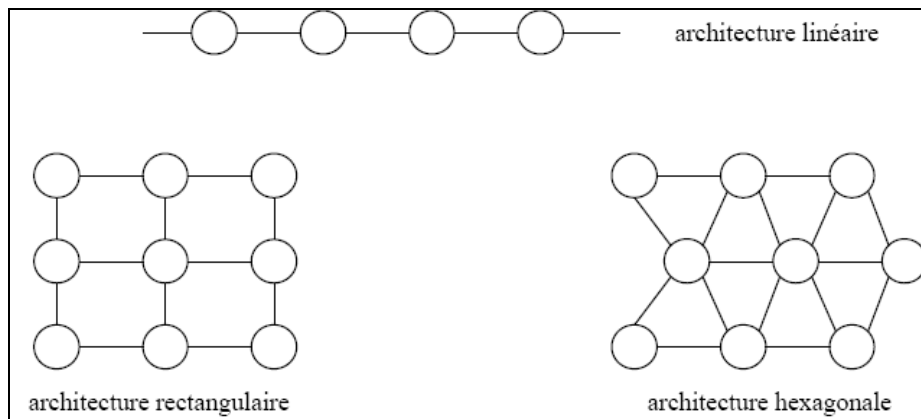
b. Notions de voisinage :

Comme dit précédemment, la carte SOM repose sur le principe d'un apprentissage compétitif. Cependant, Kohonen a proposé de compléter cette méthode d'apprentissage par l'adjonction de relations de voisinage entre les cellules.

Se basant sur les modèles biologiques, dans lesquels les cellules nerveuses ne sont pas des entités indépendantes et isolées mais un ensemble fortement interconnecté, il définit un

ensemble de liens entre les cellules. Cet ensemble ainsi que le nombre de cellules employées constituent une architecture. La méthodologie employée pour relier les cellules entre elles va permettre d'organiser la disposition des vecteurs prototypes tout en préservant les relations de voisinage. Les cellules sont disposées physiquement aux noeuds d'un réseau dont les schémas de connexion les plus classiques: linéaire, rectangulaire, hexagonal, sont présentés dans la Figure 3.2. Le maillage défini par les relations inter-cellules, appelées relations de voisinage, est appelé carte. [21]

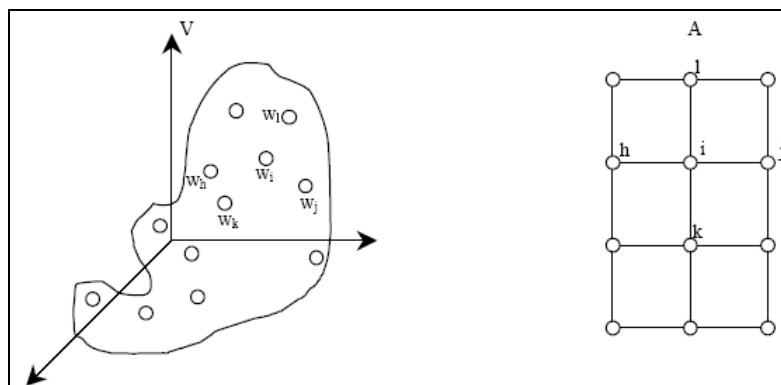
Figure II.2 : exemple de formes possibles des cartes de Kohonen.



Une cellule i est représentée par ses coordonnées $i = (i_1, i_2)$ (ici dans le cas d'un voisinage bidimensionnel) dans un espace de sortie discret A , que l'on peut représenter comme étant une grille, chaque intersection de celle-ci représentant une position possible pour une cellule.

Chaque cellule est représentée par un vecteur prototype dans la base de données et par des coordonnées dans A (Figure II.3).

Figure II.3 : Représentation des vecteurs poids dans l'espace d'entrée V (à gauche), et sur la grille A



Afin de mieux comprendre les relations de voisinages liant les neurones de la carte, il est indispensable de présenter quelques définitions importantes:

- **Distance dans A entre deux cellules i et j :** La carte est un système de coordonnées mono ou bi-dimensionnelles dans la majeure partie des cas, sur lequel une distance peut être définie. La plus commune est la distance euclidienne. Il est également possible de définir la distance dans A comme étant le nombre minimal de liens reliant les cellules i et j .
- **Voisine de rang k dans A d'une cellule :** La cellule i possède comme voisine de rang k la cellule j si le nombre minimal de lien reliant ces deux cellules est égal à k . Il peut exister plusieurs cellules voisines de même rang.
- **Voisinage de rang k d'une cellule i :** Comme le montre la Figure 3.2 une cellule possède en général plusieurs voisines de même rang. Son voisinage de rang k sera constitué par l'ensemble des cellules voisines de rang inférieur ou égal à k .

L'adjonction des notions voisins et voisinage introduit une modification dans l'algorithme d'apprentissage par compétition. Au lieu de laisser la cellule gagnante être la seule à modifier son vecteur prototype, les relations de voisinage permettent aux diverses voisines de cette cellule de bénéficier de son apprentissage en modifiant leurs vecteurs prototypes. La règle (1) va être modifiée en *winner take most* (le gagnant prend l'essentiel), l'idée étant d'amener une relaxation locale (effet de lissage) sur les vecteurs prototypes voisins de la cellule gagnante, de telle sorte qu'une organisation globale apparaisse. [21]

L'activation régit pendant l'apprentissage s'estompe au fur et à mesure que l'on s'éloigne du neurone gagnant. Cette fonction d'activation est appelée *fonction de voisinage*.

- *La fonction de voisinage :*

La fonction de voisinage ou fonction d'activation des SOM a la forme d'un chapeau mexicain et permet de distinguer 3 zones d'actions (figure II.4.1):

- Une région restreinte d'excitation (1) ;
- Une région d'action inhibitoire (2) ;
- Une région d'excitation plus faible (3), cette zone possède une action négligeable et est habituellement ignorée.

Cette fonction prend une valeur maximale pour le neurone gagnant et décroît au fur et à mesure que l'on s'éloigne de lui.

Figure II.4.1 : Fonction de voisinage (ou d'activation) « chapeau mexicain »

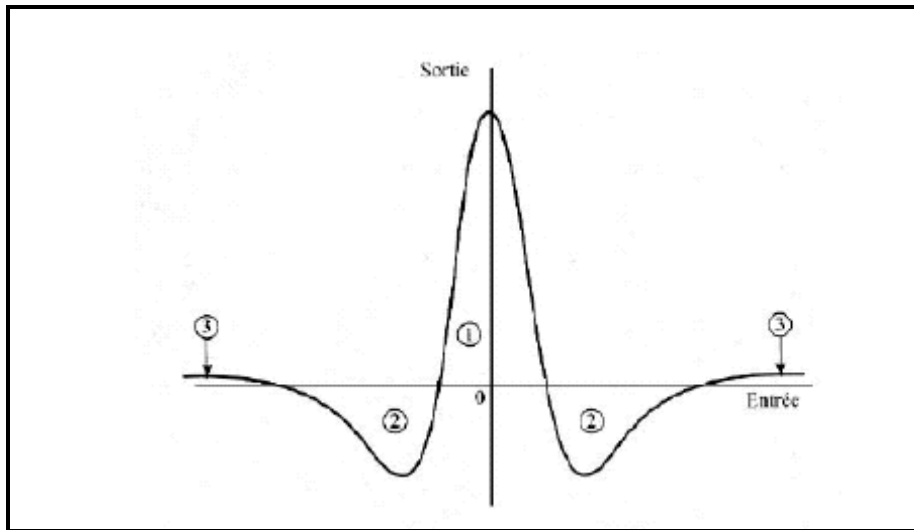
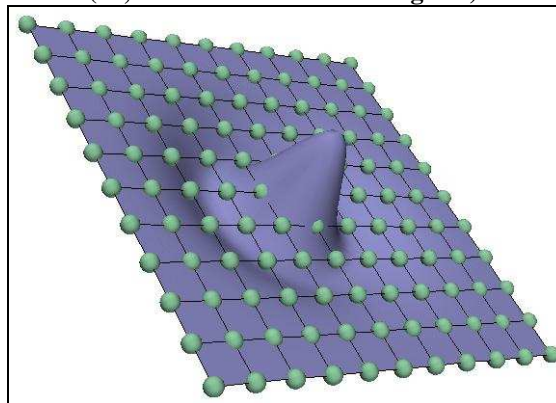


Figure II.4.2 : Activité en chapeau mexicain autour du neurone le plus excité (ici, le neurone central de la grille).



Il existe plusieurs fonctions de voisinage qui décroissent pendant le temps. Les fonctions de voisinage les plus importantes sont:

- Bubble : $h_{ck}(t) = f(\sigma_t - d_{ck})$;
- Gaussienne : $h_{ck}(t) = e^{-d_{ck}^2 / 2\sigma_t^2}$;
- Cut-gauss : $h_{ck}(t) = e^{-d_{ck}^2 / 2\sigma_t^2} f(\sigma_t - d_{ck})$
- Ep : $h_{ck}(t) = \max [0, 1 - (\sigma_t - d_{ck})/2]$.

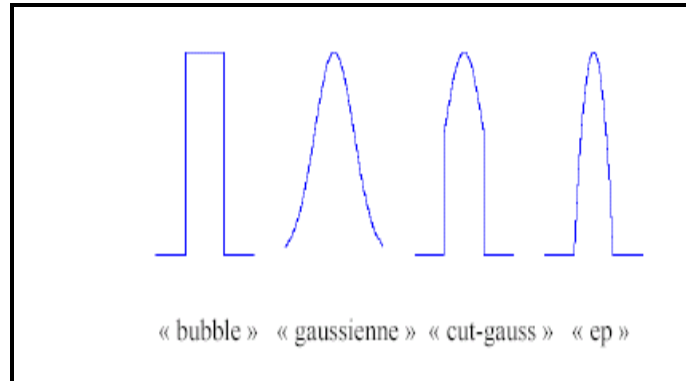
Où σ_t est le rayon de voisinage au temps t qui est défini par :

$$\sigma(t) = \sigma_{initial} \left(\frac{\sigma_{final}}{\sigma_{initial}} \right)^{\frac{t}{t_{max}}}$$

$\sigma_{initial}$ et σ_{final} sont choisies par l'utilisateur ;

$d_{ck} = \|r_c - r_k\|$ est la distance entre les unités c et k sur la grille et $f(.)$ est une fonction à seuil.

Figure II.5 : différents types de fonctions de voisinage.



La fonction de voisinage généralement adoptée est la fonction Gaussienne. Cette fonction est invariante par translation : elle ne dépend pas du neurone vainqueur et présente une rapidité de convergence qui la rend privilégiée. Elle est centrée sur le neurone déclaré vainqueur après la phase de compétition.

c. Règle d'apprentissage :

Le modèle des cartes de Kohonen simule le mécanisme d'apprentissage sous la forme d'un processus de classification topographique.

Ce mécanisme comprend trois phases pour chaque vecteur de l'espace d'entrée: une phase dite d'initialisation, suivie d'une phase de compétition entre les neurones de la carte et enfin, une phase d'adaptation des poids des neurones.

- **Initialisation** : cette étape s'effectue en choisissant aléatoirement les vecteurs poids initiaux.

- **Compétition** : X est l'espace d'entrée, chaque entrée est représentée par un vecteur $x \in \mathbb{R}^p$. L'espace d'entrée se présente de la façon suivante :

$$X = \begin{cases} x(1) = [x_{11}, x_{12}, \dots, x_{1p}] \\ x(2) = [x_{21}, x_{22}, \dots, x_{2p}] \\ \vdots \\ x(m) = [x_{m1}, x_{m2}, \dots, x_{mp}] \end{cases}$$

Soit W une grille de neurones de sortie de dimension N , chaque neurone i est associé à un vecteur poids synaptique de dimension p , $w(i) = [w_{i1}, w_{i2}, \dots, w_{ip}]$, où p correspond à la dimension du vecteur d'entrée.

Une fois qu'un vecteur d'entrée $x(1) = [x_{11}, x_{12}, \dots, x_{1p}]$ soit aléatoirement sélectionné dans la distribution de l'espace d'entrée, tous les neurones de la grille sont mis en *compétition*.

Le vecteur de l'ensemble d'apprentissage $x(I)$ est comparé à chacun des vecteurs poids. Le neurone vainqueur sera celui qui se rapprochera le plus du vecteur d'entrée courant, c'est-à-dire celui dont la distance entre son vecteur poids synaptique et le vecteur d'entrée est la plus courte. Le neurone vainqueur j est identifié à l'aide d'une mesure de similarité habituellement définie comme étant la distance euclidienne entre les deux vecteurs : [21]

$$j = \arg \min \|x - w_i\|, I = 1, 2, \dots, N$$

Où $\|\cdot\|$ est la norme euclidienne telle que définie par :

$$\|x\| = \sqrt{\sum_{j=1}^p x_j^2}.$$

• **Adaptation des poids** : cette dernière phase sert à accroître la ressemblance des neurones avec les vecteurs d'entrée.

En effet, plus un neurone est proche du neurone vainqueur, plus il sera influencé. A mesure que l'apprentissage progresse, de moins en moins de neurones seront influencés par le neurone vainqueur réduisant de la sorte la taille du voisinage de ce dernier.

L'adaptation des poids de chaque neurone est donc fonction de la position d'un neurone dans la grille W par rapport au neurone vainqueur.

La règle d'apprentissage pour un neurone k appartenant au voisinage topologique du neurone vainqueur j s'énonce comme suit :

$$W_k(t+1) = \begin{cases} W_k(t) + \alpha(t) h_{jk}(t) [x - W_k(t)] & \text{ssi } k \in \Lambda_{j(x)}(t) \\ W_k(t) & \text{ssi } k \in \mathcal{A}_{j(x)}(t) \end{cases}$$

Avec :

- t le temps ;
- $\alpha(t)$ le taux d'apprentissage ;
- $h_{jk}(t)$ la fonction de voisinage ;

○ $\Lambda_{j(x)}(n)$ le voisinage topologique autour du neurone vainqueur j au temps discret t .

Le taux d'apprentissage $\alpha(t)$ est une fonction décroissante, monotone et positive de t , telle que :

$$\alpha(t) = \frac{\alpha_0}{1 + 100t/T} \quad \text{et} \quad 0 \leq \alpha(t) \leq \alpha_0 < 1$$

Où α_0 représente la valeur initiale de $\alpha(t)$ et T est le temps total.

$\alpha(t)$ vérifie les conditions suivantes :

$$\lim_{t \rightarrow \infty} \alpha(t) = 0 \quad \text{et} \quad \lim_{T \rightarrow \infty} \sum_{i=1}^T \alpha(t) = \infty$$

La fonction de voisinage $h_{jk}(t)$ prend les valeurs suivantes :

$$\left\{ \begin{array}{ll} h_{ji}(t) = 1 & \text{si } i = j ; \\ 0 < h_{ji}(t) < 1 & \text{si } i \in \Lambda_{j(x)}(t) ; \\ h_{ji}(t) = 0 & \text{sinon.} \end{array} \right.$$

Si la fonction de voisinage est de type gaussien, elle sera décrite par :

$$h_{ji}(t) = e^{-\frac{\|r_j - r_i\|^2}{2\sigma^2(t)}}$$

Où r_j et r_i sont les positions des neurones j et i dans la carte SOM et $\sigma(t)$ est le rayon de voisinage.

Le rayon de voisinage doit être également une fonction décroissante du temps. Il est défini par :

$$\sigma(t) = \sigma(t)_{\text{initial}} \left(\frac{\sigma(t)_{\text{final}}}{\sigma(t)_{\text{initial}}} \right)^{\frac{t}{t_{\text{max}}}}$$

$\sigma(t)_{\text{initial}}$ et $\sigma(t)_{\text{final}}$ sont des paramètres fixés à l'avance.

Soit $\eta(t) = \alpha(t) h_{jk}(t)$, $\eta(t)$ est appelé le pas d'adaptation. [43]

Selon ce processus d'apprentissage, les données d'entrée sont donc ramenées à des classes qui s'organisent généralement selon une structure bidimensionnelle de neurones, pour laquelle les relations de voisinage sont définies.

Le processus de classification topographique du modèle SOM combine donc une étape de classification avec une étape de projection des données.

L'étape de projection des données fait rappeler l'ACP. En effet, chercher la plus petite distance entre x et les w_i en utilisant la distance euclidienne revient à maximiser la quantité $\sum x_j w_{ij}$, et à chercher la droite qui permet d'avoir un allongement maximum. La différence

ente les SOM et l'ACP réside dans le fait que cette étape de projection se fait d'une manière aléatoire à chaque itération dans l'algorithme de Kohonen alors que pour l'ACP, à chaque étape la projection se fait sur un axe perpendiculaire à celui de l'étape précédente.

Le processus d'apprentissage et les différentes étapes d'adaptation sont résumés dans l'algorithme d'apprentissage.

d. Algorithme d'apprentissage :

L'algorithme d'apprentissage de Kohonen permet d'accentuer la ressemblance entre une donnée et le vecteur de poids dont elle est la plus proche.

Les différents paramètres à fixer avant de lancer cet algorithme sont :

- La taille de la carte, en précisant le nombre de neurones et la structure de la carte ;
- Le voisinage, en indiquant le type et le seuil (rayon maximum de voisinage) ;
- Le temps, en spécifiant deux valeurs entre lesquelles cette fonction va décroître exponentiellement ;
- Et enfin, le nombre d'itérations. [26]

Un cycle de l'algorithme SOM peut être décrit de la manière suivante :

1. $t \leftarrow 0$, **initialiser** aléatoirement les vecteurs poids initiaux et le nombre maximum d'étapes d'apprentissage t_{\max} .

2. **Présenter** un vecteur x pris aléatoirement dans l'ensemble d'apprentissage.

3. **Similarité** : recherche du neurone vainqueur j en utilisant le critère de la distance euclidienne minimale (neurone plus proche) :

$$\|x - w_j\| = \min \|x - w_i\|, i=1, 2, \dots, N$$

4. **Modification des poids** : ajustement des poids des neurones appartenant au voisinage topologique $A_j(n)$ suivant le pas d'adaptation $\eta(n)$.

$A_j(n)$ et $\eta(n)$ varient dynamiquement durant l'apprentissage pour donner un meilleur résultat :

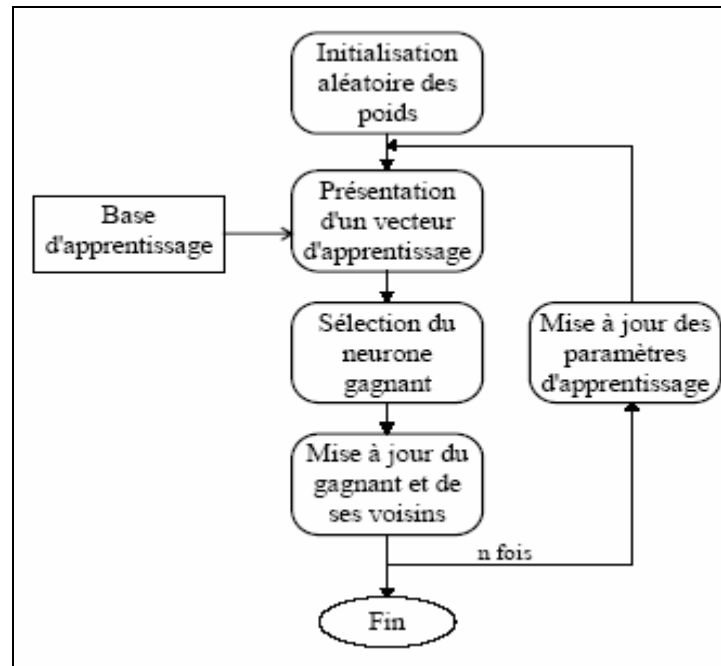
$$W_i(t+1) \leftarrow W_i(t) + \eta(t) \cdot [x - W_i(t)]$$

5. **Itération** : si $t < t_{\max}$ alors $t \leftarrow t+1$, retourner au point 2.

Sinon fin [31].

Le schéma suivant, récapitule les différentes étapes de l'algorithme de Kohonen.

Figure II.6 : Schéma de l'algorithme d'apprentissage de Kohonen.



En fin d'apprentissage, les poids des neurones vont converger vers des valeurs telles qu'un neurone ne sera actif que pour un sous ensemble précis de la base d'apprentissage, et chaque élément de la base n'activera qu'un seul neurone [31].

Cette valeur représente une observation *moyenne* résumant le sous-ensemble des observations qui lui sont affectées [26].

La marge d'erreur de l'algorithme exprime la distance moyenne de chaque observation de la base de données par rapport au vecteur prototype de sa classe d'appartenance. Cette erreur peut être calculée par la formule suivante :

$$\text{Erreur} = \frac{1}{N} \left(\sum_{i=1}^N \min_j \| x_i - w_j \| \right)$$

- *Convergence de l'algorithme :*

Malgré ses nombreuses applications et utilisations, la carte de Kohonen résiste à toute démonstration mathématique de sa convergence.

Des tentatives de démonstrations ont été données par T.Kohonen lui même en 1984, mais la démonstration complète ne s'est faite que plus tard (Cottrell et Fort 1987). Cette démonstration s'est limitée au cas d'une carte monodimensionnelle : l'espace d'entrée comporte une dimension unique et la carte est une carte linéaire dont les neurones ne sont disposés que le long d'une seule dimension (une ficelle).

Par la suite, la démonstration a été généralisée pour d'autres distributions et avec d'autres fonctions de voisinage (Erwin et al 1992, Bouton et Pagès 1993).

Ce qui rend difficile l'étude théorique de l'algorithme dans le cas général, est le fait qu'il n'existe pas de potentiel (ou énergie) associé lorsque les entrées sont tirées suivant une loi de probabilité P continue.

Dans le cas discret, Ritter et al (1992), introduisent une fonction de potentiel, de façon à ce que l'algorithme de Kohonen réalise une descente de gradient stochastique, du moins pour chaque valeur constante de la fonction de voisinage.

Cette fonction de potentiel est définie par:

$$V(x_1, x_2, \dots, x_N, C_1, C_2, \dots, C_N) = \sum_{i=1}^n \sum_{k \in V_r(i)} \sum_{x_j \in \Gamma_k} \|x_j - C_i\|^2$$

Ce potentiel généralise la distorsion quadratique ou la variance intra-classes qui comprend la somme des carrés des distances de chaque observation non seulement à son vecteur code mais aussi aux vecteurs codes des classes voisines.

L'étude mathématique de ce potentiel n'est pas simple non plus, car il n'est pas partout différentiable.

Pour des cartes multidimensionnelles, les résultats ne sont que partiels. La difficulté majeure réside en fait dans le manque d'outils mathématiques pour définir un ordonnancement correct dans des espaces de dimension supérieure à 1.

L'alternative du calcul de la convergence de l'algorithme est la visualisation des résultats de l'apprentissage. Lorsque les entrées semblables appartiendront aux mêmes classes et que des classes voisines regrouperont des éléments qui se ressemblent, l'algorithme de Kohonen aura convergé. [8]

Une fois que l'algorithme de Kohonen ait convergé, la carte de Kohonen démontre des propriétés statistiques importantes qui confirment les liens forts entre l'espace d'entrée et les sorties de la carte. Ces propriétés feront l'objet de la section suivante.

4) *Propriétés de la carte SOM :*

Les propriétés importantes de la carte de Kohonen sont : l'approximation de l'espace d'entrée, la reproduction des variations statistiques de l'espace d'entrée et enfin la conservation de l'ordre topologique.

• *Approximation de l'espace d'entrée :*

Afin de bien expliquer cette première propriété de l'algorithme, le cas où les données d'entrées sont uniformément distribuées est considéré.

L'application de l'algorithme à cette base de donnée, génère une carte dont les unités sont aussi uniformément distribuées.

Le résultat d'apprentissage est donc un ensemble *représentatif* de l'ensemble d'apprentissage (les entrées).

A la convergence, les neurones représentent des zones assez équiprobables, régulièrement réparties et de même surface (en moyenne). Chaque neurone a la même probabilité d'être sélectionné. Ces zones sont des ensembles mutuellement exclusifs dont l'union comprend tous les points de l'espace d'entrée.(voir figure 3.7)

• *Reproduction des variations statistiques de l'espace d'entrée:*

La densité de la répartition des neurones dans l'espace de sortie reflète les variations statistiques de l'espace d'entrée : une région dense de données attire davantage de neurones qu'une région moins dense. Autrement dit, les zones de représentation sont grandes dans les régions à forte densité de probabilité et inversement. [9]

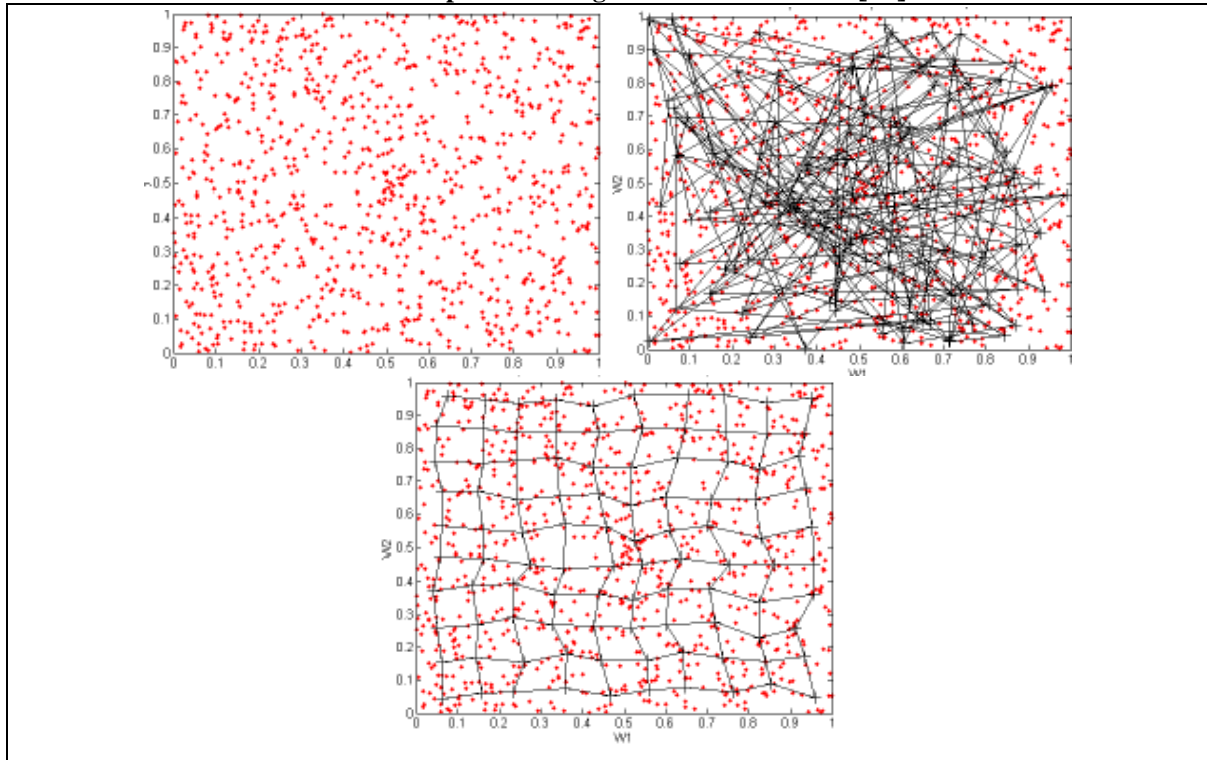
• *Conservation de l'ordre topologique :*

La propriété de l'ordre topologique est une conséquence directe de la mise à jour des vecteurs poids qui force le neurone gagnant et tous les neurones appartenant à son voisinage topologique de se déplacer vers le vecteur d'entrée associé.

En effet, les vecteurs semblables de l'ensemble de départ sont cartographiés sur des neurones se trouvant près les uns des autres.

Cette propriété reflète le comportement du système neuronal humain : de même que dans le cerveau, des stimulus similaires excitent des endroits voisins, deux vecteurs d'entrées proches activeront deux neurones voisins. C'est pourquoi on dit que la carte de Kohonen conserve la proximité topologique. [11]

Figure II.7 : De gauche à droite : la première figure représente l'ensemble des entrées uniformément distribuées; la 2^{ème} figure montre les entrées et les vecteurs poids au début de l'apprentissage qu'on relie en respectant l'ordre de leur indice ; et enfin dans la dernière figure à la fin de l'apprentissage, les vecteurs poids sont organisés uniformément. [11]



Connaissant le fonctionnement d'un réseau de Kohonen ainsi que ses propriétés, les avantages et les limites de l'algorithme de Kohonen peuvent maintenant être étudiés.

5) Principaux avantages et limites des cartes de Kohonen :

Grâce à ses nombreux avantages, la carte SOM est utilisée dans différents domaines.

Les principaux avantages d'utilisation des cartes de Kohonen sont :

- La projection du résultat de la classification dans un espace de dimension réduite ;
- L'économie dans le temps de calculs et ce grâce au tirage on line de l'algorithme : une seule observation est choisie par itération ;
- La visualisation directe des relations entre les données analysées [51];
- La facilité d'application ne nécessite pas une compréhension approfondie;
- La représentation graphique simple des prototypes et des classes ;
- L'apprentissage rapide car le rayon de voisinage commence par une grande valeur ensuite décroît dans le temps assurant ainsi un meilleur ajustement des poids ;
- La projection d'un point sur une carte déjà tracée est assez rapide [39] ;
- Les groupes sont déterminés automatiquement par le réseau.

Néanmoins, cette approche n'est pas exempte de défauts. Ses principales limitations peuvent être regroupées en 4 points :

1. la qualité de modélisation dépend étroitement de la dimension de la carte de sortie et donc de la qualité de préservation de l'ordre topologique qui est fixée à priori sans idée de la topologie des données.
2. la topologie de la carte est stationnaire, elle n'évolue pas dans le temps.
3. les règles d'apprentissage sont toutes fixées à l'avance, elles ne dépendent aucunement de la particularité de chaque processus de convergence, qui réside dans la séquence complète de déplacement des poids w_i . [51]
4. la carte de sortie est non orientée contrairement à la physiologie d'un cortex. [23].

6) *Les extensions de la Cartes de Kohonen :*

Afin d'améliorer les performances des cartes SOM, plusieurs arrangements ont été introduits à l'algorithme donnant naissance à plusieurs autres méthodes dites extensions de la carte.

Ainsi, ces extensions s'éloignent parfois grandement soit du modèle biologique, soit de l'algorithme en tant que tel.

Dans ce qui suit, seront présentés brièvement les développements récents dans la littérature.

- *Cartes hyperboliques :*

La topologie rectangulaire est celle utilisée généralement dans les cartes de Kohonen. Cependant, elle peut ne pas être la meilleure représentation pour un espace d'entrée complexe, de haute dimension.

À cette question, Ritter et al apportent leur élément de réponse avec un algorithme concret HSOM. La carte de neurones employée consiste en une projection d'un espace hyperbolique sur R^2 . Ritter conclut à une meilleure représentation pour un espace d'entrée à haute dimension, mais surtout à la possibilité de former une structure hiérarchique dans la carte de neurones. Un des inconvénients de cette approche est que la topologie de la carte est toujours spécifiée à priori. [16]

- *Tree Structured Self-Organizing Map, le TS-SOM:*

Cet algorithme s'appuie sur une géométrie d'arbre en croissance. Une spécification progressive de l'échelle d'apprentissage de l'algorithme se fait au fur et à mesure que l'on

ajoute des neurones. La structure d'arbre produit des catégories et des sous catégories, la modélisation en sortie est donc hiérarchisée : on peut lire la carte à différents niveaux d'arborescence. Cette approche ouvre la voie des algorithmes constructifs, qui sont tels que le nombre de neurones augmente au fil de l'apprentissage. [32]

- ***La famille d'algorithmes GSOM ou le plus connu d'entre eux, le GNG (Growing Neural Gas) :***

Ces algorithmes ont montré des performances extrêmement bonnes.

Dans le GNG, les neurones sont ajoutés au réseau avec une fréquence régulière. Pour savoir où ajouter un nouveau neurone, on se base sur une fonction d'erreur attribuée à chaque neurone. Plus celle-ci est grande, plus la région concernée a besoin de renfort neuronal. Avec cet algorithme, le problème de fixer une topologie par avance est résolu et l'erreur de modélisation est très faible. Les neurones intermédiaires disparaissent de la carte finale, la topologie est apprise automatiquement. Le problème de cette approche est évidemment qu'elle fait abstraction totale de la réalité biologique. [36]

- ***Receptive-Field Laterally Interconnected Synergetically Self-Organizing Map RFLISSOM:***

À l'inverse du GNG, cet algorithme reproduit assez fidèlement les observations expérimentales dans le cortex visuel primaire.

Dans ce modèle, la dynamique des neurones est régie par un système d'équations différentielles, les neurones étant de type « sommation et activation ». L'apprentissage est non supervisé de type Hebbien, et la connectivité du réseau est dynamique car on enlève au fur et à mesure les connexions inutiles. Tous les ingrédients sont donc réunis pour reproduire le plus fidèlement le comportement biologique des neurones avec les connaissances et techniques d'aujourd'hui.

Ce modèle contient tout de même des défauts, notamment des contraintes énormes sur la topologie initiale (connexions excitatrices et inhibitrices distribuées avec une parfaite régularité spatiale, graphe complet). [47]

- ***L'utilisation des algorithmes génétiques sur les SOM :***

Ceux-ci sont des algorithmes de recherche d'optimum global. On dispose d'une population de *chromosomes*, chacun d'entre eux étant un codage binaire d'une combinaison particulière de paramètres, correspondant ou « codant » pour une solution

particulière du problème. La validité de la solution est mesurée par une fonction de coût ou d'énergie. Le problème étant la recherche de l'optimum global, il faut trouver le chromosome qui code cette solution optimale. Avec des mutations, des recombinaisons et un critère de sélection naturelle, la population génétique évolue en favorisant les meilleurs chromosomes mais sans rester prisonnière des optima locaux.

Bien entendu, cette technique est lourde et complexe (codage, choix de la fonction de coût et temps de convergence sont des obstacles de taille) et ne permet pas de comprendre pourquoi un processus est meilleur qu'un autre, elle se contente de donner le meilleur. [44]

- **Auto SOM (Haese et Googhill, 2001):**

Dans cette méthode, la carte SOM est enrichie d'un algorithme qui estime automatiquement le coefficient d'apprentissage. L'algorithme consiste en un filtre de Kalman³ couplé avec une méthode d'estimation récursive de paramètres.

Le filtre calcule l'adaptation des poids du réseau en estimant le coefficient d'apprentissage qui minimise la variance de l'erreur d'estimation. Alors que la méthode d'estimation, calcule la largeur de la fonction de voisinage qui minimise la variance de l'erreur de prédiction du filtre de Kalman.

A cela s'ajoute une « topographic function » qui permet de préserver le voisinage et prévient ainsi la non convergence de la carte.

Comme cette carte vise à préserver les propriétés topologiques de l'espace d'entrée, la qualité de la topographie est mesurée pendant la phase d'apprentissage, soit en utilisant le produit topographique, soit en employant une fonction topologique. [17]

- **Batch map :**

Dans cette version de SOM, le taux d'apprentissage n'est pas utilisé. Contrairement à l'algorithme standard de Kohonen où on présente un vecteur de données à la fois, l'ensemble de la base de données est présenté pour adaptation au cours de la phase d'apprentissage dans la version Batch map.

A chaque étape d'apprentissage, la base de données est partitionnée selon les régions de Voronoi de la carte des vecteurs poids et de nouveaux vecteurs poids sont calculés. Le nouveau vecteur poids est la moyenne des poids des exemples sachant que le poids de chaque exemple est la valeur de la fonction de voisinage au point du vecteur gagnant. [38]

³ Le filtre de Kalman est un estimateur récursif qui permet d'estimer l'état courant. Seuls l'état précédent et les mesures actuelles sont nécessaires. L'historique des observations et des estimations n'est ainsi pas requis.

- Analyse en Composantes Curvilignes (ACC) [17]:

Cette méthode combine les avantages de l'ACP et des cartes auto-organisatrices de Kohonen.

Le principe général de la méthode est de minimiser un critère expliquant les différences de topologie entre l'espace initial et l'espace de projection.

L'algorithme de l'ACC est constitué de 2 étapes indépendantes :

1. Une étape de quantification vectorielle des individus :

Cette étape fournit les vecteurs prototypes de la distribution des vecteurs d'entrée.

Pour cette étape on peut utiliser soit un réseau de Kohonen soit une autre méthode de quantification vectorielle comme le Neural Gaz ou les K-means.

2. Une étape de réduction de dimension :

Cette étape permet de conserver au maximum la topologie des vecteurs d'entrée.

L'opération recherchée est alors une projection non-linéaire d'un espace de dimension p vers un espace de dimension inférieure q.

L'algorithme de projection est basé sur la minimisation du critère de distorsion⁴. La minimisation de ce critère, se fait par une descente de gradient.

Après convergence, les N prototypes vont permettre non seulement l'interpolation mais aussi l'extrapolation de tout vecteur n'ayant pas participé à la phase d'identification du modèle.

- Version auto adaptative de la carte [31]:

La version auto-adaptative de la carte de Kohonen augmente le nombre de neurones de la couche de sortie jusqu'à dépasser le nombre de comportements types.

Dans une première phase, un apprentissage classique se fait (car on a besoin d'un début d'organisation et le nombre initial de neurones peut-être suffisant).

Si l'apprentissage de la première phase n'est pas couronné de succès, on passe à la deuxième phase : l'ajout d'une colonne ou d'une ligne de neurones à la carte. La position d'insertion est déterminée par le neurone le plus souvent élu et le plus proche de ses voisins directs. Les poids des neurones insérés sont initialisés à la moyenne algébrique des poids de leurs voisins.

⁴ Par exemple $E = \sum_{i,j} (X_{ij} - Y_{ij})^2 F(Y_{ij})$ avec F fonction monotone positive et décroissante, qui favorise le respect local de la topologie.

Toutes ces méthodes sus mentionnées présentent plusieurs avantages : une recherche plus rapide du centre d'excitation, une convergence plus rapide et des capacités de généralisation améliorées.

Cependant, l'aboutissement à des architectures adéquates se fait souvent avec un prix élevé en temps de calcul et avec un nombre important de données d'apprentissage.

7) *Classification de Kohonen par rapport aux autres méthodes :*

Si la notion de voisinage lui est particulière, la classification de Kohonen a des analogies avec certaines autres méthodes de classification et des propriétés qui permettent d'envisager d'éventuels couplages avec elles.

Parmi ces méthodes, l'algorithme de Lloyd (Simple Competitive Learning⁵) est exactement celui de Kohonen à 0 voisin. Ces deux techniques de classification sont *mieux adaptées aux grandes bases de données* que les algorithmes de type nuées dynamiques – qui obligent à classer tous les individus à chaque itération – ou celui de la classification hiérarchique avec la distance de Ward – dans lequel il faut calculer toutes les distances entre les centres de classes.

La classification de Kohonen est *robuste* – au sens où le résultat ne peut être grandement modifié par l'ajout d'un nouvel élément à la base de donnée si celui-ci n'est pas trop extravagant (valeur erronée ou aberrante). Cette propriété est aussi vérifiée par la méthode des centres mobiles mais n'est pas partagée par la classification ascendante hiérarchique dont le résultat peut être remis en cause par l'apport d'un individu supplémentaire. Par contre, cette dernière est la seule à fournir exactement le même résultat quand on relance l'algorithme car les autres –dépendent de l'ordre de présentation des individus et de l'initialisation.

Ces méthodes peuvent être complémentaires et donner naissance à des *combinaisons hybrides* du type *centres mobiles – classification hiérarchique, carte de Kohonen – Lloyd* ou *carte de Kohonen – classification hiérarchique* [38].

⁵ Cet algorithme représente la version stochastique de l'algorithme des nuées dynamiques la plus simple. C'est pourquoi on l'appelle Algorithme d'Apprentissage Compétitif Simple (abrégé en SCL).

CHAPITRE III

APPLICATION DES CARTES AUTO

ORGANISATRICES DE KOHONEN

Introduction :

Ce troisième et dernier chapitre se divise deux parties. La première partie est consacrée à la mise en application de l'algorithme de Kohonen, aux méthodes de visualisation et aux champs d'applications de la carte. Deux applications bien particulières du réseau de Kohonen seront également décrites (application à la validation des données et application sur une base de données incomplète).

Dans la deuxième partie, une application des cartes de Kohonen à des données réelles est faite. Elle intégrera la programmation de l'algorithme à l'aide du logiciel Matlab. Deux bases de données seront considérées. La première base utilisée concerne l'indicateur composite IDH (indicateur de Développement Humain) de 118 pays. Deux cas sont considérés: le cas à une seule dimension en considérant la variable IDH, et le cas à deux dimensions en prenant en compte deux des composantes de l'IDH : l'espérance de vie à la naissance et le PIB.

La deuxième base de données décrit quant à elle les 48 wilayas d'Algérie selon quatre variables : le taux d'analphabétisme féminin, le pourcentage de chefs de ménages sans instruction, la part chefs de ménages actifs et le pourcentage de ménages propriétaires de leur logement de chaque wilaya.

Première partie : Mise en application et visualisation de l'algorithme de Kohonen

1) Mise en application d'un réseau de Kohonen :

Afin de mettre en application un réseau de Kohonen, les ajustements nécessaires à l'élaboration de celui-ci seront définis ainsi que les méthodes d'évaluation de la qualité de représentation.

A. Réglages :

Lors de la conception d'un réseau de Kohonen, le respect de certaines conditions contribuera au bon fonctionnement du modèle. Ces conditions concernent :

- *Le choix du nombre de neurones de sortie :*

Soient C le nombre de classes prévu et que le nombre de neurones de sortie est N , deux situations problématiques peuvent se présenter :

- 1) $N \ll C$: le réseau comporte trop peu de neurones et des vecteurs d'entrée appartenant à des classes différentes se trouveront représentés dans une même classe.
- 2) $N \gg C$ le réseau comporte trop de neurones et effectuera des distinctions qui n'ont pas lieu d'être.

Un bon choix pour le nombre neurone serait de prendre $2C < N \ll M$, où M est le nombre de modèles du vecteur d'entrée. T. Kohonen propose d'appliquer l'heuristique $N = 5 \times (\text{nombre d'exemples})^{0.54321}$

- *Le choix aléatoire des poids :*

Les valeurs des poids doivent être différentes sinon chacun des vecteurs initiaux seraient à une distance égale de tous les vecteurs d'apprentissage, empêchant ainsi toute possibilité de discrimination entre les vecteurs de poids. Il est préférable d'avoir des petites valeurs pour les poids

- *Le taux d'apprentissage $\alpha(t)$:*

Le taux d'apprentissage doit être une fonction positive, décroissante et monotone. Ceci permet aux poids w_i de s'ordonner pendant les premières étapes et de consacrer le reste des étapes à l'affinage des positions. On notera α_0 sa valeur initiale. Kohonen recommande de prendre α_0 proche de 1 pour les 1000 premières étapes puis de le laisser décroître. La

forme de la fonction (linéaire, exponentielle ou inversement proportionnelle à t) n'a pas d'importance.

● *La fonction de voisinage :*

Il est nécessaire d'utiliser une fonction de voisinage convexe pour éviter que la carte ne passe, en cours d'apprentissage, par des états stables, alors que les vecteurs prototypes n'ont pas encore atteint leur position finale. Une telle situation peut amener un blocage de l'organisation alors qu'elle n'est pas terminée. La fonction la plus employée est de type gaussien.

● *Le rayon d'influence :*

Le rayon d'influence du voisinage autour du neurone vainqueur $\sigma(t)$ doit être également une fonction décroissante du temps. Ce rayon sera tout d'abord grand pour permettre à la carte de se déplier puis se restreindra au seul neurone vainqueur, ou à ses voisins directs, pour affiner le placement des vecteurs prototypes.

Le choix du rayon d'influence a une grande importance car, si l'étendue du voisinage de départ est trop restreinte, la carte ne pourra pas s'ordonner globalement. Pour éviter de voir des zones de la carte désordonnées il est recommandé de prendre une valeur initiale très grande, voire même plus grande que la taille de la carte et de la laisser décroître jusqu'à 1 au cours de l'apprentissage.

Chaque neurone trouvé dans ce rayon est considéré comme appartenant au voisinage du neurone vainqueur

● *Le nombre d'échantillons à présenter t_{max} :*

Pour obtenir une bonne précision statistique, il est recommandé de prendre le nombre d'échantillons t_{max} au moins égal à 6 fois le nombre d'observations de la base de données.

● *Dimension de la carte :*

D'une manière générale, les cartes de petites tailles (de dimension 3x3, 2x2 et 1) sont relativement stables et simples à visualiser.

● *Le nombre de classes :*

En général, les méthodes d'apprentissage non supervisé donnent des résultats reproductibles s'il y a moins de 10 classes.

Cependant, le choix initial du nombre de classes est arbitraire et il n'existe pas de méthode sûre pour choisir la taille de la grille.

Pour exploiter les caractéristiques stochastiques de l'algorithme SOM, et obtenir une bonne classification et une bonne organisation, il est préférable de travailler sur des grandes cartes. Mais on peut penser que le nombre significatif de classes sera souvent plus petit que la taille de la grille ($n \times n$). D'un autre côté, il n'est ni facile ni utile d'interpréter et de décrire un trop grand nombre de classes. [21]

B. Calcul de l'erreur :

Afin d'évaluer quantitativement les représentations fournies par les cartes SOM, il est nécessaire de trouver des méthodes d'évaluation exploitables. Ces méthodes doivent garantir que :

- la carte SOM représentera de manière correcte la distribution des entrées si les vecteurs prototypes approximent la densité de probabilité des vecteurs d'entrée ;
- les relations de voisinage de l'espace d'entrée sont respectées si deux entrées proches sont représentées par deux cellules proches sur la carte ;
- les vecteurs prototypes seront ordonnés et la carte organisée si deux cellules proches possèdent deux vecteurs prototypes proches.

On trouve dans la littérature plusieurs approches. Parmi ces dernières on peut citer l'approche géométrique, l'approche discrète et une approche par ensembles ordonnés.

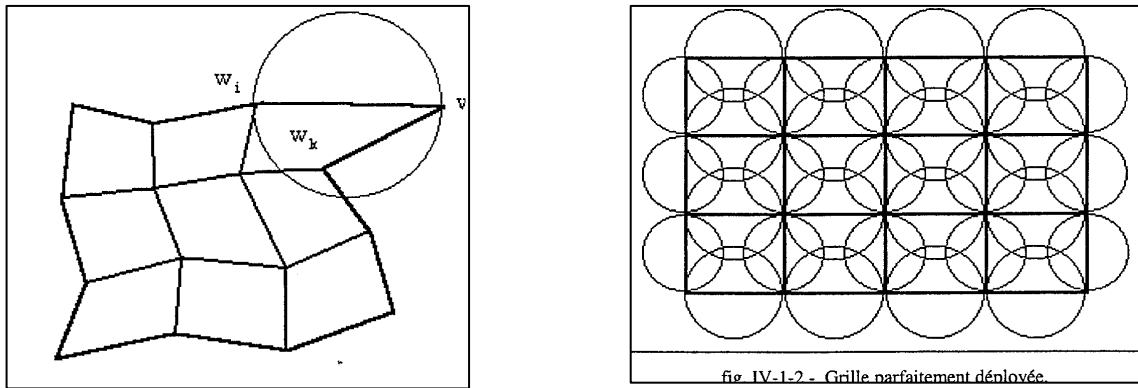
- Approche géométrique :

Pour deux neurones N_i et N_j proches dans la carte, on construit dans R^p l'hypersphère $HS(w_i, w_j)$ dont l'un des diamètres est le segment de droite rejoignant les deux représentations w_i et w_j de ces deux neurones dans R^p

Puis on recherche tous les neurones dont la représentation est incluse dans cette hypersphère et dont la distance sur la carte est supérieure à la distance de ces deux neurones.

Tous les neurones de cet ensemble violent la relation d'intermédiarité. L'évaluation de la qualité est liée à ce nombre d'intrus. [49]

Figure III.1: A droite, existence d'intrus dans la sphère la grille n'est pas tout à fait ordonnée. A gauche : une grille parfaitement organisée.



- *Approche discrète:*

Villmann propose de discrétiser l'espace des données pour faciliter la comparaison avec la carte en utilisant ce que l'on appelle les régions de Voronoï et le graphe de Delaunay.

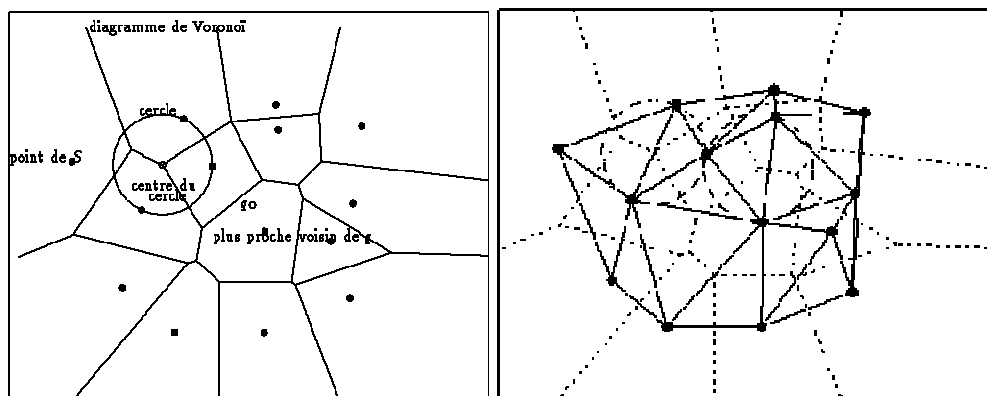
Une région de Voronoï associée à un élément P d'un espace euclidien S , est l'ensemble des points qui sont les plus proches de p que tout autre point de S .

Le graphe de Delaunay consiste à faire une triangulation particulière de l'espace S . Ceci revient à faire un découpage de l'enveloppe convexe de S en triangles dont les sommets sont des points de S et tel que les intérieurs de ces triangles sont disjoints. Une des propriétés de ce graphe : seuls les sommets contenus dans le cercle circonscrit d'un triangle quelconque de la triangulation sont les trois sommets de ce triangle.

Le graphe de Delaunay s'obtient en traçant une arête entre deux sommets du graphe de Voronoï.

Si la triangulation de Delaunay est vérifiée sur l'espace des entrées couplé des neurones de sorties, la carte sera organisée. [45]

Figure III.2: A gauche Diagramme de Voronoï et à droite, Triangulation de Delaunay (trait pleins) et diagramme de Voronoï (pointillé)



- **Approche par ensembles ordonnés :**

Si la valeur des dissimilarités entre individus est conservée alors la topologie de l'ensemble des données existant dans l'espace de départ est conservée dans l'ensemble d'arrivée.

L'évaluation du désaccord entre les deux ordres se fait par les formules suivantes : [18]

$$C = \sum_{i < j} d_X(x_i, x_j) d_C(x_i, x_j)$$

Sachant que :

$$d_X(x_i, x_j) = \|x_i - x_j\| \quad \text{et} \quad d_C(x_i, x_j) = \begin{cases} 1 & \text{si } x_i \text{ et } x_j \text{ sont voisins} \\ 0 & \text{sinon} \end{cases}$$

Même si ces trois approches sont intéressantes, elles restent difficiles à appliquer si la dimension de l'espace d'entrée est assez grande. C'est pourquoi, en général, on se contente de calculer l'erreur de dissimilarité entre chaque observation et le vecteur poids de sa classe:

$$\text{Erreur} = \frac{1}{N} \left(\sum_i \min_j \|x_i - w_j\| \right)$$

2) Méthodes d'analyse et de visualisation de la carte SOM :

La visualisation d'une carte de SOM est utilisée pour interpréter les résultats obtenus à partir du processus d'apprentissage. Celle-ci permet de discerner le nombre de victoires appartenant à chaque neurone dans la carte.

Plusieurs méthodes de visualisation peuvent être utilisées pour représenter les résultats obtenus. Cette section présente les quatre méthodes les plus utilisées : les U-matrices, les histogrammes de densité des données, les plans, et le mappage de Sammon.

- *Les U-matrices :*

Les U-matrices permettent de visualiser les groupes implicites contenus dans l'échantillon. Le principe est de calculer une matrice des distances entre les vecteurs poids des neurones adjacents. Il suffit alors de représenter ces distances dans un espace à deux dimensions. Des niveaux de gris sont généralement utilisés. Les zones claires séparent des neurones proches et les zones foncées des neurones éloignés. (Figure (a))

- *Les histogrammes de densité des données :*

Les histogrammes de densité des données permettent, lorsque l'apprentissage est terminé, de représenter sur un plan le nombre de victoires de chaque neurone.

Le nombre de victoires par neurone déterminera sa couleur (par exemple foncé si le neurone a eu plusieurs victoires), sa taille ou bien sa hauteur (dans le cas d'une représentation tridimensionnelle du SOM). (Figure (b))

- *Les plans :*

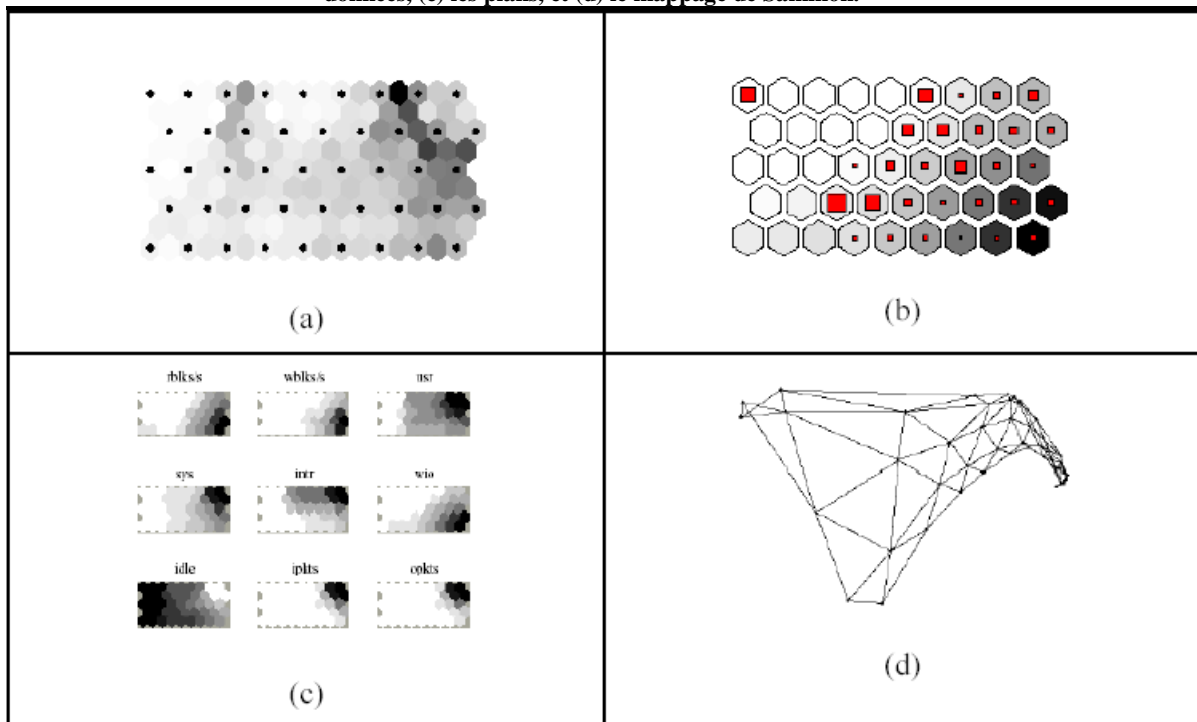
Les plans permettent de visualiser les valeurs relatives d'une des composantes des vecteurs de poids sur une carte 2D. Comme pour les U-matrices, on utilise généralement des représentations par niveaux de gris. On peut par exemple associer les petites valeurs au gris clair et les grandes valeurs au gris foncé. (Figure c)

- *Le mappage de Sammon :*

Le mappage de Sammon est surtout utilisé lors des phases d'élaboration des cartes auto organisatrices. C'est une méthode itérative de représentation de données multidimensionnelles en deux dimensions.

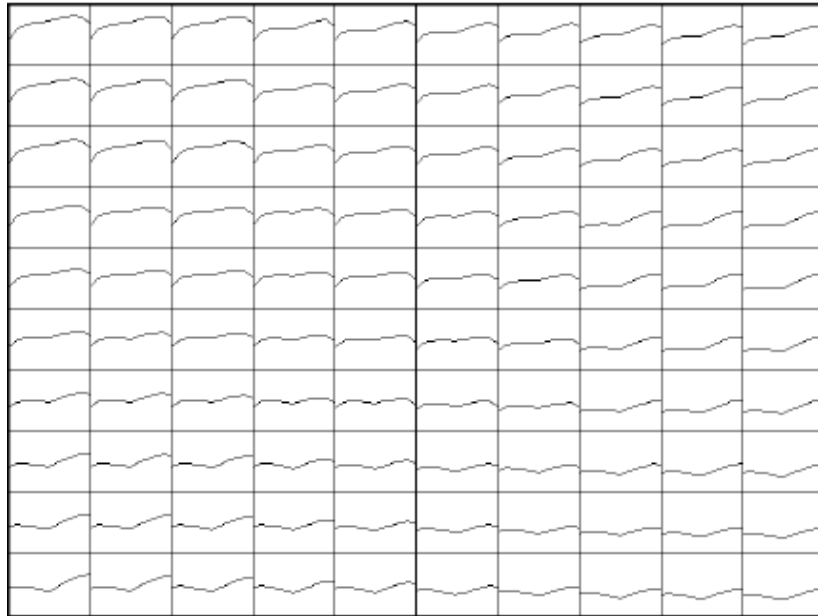
Le mappage de Sammon est une méthode antérieure et indépendante des cartes de Kohonen. Cependant la représentation des vecteurs de poids sur un plan, en connectant les neurones voisins par des lignes, permet d'obtenir une représentation ordonnée reproduisant la forme des données utilisées en entrée. (Figure (d)) [46]

Figure III.3: Différentes méthodes de visualisations de SOM : (a) U-matrices, (b) Histogrammes de densité des données, (c) les plans, et (d) le mappage de Sammon.



Une autre méthode de visualisation des cartes SOM organisées en ficelle ou en grille, serait de présenter chaque neurone par une case et de dessiner à l'intérieur de chaque case les valeurs des composantes de chaque neurone.

Figure III.4 : Visualisation des contenus des classes d'une carte (10 x 10)



Dans ce qui précède, les différents aspects de l'algorithme de Kohonen ont été traités sans tenir compte de la nature des données. Cependant le traitement des données qualitatives est spécifique. Les caractéristiques de ce traitement feront l'objet de la section suivante.

3) *Traitement de données qualitatives par l'algorithme Kohonen :*

Dans sa forme originelle, l'algorithme de Kohonen traite des données quantitatives à valeurs continues réelles, pour lesquelles chaque observation est décrite par un vecteur réel. Cependant dans les applications réelles, les individus peuvent aussi être décrits par des variables de nature qualitative (ou catégorielle). C'est le cas par exemple, pour les données d'enquêtes, où les réponses aux questions peuvent avoir un nombre fini de modalités de réponse (par exemple : sexe, catégorie socioprofessionnelle, niveau de revenu, type d'emploi, lieu d'habitation, niveau d'éducation, etc...).

La plupart du temps, les variables qualitatives ne peuvent pas être utilisées telles quelles. En particulier pour faire un apprentissage de Kohonen, cela n'a aucun sens de les utiliser comme des variables numériques même si les modalités sont codées par des nombres. Ainsi les données qualitatives nécessitent un traitement spécifique.

Dans ce qui suit, seront traités successivement le cas où la base de données contient uniquement des variables qualitatives, et le cas où elle comporte des données qualitatives et quantitatives.

3.1) Base de données constituée uniquement de variables qualitatives :

L'application de l'algorithme de Kohonen aux données qualitatives requiert un traitement spécial de la base de données. Pour ce faire, il est nécessaire de définir deux sortes de tableaux d'analyse des données : le tableau disjonctif complet et la table de Burt. La réalisation de ces deux tableaux est présentée dans le détail en annexe. [46]

3.1.1) L'algorithme KACM (Kohonen Analyse des Correspondances Multiples) :

Cet algorithme ne prend en considération que les modalités. Soit un réseau de Kohonen $n \times n$ (grille), muni de sa topologie habituelle et la table de Burt corrigée B^c relative à la base de données utilisée.

L'apprentissage à chaque étape consiste à :

- Présenter au hasard une entrée $r(j)$ (une ligne de la *matrice corrigée* B^c) ;
- Rechercher son unité gagnante u_0 qui minimise $\|r(j) - C_u\|^2$ pour toutes les unités u ;
- Mettre à jour les vecteurs poids de l'unité gagnante et de ses voisines :

$$C_u^{\text{nouveau}} = C_u^{\text{ancien}} + \varepsilon \sigma(u, u_0) (r(j) - C_u^{\text{ancien}})$$

Où ε est le paramètre d'adaptation et σ la fonction de voisinage définis de la même manière que précédemment.

Après apprentissage, chaque profil ligne $r(j)$ est représenté par son unité gagnante.

Grâce à la propriété de conservation de la topologie de l'algorithme de Kohonen, la représentation des M entrées sur la grille ($n \times n$) souligne la *proximité* entre les modalités des K variables.

Une fois la convergence réalisée, une classification organisée de toutes les modalités est obtenue, où les modalités liées appartiennent à la même classe ou à des classes voisines.

Cette méthode (KACM) est une alternative très intéressante de l'AFCM classique. [46]

3.1.2) Traitement des individus/modalités :

Afin de faire une analyse sur les individus, l'application de l'algorithme de Kohonen aux lignes du tableau disjonctif complet D^c , donne une carte SOM qui analyse les individus seulement. Mais pour représenter simultanément les modalités et les individus, il est nécessaire d'utiliser d'autres algorithmes :

- **KACM_ind** (*Kohonen Analyse des Correspondances Multiples avec individus*) :

Les modalités sont affectées aux classes après apprentissage, comme données supplémentaires.

Une carte de Kohonen est construite avec les individus en utilisant l'algorithme SOM appliqué au Tableau Disjonctif Complet Corrigé et les modalités sont projetées comme des données supplémentaires.

L'algorithme SOM est appliqué aux lignes de ce tableau corrigé. Chaque modalité j est représentée par un M vecteur, qui est le *vecteur moyen* de tous les individus partageant cette modalité.

Ses coordonnées sont:

$$\frac{b_{jl}}{b_j \sqrt{b_l} \sqrt{K}} \quad \text{Pour } l=1, \dots, M$$

Chaque vecteur moyen est affecté à la classe de Kohonen de son plus proche code vecteur. KACM_ind nécessite plus d'itérations que KACM. En effet, les individus (qui sont généralement plus nombreux que les modalités) sont classés en premier lieu. Ensuite les vecteurs poids sont calculés pour être utilisés comme prototypes et assigner les modalités considérées comme données supplémentaires.

La visualisation n'est pas toujours utile, surtout lorsqu'il y a trop d'individus. Mais la carte peut montrer des associations entre modalités ou groupes de modalités et sous-ensembles d'individus. [46]

- **KDISJ** (*Kohonen sur tableau Disjonctif*) :

Deux algorithmes SOM sont utilisés sur les lignes (individus) et sur les colonnes (modalités) de D^c (*tableau disjonctif complet corrigé*), l'association entre modalités et individus est contrainte durant tout l'apprentissage.

Soient un tableau disjonctif complet corrigé (D^c) et un réseau de Kohonen. A chaque unité u est associé un vecteur code C_u formé de $(M + N)$ composantes, les M premières évoluent dans l'espace des individus (représentées par les lignes de D^c), les N dernières dans l'espace des modalités (représentées par les colonnes de D^c). Soit :

$$C_u = (C_M, C_N)_u = (C_{M,u}, C_{N,u})_u$$

Les étapes d'apprentissage du réseau de Kohonen sont doubles :

Une ligne de D^c (c'est-à-dire un individu i) est tirée, puis une colonne (c'est-à-dire une modalité j). A chaque individu i tiré, est associée la modalité $j(i)$ définie par :

$$j(i) = \text{Arg max}_j d_{ij}^c = \text{Arg max}_j \frac{d_{ij}}{\sqrt{Kd_{.j}}}$$

Ensuite, un vecteur individu étendu est créé $X = (i, j(i)) = (XM, XN)$, de dimension $(M + N)$. L'unité gagnante est définie comme étant celle dont le vecteur code est le plus proche, au sens de la distance euclidienne restreinte aux M premières composantes.

Notons u_0 cette unité gagnante. Les vecteurs codes de l'unité u_0 et ses voisins sont rapprochés du vecteur complété $(i, j(i))$, selon la loi usuelle de Kohonen :

$$\left\{ \begin{array}{l} u_0 = \text{Arg min}_u \|X_M - C_{M,u}\| \\ C_u^{\text{nouveau}} = C_u^{\text{ancien}} + \varepsilon \sigma(u, u_0)(X - C_u^{\text{ancien}}) \end{array} \right.$$

Où ε est le paramètre d'adaptation (positif, décroissant), et σ est la fonction de voisinage telle que $\sigma(u, u_0)$ vaut 1 si u et u_0 sont voisines sur le réseau de Kohonen, et 0 sinon.

Quand une modalité j de dimension N (une colonne de D^c) est tirée, aucun individu ne lui est associé. En effet, par construction, il existe plusieurs individus ex-æquo et le choix serait arbitraire.

L'unité gagnante v_0 est définie comme étant celle dont le vecteur code est le plus proche, au sens de la distance euclidienne restreinte aux N dernières composantes.

Les N dernières composantes du vecteur code gagnant associé à v_0 et de ses voisins sont rapprochés de celles du vecteur modalité j , sans modifier les M premières composantes.

Soit Y le vecteur colonne de dimension N correspondant à la modalité j . Cette étape peut s'écrire :

$$\left\{ \begin{array}{l} v_0 = \text{Arg min}_u \|Y - C_{N,u}\| \\ C_{N,u}^{\text{nouveau}} = C_{N,u}^{\text{ancien}} + \varepsilon \sigma(u, v_0)(Y - C_{N,u}^{\text{ancien}}) \end{array} \right.$$

Où les M premières composantes ne sont pas modifiées.

Un classement classique de Kohonen est fait sur les individus, un autre sur les modalités, tout en les maintenant associés. Après convergence, les individus et les modalités sont classés dans les classes de Kohonen. Des individus ou modalités « proches » sont classés dans la même classe ou dans des classes voisines. Il suffit en général de faire $20(M+N)$ itérations pour obtenir la convergence. [46]

Le cas traité dans la section suivante, considère, une base de données constituée d'un mélange de variables quantitatives et qualitatives.

3.2) Base de données constituée de variables quantitatives et qualitatives:

Quand la base des données inclut des variables quantitatives et qualitatives, la première idée est de faire une classification sur les variables quantitatives, et puis de croiser cette classification avec les autres variables qualitatives.

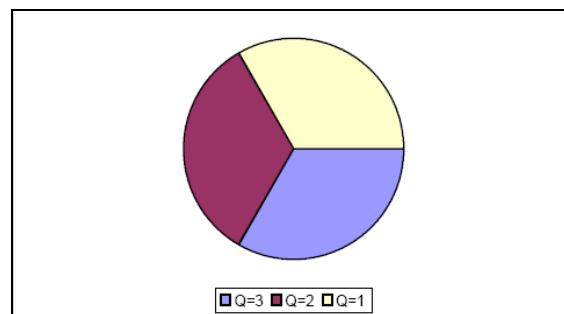
- Classification de Kohonen et variables qualitatives :

Soit un algorithme de Kohonen, appliqué à une base de données d'individus décrits par des variables quantitatives.

Pour interpréter les classes selon des variables qualitatives non utilisées dans l'algorithme de classement de Kohonen, il peut être intéressant d'étudier la répartition de leurs modalités dans chaque classe.

Les statistiques élémentaires sont calculées dans chaque classe et à l'intérieur de chaque cellule on constitue un camembert montrant comment sont réparties les modalités de chacune des variables qualitatives. [46]

Figure III.5: Camembert d'une variable qualitative Q possédant trois modalités.



4) Champs d'applications des cartes de Kohonen :

La méthode SOM est, en premier lieu, une méthode de représentation de données. Cependant, les propriétés intéressantes de la carte font qu'elle est utilisée dans beaucoup de domaines, comme la classification, la discrimination, l'optimisation et la prévision.

- Classification :

L'application des cartes SOM comme outils de classification se fait dans plusieurs domaines, dont, à titre d'exemples, la médecine et l'économie.

Dans le domaine médical, les cartes SOM ont été notamment utilisées pour classer des tumeurs cérébrales dans des bases de données de grande taille, et pour classer des électroencéphalogrammes épileptiques (EEG).

En économie, les chercheurs du Samos ont réalisé une étude sur une approche dynamique des styles des gérants en France, l'objectif étant d'analyser les comportements d'une sélection d'actifs et leur évolution sur une période. L'étude a permis de donner une nouvelle classification des styles des gérants suivant le positionnement sur le marché français [30].

Une autre application dans le domaine économique : une classification des pays européens selon leurs performances économiques dans le temps [1].

– Discrimination :

La première application des SOM sur des données réelles et complexes fut le Phonetic Typewriter de Kohonen. Toutes sortes d'applications sur le traitement du langage et la parole ont suivi : reconnaissance de formes statistiques (parole) et sémantique. Une application de SOM et de chaînes de Markov permet de faire de la reconnaissance de phonèmes.

– Optimisation :

Un domaine d'application plus marginal des SOM est celui de l'optimisation. Une carte monodimensionnelle peut être assimilée à un parcours quasi optimal passant par des points définis. L'idée de résoudre le problème du voyageur de commerce en découle.

– Prévision :

Les cartes de Kohonen peuvent aussi être utilisées comme outil de prédiction.

Un exemple de cette application : la prédiction de séries temporelles financières par double carte de Kohonen. En effet, une équipe française a travaillé sur la prédiction de l'indice boursier DAX30 en utilisant les cartes auto organisatrices de Kohonen [2]. Dans le domaine médical, l'article proposé par *B. Wynsa et al*, montre que les cartes SOM sont utilisées pour prédire, avant son apparition, l'arthrite chronique. Dans cet article, les cartes SOM donnent de très bons résultats contrairement aux autres méthodes neuronales [12].

Ces domaines d'application ne sont pas les seuls, il en existe bien d'autres. Dans ce qui suit, deux applications bien particulières des cartes de Kohonen seront présentées. Elles

sont relatives aux pré-traitements des données : la validation de données multi paramètres et le traitement des données manquantes.

4.1) Application à la validation des données :

Les cartes topologiques de Kohonen ne sont pas employées seulement comme une méthode de visualisation et de classification des données de grandes dimensions, elles peuvent être utilisées aussi pour détecter des données atypiques en contrôlant la distance entre chaque vecteur d'entrée et l'unité gagnante.

- Principe de la méthode :

Soit un ensemble de N prototypes ou vecteurs de référence $\{w_1, \dots, w_N\}$. Chaque prototype w_k est représentatif d'une classe W_k de vecteurs de X . L'espace de référence est ainsi divisé en N classes.

Soit $K_i(x)$ l'activation du prototype i pour le vecteur d'entrée x :

$$K_i(x) = \exp \left(\frac{-1}{2\sigma_i^2} \|x - w_i\|^2 \right)$$

Où σ_i^2 est un paramètre définissant la région d'influence du prototype i . et est égale à la moyenne empirique des variances des n vecteurs d'entrées activant le prototype i . Plus σ_i^2 est grand, plus la zone d'influence de w_i est grande et donc, plus l'activation de $K_i(x)$ est proche de 1.

Le vecteur x est considéré comme invalide si l'activation $K_b(x)$ du prototype le plus proche w_b est inférieure à un certain seuil fixé α .

Le rejet peut être dû à une limitation de l'ensemble d'apprentissage. Il est donc nécessaire de stocker ces vecteurs rejetés pour permettre une interprétation ultérieure et un éventuel ré-apprentissage du système.

Pour déterminer plus précisément quelles sont les composantes invalides du vecteur x , la contribution de chaque composante est examinée séparément.

Pour chaque composante j du vecteur x , la distance avec les composantes du plus proche prototype est calculée: $|x_j - w_{bj}|$. La composante ayant la distance la plus élevée est déconnectée et un nouveau prototype gagnant est calculé uniquement avec les composantes valides du vecteur x .

Cette procédure de rejet des données invalides permet d'éliminer les vecteurs x qui paraissent peu vraisemblables pour chacune des densités de classe, et permet aussi la

détection de nouveaux échantillons qu'il faut peut-être inclure dans l'ensemble d'apprentissage.

Un paramètre important à définir lors de la validation des données, le paramètre α .

– *Définition du seuil α :*

Le paramètre α représente le seuil en dessous duquel l'activation $K_b(x)$ du prototype le plus proche w_b permet de considérer le vecteur x comme invalide.

Soit l'ensemble de n vecteurs x de l'ensemble d'apprentissage. Supposons que 1% de ces données soient invalides ou aberrantes. L'activation $K_b(x)$ du plus proche prototype est calculée pour ces n vecteurs. 1% des données de l'ensemble d'apprentissage est donc considéré comme ayant une activation trop faible pour être considéré comme valide. Les données sont classées par ordre croissant d'activation. Le seuil α correspond donc à l'activation $K_b(x)$ de rang $n/100$. [10]

Après avoir présenté l'application de l'algorithme SOM à la détection de données atypiques, la section suivante abordera la problématique du traitement de données manquantes.

4.2 Adaptation de l'algorithme de Kohonen à des données incomplètes:

Le traitement des données avec observations manquantes est un problème concret et toujours embarrassant lorsqu'il s'agit de données réelles. En effet, on est très souvent en présence d'observations pour lesquelles l'ensemble des valeurs des variables descriptives n'est pas disponible, et ceci se produit pour de nombreuses raisons : erreurs de saisie, rubriques non renseignées dans des enquêtes, valeurs aberrantes qu'on préfère supprimer, données recueillies difficilement, statistiques officielles non disponibles, etc.

La plupart des logiciels statistiques suppriment purement et simplement les observations incomplètes, mais si cela n'a pas de conséquences pratiques lorsque les données sont très nombreuses, il peut enlever tout intérêt à l'étude si le nombre de données restantes est trop faible.

Pour éviter de supprimer ainsi les données, une valeur manquante peut être remplacée par la moyenne de la variable correspondante, mais cette moyenne peut être une très mauvaise approximation dans le cas où la variable présente une grande dispersion.

L'algorithme de Kohonen supporte parfaitement la présence de données manquantes et permet même leur estimation.

- Principe de la méthode :

1^{ère} étape : « l'application de l'algorithme »

Soient des vecteurs d'entrée à valeurs réelles de dimension p . Parmi ces observations, se retrouvent des vecteurs incomplets.

Si un vecteur de données incomplet x est présenté, l'ensemble M_x des numéros des composantes manquantes est d'abord déterminé. M_x est un sous ensemble de $\{1, 2, \dots, p\}$.

Si (C_1, C_2, \dots, C_n) est l'ensemble des vecteurs-codes à cet instant, le vecteur-code gagnant $i_O(C, x)$ associé à x est calculé en posant :

$$i_O(C, x) = \text{Arg min}_i \left\| x - c_i \right\|$$

Où la distance

$$\left\| x - c_i \right\|^2 = \sum_{k \notin M_x} (x_k - C_{i,k})^2$$

est calculée sur les composantes présentes dans le vecteur x .

La mise à jour des poids peut se faire de deux manières :

- a) La mise à jour des vecteurs codes (le gagnant et ses voisins) ne porte que sur les composantes présentes dans le vecteur ;
- b) Si la base de données est suffisamment grande, la construction de la carte peut se faire sans l'utilisation des vecteurs incomplets.

Après construction de la carte, les vecteurs incomplets sont classés en les affectant dans la classe dont le vecteur code est le plus proche, au sens de la distance restreinte aux composantes présentes.

Cela donne d'excellents résultats, dans la mesure bien sûr où une variable n'est pas complètement absente ou presque, et aussi dans la mesure où les variables sont corrélées, ce qui est le cas dans la plupart des jeux de données réelles. **[10]**

2^{ème} étape : « Estimation des valeurs manquantes » :

Une des applications les plus intéressantes de l'algorithme, est qu'il est possible d'estimer à posteriori les valeurs manquantes. Plusieurs méthodes peuvent être utilisées.

1^{ère} méthode : Si M_x est l'ensemble des numéros des composantes manquantes de l'observation x , et si x est classé dans la classe i , pour chaque indice k de M_x , x est estimé par :

$$\hat{x}_k = c_{i,k}$$

Comme la fin de l'apprentissage de l'algorithme de Kohonen se fait à « zéro voisin », les vecteurs codes sont asymptotiquement proches des moyennes de leur classe.

Cette méthode d'estimation consiste donc à estimer les valeurs manquantes d'une variable par sa moyenne dans la classe. Il est clair que cette estimation est d'autant plus précise que les classes formées par l'algorithme sont homogènes et bien séparées les unes des autres.

De nombreuses simulations ont montré qu'en présence de variables corrélées, la précision de ces estimations est remarquable [43].

2^{ème} méthode :

Soit x un nouveau vecteur, composé de deux parties x_o et x_m , contenant respectivement les variables observées et manquantes. L'objectif est de reconstruire x_m à l'aide des informations fournies par la carte topologique de Kohonen. Intuitivement, la méthode proposée dans ce paragraphe, repose sur des similitudes entre ce nouveau vecteur $x = (x_o, x_m)$ et les vecteurs de référence ou prototypes w_k .

Cette méthode simple, estime les données manquantes par la valeur de la composante correspondante du prototype gagnant w_b :

$$\forall l \in M(x), \hat{x}_l = w_{bl}$$

Cependant, elle est très sensible au changement de prototypes entre deux vecteurs successifs x .

3^{ème} méthode :

La troisième méthode n'est qu'une amélioration de la deuxième. En effet, elle permet de remédier au problème du changement de prototypes entre deux vecteurs successifs.

Cette méthode prend en compte l'influence des k plus proches prototypes. Chaque composante manquante de x_m est estimée par une combinaison des composantes correspondantes des k plus proches prototypes :

$$\forall l \in M(x), \hat{x}_l = \frac{\sum_{i=1}^k K_i(x^0) w_{il}}{\sum_{i=1}^k K_i(x^0)}$$

Cette méthode propose donc, d'estimer toutes les valeurs manquantes d'un vecteur dans un même système, quel que soit le nombre de variables disponibles. Mais il est bien évident que la précision de la reconstruction dépendra du nombre de variables disponibles par rapport au nombre de variables manquantes.

Après avoir vu les différentes possibilités qu'offre l'utilisation de la carte SOM, trois applications à données réelles seront étudiées dans la deuxième partie de ce chapitre.

Deuxième partie

Application de l'algorithme de Kohonen à des données réelles

Introduction :

Trois applications de l'algorithme de Kohonen sont présentées dans cette partie. Pour les deux premières applications, une base de données du PNUD relative à l'année 2004 pour 118 pays sera utilisée.

Dans un premier temps, la variable de Développement Humain (IDH) seulement des 118 pays sera considérée. Par la suite, une classification à deux dimensions sera faite en prenant en compte le PIB et l'espérance de vie à la naissance. Cette 2^{ème} classification permettra de voir comment la carte se déploie pendant l'apprentissage.

La dernière application de l'algorithme concernera les 48 wilayas de l'Algérie en utilisant les données relatives au Recensement Général de la population de l'année 1998.

Mais avant de travailler ces deux bases de données, une présentation des données ainsi que l'algorithme élaboré à l'aide du logiciel Matlab doit être faite.

1. Bases de données:

Deux bases de données sont utilisées. Une relative au RGPH de 1998 et une autre relative aux données du PNUD.

PNUD :

Chaque année, le programme des Nations Unies pour le Développement (PNUD) établit un classement des pays suivant un indicateur composite « l'IDH ».

L'Indice de Développement Humain (IDH) évalue le niveau de développement humain des pays du monde en se basant sur des données d'organismes nationaux ou internationaux.

L'IDH se calcule comme moyenne des indices de longévité, niveau d'éducation et niveau de vie.

La classification effectuée par le PNUD et la méthode de calcul de l'IDH sont affichées en annexe.

Les pays retenus pour la classification de Kohonen (au nombre de 118) sont décrits par 3 variables : l'IDH, l'espérance de vie à la naissance et le PIB (produit intérieur brut).

RGPH 1998 :

Pour cette base de données, les variables considérées sont : le taux d'analphabétisme féminin, le pourcentage de chefs de ménages sans instruction, la part chefs de ménages actifs et le pourcentage de ménages propriétaires de leur logement de chaque wilaya.

2. Programmation de l'algorithme de Kohonen :

Afin de mettre en application l'algorithme de Kohonen, une programmation de cet algorithme à l'aide du logiciel Matlab a été faite.

Le programme ne tient compte que des cartes carrées et ce pour faciliter la définition du voisinage de la carte de Kohonen,

Les hypothèses établies pour l'algorithme sont :

- La dimension des différentes cartes a été choisie selon l'heuristique suivante : $Ne=5*N^{0.54321}$. Avec N le nombre d'individus de la base de données. Les cartes utilisées pour traiter la base de données du PNUD comprennent 64 neurones (8 x 8), et celle relative aux données du RGPH contient 25 neurones (5 x 5) ;

- La distance choisie est la distance euclidienne ;

- Le voisinage est de type rectangulaire. Les coordonnées sur la carte de Kohonen des voisins du neurone vainqueur $w(i, j)$ sont :

$$\text{Voisins } (w(i, j)) = \{w(i-m, j), w(i+m, j), w(i, j+m), w(i, j-m) \text{ tel que } 0 < i-m, i+m, j-m, j+m < n\} ;$$

- La fonction de voisinage retenue est la fonction gaussienne. Cette fonction est invariante par translation, ne dépend pas du neurone vainqueur et présente une rapidité de convergence qui la rend privilégiée ;

- Le taux d'apprentissage appliqué est défini par :

$$\alpha(t) = \frac{\alpha_0}{1 + 100t/T} \quad \text{et } 0 \leq \alpha(t) \leq 0.9999$$

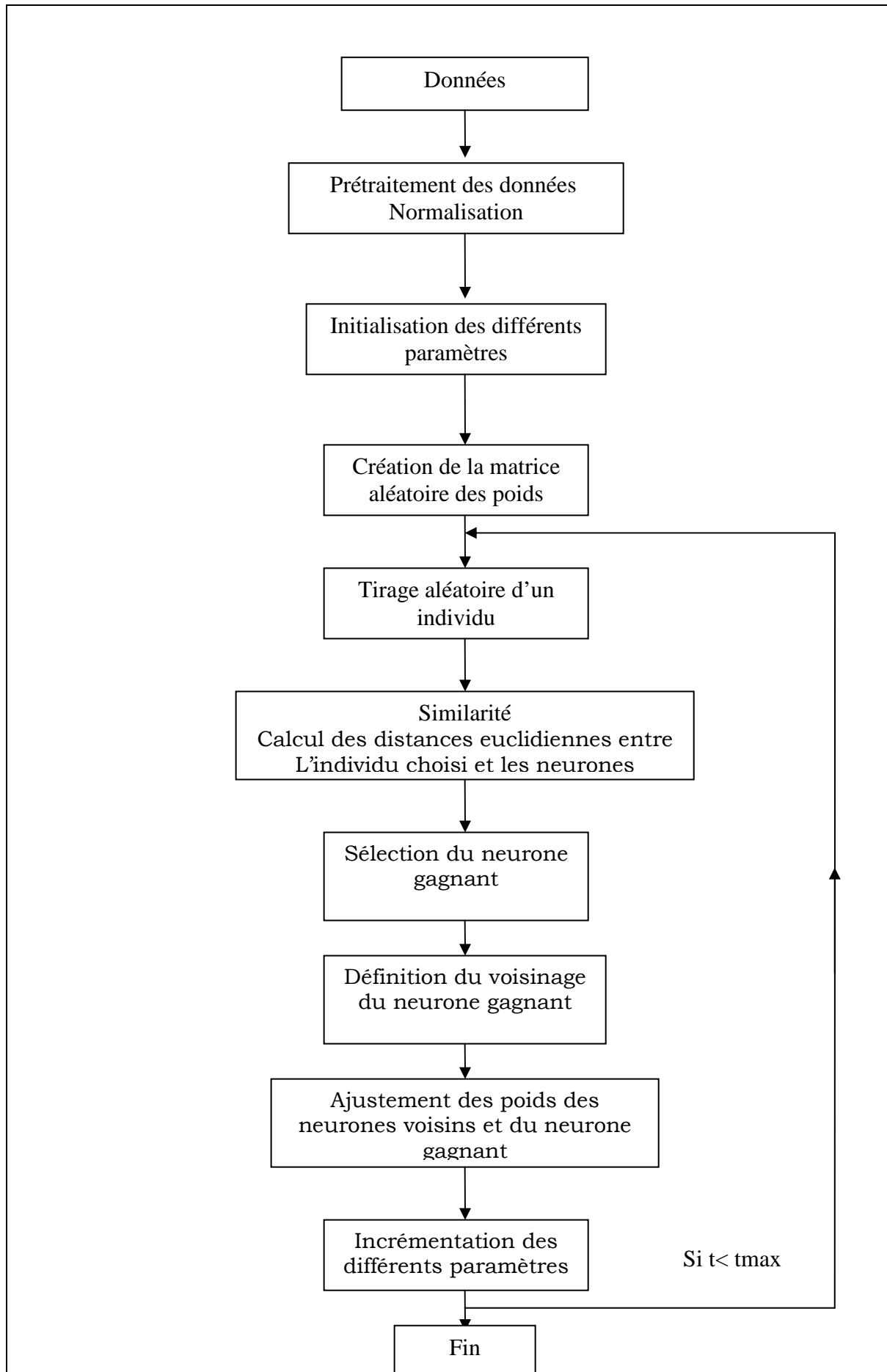
- Le rayon de voisinage est déterminé par la relation :

$$R(t) = R(t)_{\text{initial}} \left(\frac{R(t)_{\text{final}}}{R(t)_{\text{initial}}} \right)^{\frac{t}{t_{\text{max}}}}$$

Le rayon initial est égal à la taille de la carte (Ne).

- Les données sont centrées et réduites.

- Le nombre maximum d'itérations est de 700 itérations pour les données du PNUD et de 300 pour celles du RGPH ($T_{\max}=6 \times N$).



3. Classification des 118 pays :

Dans cette section, les pays sont décrits par des variables permettant de cerner leur niveau de développement humain : l'IDH, l'espérance de vie à la naissance et le PIB.

Si l'espérance de vie et le Pib sont des variables connues, il n'en est pas de même pour l'IDH qui mérite d'être défini avant d'être utilisé.

- Définition de l'IDH :

L'indice de développement humain ou IDH, est un indice statistique composite, créé par le Programme des Nations Unies pour le Développement (PNUD), évaluant le niveau de développement humain des pays du monde.

L'IDH est un indice sans unité, compris entre 0 (exécrable) et 1 (excellent) et est calculé par la moyenne de trois indices quantifiant respectivement :

- La santé/ longévité : mesurée par l'espérance de vie à la naissance, qui permet de mesurer la satisfaction des besoins matériels essentiels tels que l'accès à une alimentation saine, à l'eau potable, à un logement décent, à une bonne hygiène de vie et aux soins médicaux ;
- Le savoir ou le niveau d'éducation. Il est mesuré par le taux d'alphabétisation des adultes (pourcentage des 15 ans et plus sachant écrire et comprendre aisément un texte court et simple traitant de la vie quotidienne) et le taux brut de scolarisation (mesure combinée des taux pour le primaire, le secondaire et le supérieur). Il traduit la satisfaction des besoins immatériels tels que la capacité à participer aux prises de décision sur le lieu de travail ou dans la société ;
- le niveau de vie (logarithme du produit intérieur brut par habitant en parité de pouvoir d'achat), afin d'englober les éléments de la qualité de vie qui ne sont pas décrits par les deux premiers indices tels que la mobilité ou l'accès à la culture.

L'IDH vaut :

$$IDH = \frac{A + D + E}{3}$$

Où A, D et E sont respectivement les indices de longévité, niveau d'éducation et niveau de vie. Le calcul des indices est donné dans le tableau ci-dessous :

Tableau 1 : Calcul des indices composant l'indice de développement humain.

Indice	Mesure	Valeur minimale	Valeur maximale	Formule
Longévité	espérance de vie à la naissance (EV)	25 ans	85 ans	$A = \frac{EV - 25}{60}$
Education	Taux d'alphabétisation (TA)	0%	100%	$D = \frac{2TA + TBS}{3}$
	Taux brut de scolarisation (TBS)	0%	100%	
Niveau de vie	logarithme du PIB par habitant en parité de pouvoir d'achat	100 USD	40 000 USD	$E = \frac{\log_{10} PIB - 2}{2.60206}$

Suivant l'IDH, le PNUD établit à chaque rapport du développement humain un classement des pays. La classification établie en 2004 est affichée dans le tableau suivant :

Tableau 2 : La classification du PNUD 2004

Pays	IDH	Pays	IDH	Pays	IDH
Norvège	0,968	Jamahiriya arabe libyenne	0,818	Maroc	0,646
Islande	0,968	Oman	0,814	Inde	0,619
Australie	0,962	Roumanie	0,813	Cambodge	0,598
Canada	0,961	Arabie saoudite	0,812	Ghana	0,553
Irlande	0,959	Malaisie	0,811	Pakistan	0,551
Suède	0,956	Maurice	0,804	Mauritanie	0,55
Suisse	0,955	Albanie	0,801	Bangladesh	0,547
Pays-Bas	0,953	Brésil	0,8	Népal	0,534
Japon	0,953	Kazakhstan	0,794	Cameroun	0,532
Finlande	0,952	Venezuela	0,792	Papouasie-Nouvelle-Guinée	0,53
France	0,952	Colombie	0,791	Haïti	0,529
Etats-Unis	0,951	Ukraine	0,788	Soudan	0,526
Danemark	0,949	Thaïlande	0,781	Kenya	0,521
Espagne	0,949	Chine	0,777	Djibouti	0,516
Autriche	0,948	Arménie	0,775	Timor-Leste	0,514
Belgique	0,946	Turquie	0,775	Zimbabwe	0,513
Royaume-Uni	0,946	Pérou	0,773	Togo	0,512
Luxembourg	0,944	Jordanie	0,773	Yémen	0,508
Nouvelle-Zélande	0,943	Liban	0,772	Ouganda	0,505
Italie	0,941	Philippines	0,771	Sénégal	0,499
Allemagne	0,935	Tunisie	0,766	Érythrée	0,483
Israël	0,932	Iran	0,759	Nigéria	0,47
Grèce	0,926	Paraguay	0,755	Tanzanie	0,467
Chypre	0,903	Maldives	0,741	Guinée	0,456
Portugal	0,897	Jamaïque	0,736	Rwanda	0,452
République tchèque	0,891	Algérie	0,733	Angola	0,446
Koweït	0,891	Viet Nam	0,733	Bénin	0,437
Malte	0,878	Indonésie	0,728	Zambie	0,434
Qatar	0,875	République arabe syrienne	0,724	Côte d'Ivoire	0,432
Hongrie	0,874	Nicaragua	0,71	Congo	0,411
Pologne	0,87	Égypte	0,708	Éthiopie	0,406
Argentine	0,869	Ouzbékistan	0,702	Tchad	0,388
Émirats arabes unis	0,868	Mongolie	0,7	Mozambique	0,384
Chili	0,867	Bolivie	0,695	République centrafricaine	0,384
Bahreïn	0,866	Guatemala	0,689	Mali	0,38
Uruguay	0,852	Gabon	0,677	Niger	0,374
Bahamas	0,845	Afrique du Sud	0,674	Burkina Faso	0,37
Cuba	0,838	Tadjikistan	0,673	Sierra Leone	0,336
Mexique	0,829	Botswana	0,654		
Bulgarie	0,824	Namibie	0,65		

Pour les deux classifications effectuées dans cette section, les données utilisées sont relatives au rapport sur le développement humain du PNUD de l'année 2004. [53]

Les cartes de Kohonen créées sont carrées de dimension (8x8) soit 64 neurones.

3.1 Classification des pays selon l'indicateur IDH :

La classification des 118 pays selon la variable IDH a permis de distinguer 47 groupes de pays.

La figure suivante montre la répartition des différents pays sur la grille. Chacun des groupes, englobe un certain nombre de pays dont les valeurs de l'IDH sont proches.

En partant de gauche à droite selon deuxième diagonale, le niveau d'IDH décroît.

Ainsi, les pays riches dont l'IDH est supérieur à 0.8, sont positionnés en haut à du coté gauche de la carte. A titre d'exemple, le groupe 1 comprend : la Finlande, la France, la, les Etats-Unis le Danemark, l'Espagne, l'Autriche, le Royaume Uni, la Belgique. La Nouvelle Zélande et le Luxembourg. Les valeurs d'IDH des ces pays varient entre 0.943 et 0.952. Le groupe 2 quant à lui regroupe : la Norvège, l'Islande, l'Australie, le Canada, l'Irlande, la Suède, la Suisse, les Pays Bas et le Japon. Ces deux groupes de pays possèdent les plus fortes valeurs d'IDH (respectivement 0.968 ; 0.968 ; 0.962 ; 0.961; 0.959. 0.956 ; 0.955 ; 0.953 ; 0.953).

Les pays dont l'IDH est compris entre 0.5 et 0.8 et qui sont beaucoup plus nombreux, forment la partie centrale de la carte. L'Algérie se trouve dans le groupe 57 avec le Viet Nam. Ces deux pays possèdent la même valeur d'IDH (0.733).

La Tunisie avec un IDH de 0.766 et le Maroc (IDH de 0.646) se trouvent dans le groupe 28 et 30 respectivement.

Enfin, les pays très pauvres se trouvent en bas de la carte à droite. Par exemple, le groupe 64 est constitué du Tchad, Mozambique, République Centre Africaine, Mali, Niger, le Burkina Faso et Sierra Leone. Ces pays sont les derniers de la liste selon le classement du PNUD par la variable IDH. Les valeurs d'IDH correspondant à ses pays dépassent pas 0.388.

Figure III.6 : Classification des 118 pays selon la variable IDH.

1 Finlande France Etats-Unis Danemark Espagne Royaume Uni Autriche Belgique Luxembourg N.Zélande	9 Allemagne Italie Israël	17 Portugal, R. Tchèque Koweït Chypre	25 Libye	33 Colombie Kazakhstan Venezuela	41	49	57 Algérie VietNam
2 Islande Norvège Australie, Canada Irlande Suède Japon Pays Bas Suisse	10 Grèce	18 Bahamas Cuba	26 Arabie Saoudite Malaisie Oman Roumanie	34	42 Paraguay	50 Indonésie	58
3	11 Argentine Bahreïn Qatar Hongrie Pologne Emirate arabes unis Malte Chili	19	27 Ukraine	35 Chine Turquie Arménie	43	51 Guatemala	59 Ouzbékistan Bolivie Mongolie
4	12 Uruguay	20 Brésil	28 Tunisie	36	44 Nicaragua Égypte	52 Inde	60
5 Mexique Bulgarie	13 Maurice	21 Iran	29	37	45	53 Népal Papouasie Nouvelle Soudan Cameroun Haïti	61 Yémen Togo
6 Albanie	14 Pérou Jordanie Liban Philippines	22	30 Maroc Namibie Botswana	38	46 Ghana Bangladesh Mauritanie Pakistan	54 Ouganda Sénégal	62 Nigeria Tanzanie
7 Thaïlande	15	23 Gabon	31 Cam bodge	39 Djibouti Timor Leste Kenya Zimbabwe	47	55 Cote d'Ivoire Bénin Zambie	63 Ethiopie
8 Maldives Jamaïque	16 République arabe syrienne	24 Tadjikistan, Afrique du Sud	32	40 Érythrée	48 Rwanda Angola Guinée	56 Congo	64 Tchad R.C.Africaine Niger Mali B.Faso Sierra Leone Mozambique

Le nombre de classes obtenues peut être réduit en regroupant ces dernières au moyen de la classification hiérarchique classique sur les vecteurs codes finaux. Les groupes issus de cette deuxième classification sont appelés « macro classes ».

Les résultats de cette classification sont affichés dans la grille suivante :

Figure III.7 : Macro classes selon la classification hiérarchique

1 Finlande France Etats-Unis Danemark Espagne Royaume Uni Autriche Belgique Luxembourg N.Zélande	9 Allemagne Italie Israël	17 Portugal, R. Tchèque Koweït Chypre	25 Libye	33 Colombie Kazakhstan Venezuela	41	49	57 Algérie VietNam
2 Islande Norvège Australie, Canada Irlande Suède Japon Pays Bas Suisse	10 Grèce	18 Bahamas Cuba	26 Arabie Saoudite Malaisie Oman Roumanie	34	42 Paraguay	50 Indonésie	58
3	11 Argentine Bahreïn Qatar Hongrie Pologne Emirate arabes unis Malte Chili	19	27 Ukraine	35 Chine Turquie Arménie	43	51 Guatemala	59 Ouzbékistan Bolivie Mongolie
4	12 Uruguay	20 Brésil	28 Tunisie	36	44 Nicaragua Égypte	52 Inde	60
5 Mexique Bulgarie	13 Maurice	21 Iran	29	37	45	53 Népal Papouasie Nouvelle Soudan Cameroun Haïti	61 Yémen Togo
6 Albanie	14 Pérou Jordanie Liban Philippines	22	30 Maroc Namibie Botswana	38	46 Ghana Bangladesh Mauritanie Pakistan	54 Ouganda Sénégal	62 Nigeria Tanzanie
7 Thaïlande	15	23 Gabon	31 Cam bodge	39 Djibouti Timor Leste Kenya Zimbabwe	47	55 Cote d'Ivoire Bénin Zambie	63 Ethiopie
8 Maldives Jamaïque	16 République arabe syrienne	24 Tadjikistan, Afrique du Sud	32	40 Érythrée	48 Rwanda Angola Guinée	56 Congo	64 Tchad R.C.Africaine Niger Mali B.Faso Sierra Leone Mozambique

Macro classe 1
 Macro classe2
 Macro classe3
 Macro classe 4

Quatre macro classes ont été créés :

- la première se trouvant en haut à gauche comprend les pays riches à développement humain élevé (IDH supérieur à 0.863);
- la deuxième regroupant les pays situés en diagonale englobe les pays en voie de développement à développement humain moyen (IDH supérieur à 0.71 et inférieur à 0.87 ;

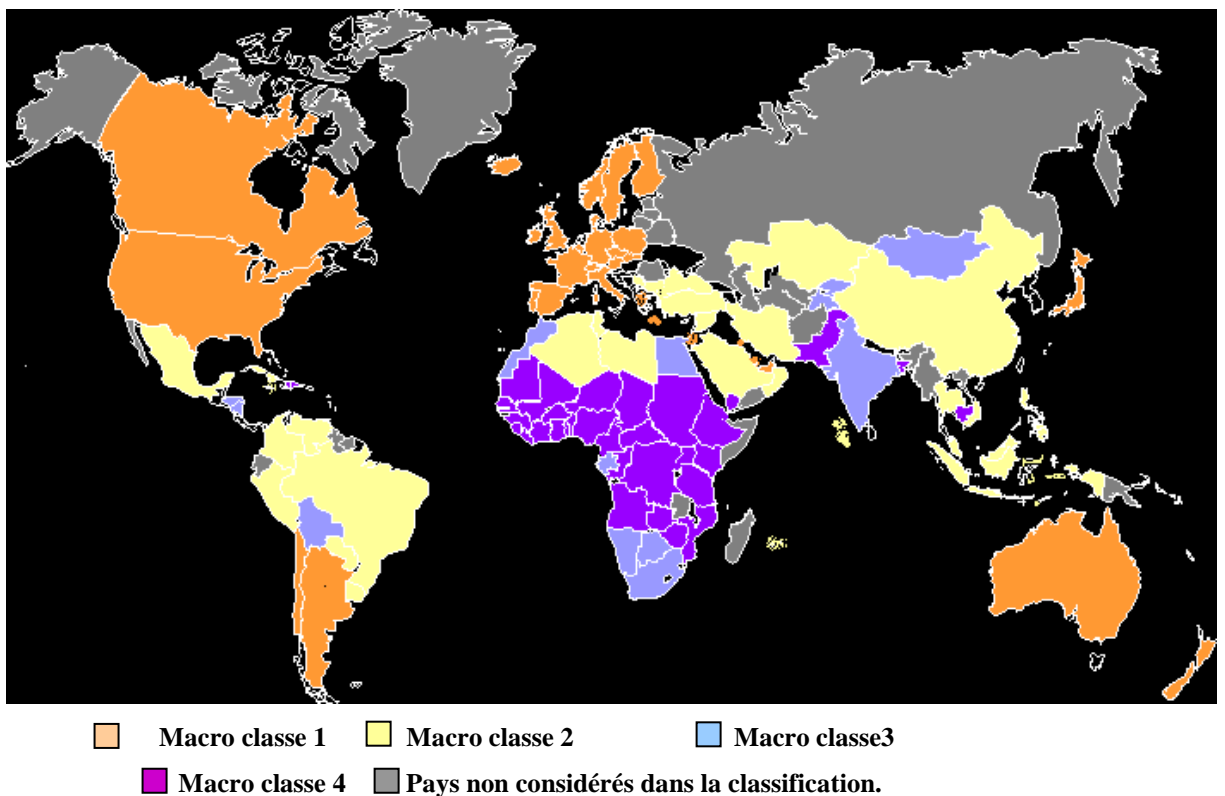
- la troisième macro classe colorée en bleu, regroupe les pays sous développés à développement humain faible (IDH compris entre 0.58 et 0.71) ;
- et enfin la dernière macro classe, en bas et à droite rassemble les pays très pauvres dont le niveau d'IDH est très faible (inférieur à 0.55)..

La valeur moyenne de l'IDH de chaque macro classe est présentée dans le tableau suivant :

Tableau 3 : valeur moyenne de l'IDH dans les macro classes

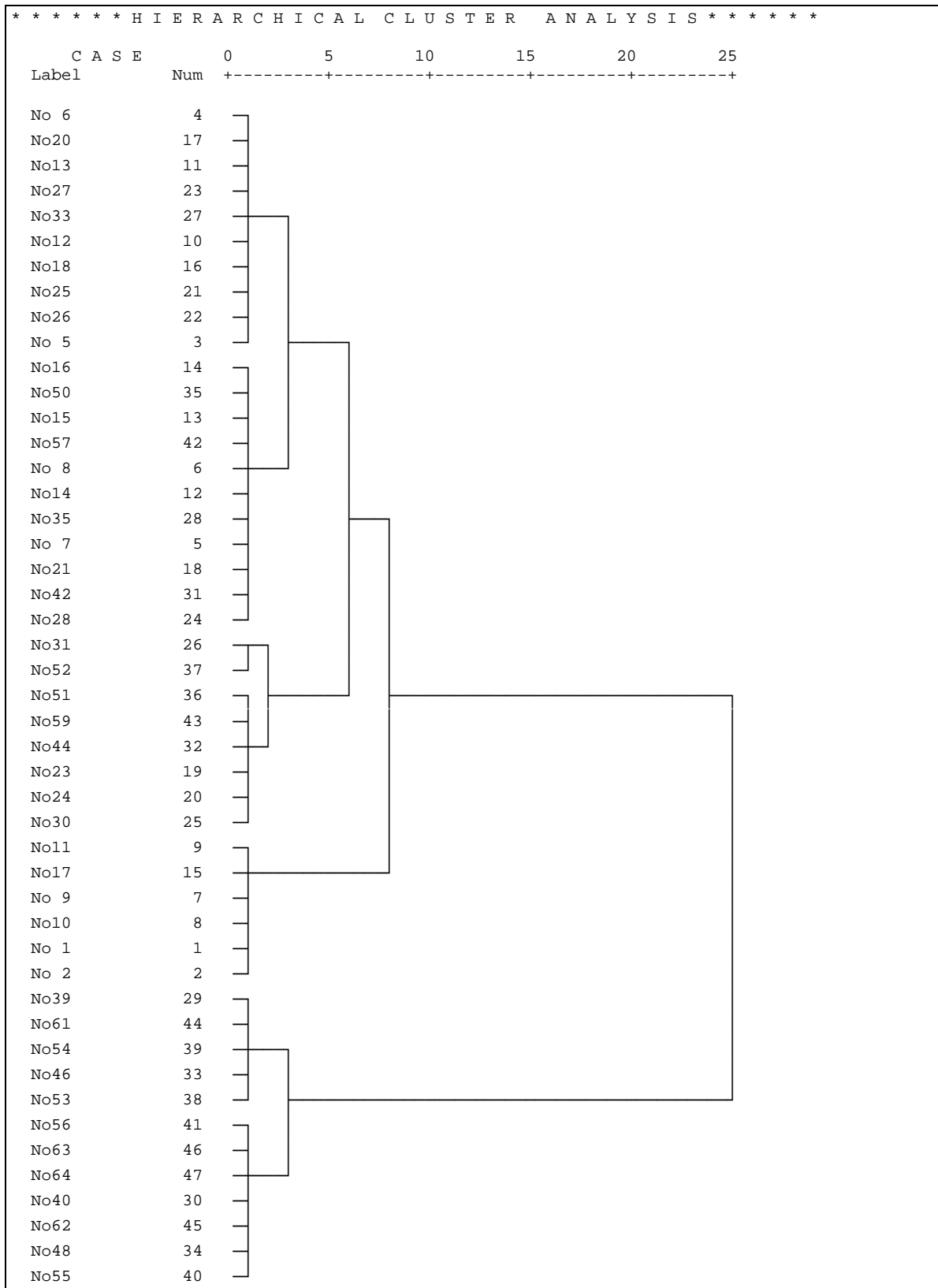
Numéro de la macro classe	1	2	3	4
Valeur moyenne de l'IDH	0.9263	0.7850	0.6367	0.5292

Figure III.8: Représentation des 118 pays classifiés sur la carte du monde.



La représentation des résultats de la classification hiérarchique sur la carte géographique du monde, montre la proximité de plusieurs classes sur les territoires. Ainsi la majorité des pays pauvres et sous développés appartient à l'Afrique noire alors que les pays riches se localisent en Amérique du nord, à l'Europe de l'ouest et en Australie.

Figure III.9 : Dendrogramme de la classification hiérarchique sur les vecteurs code de l'application IDH



3.2 Classification en deux dimensions : PIB / espérance de vie à la naissance :

En utilisant les deux variables PIB et l'espérance de vie à la naissance, la classification permet de discerner 49 classes. Les résultats de cette classification se trouvent dans la grille qui suit.

Les groupes obtenus ne sont pas très différents dans l'ensemble de ceux de la première classification car les variables choisies sont des composantes de l'indice de développement humain.

La partie du haut à gauche englobe des groupes de pays riches. Par exemple, le groupe 1 comprend l'Islande et la Suisse. Les valeurs d'espérance de vie et de PIB de ces deux pays sont parmi les plus élevées (81.5 et 81.ans pour l'espérance de vie et 36510 et 35633\$ respectivement).

La partie opposée, en bas à droite, réunit les pays très pauvres comme la Zambie, le Mozambique, le Zimbabwe, Sierra Leone et l'Angola (groupe 64). Ces pays pauvres ont une faible espérance de vie et un niveau de PIB très bas (respectivement ans et 1488.8\$ en moyenne).

La partie centrale de la carte contient les pays dont le niveau du PIB est moyen et celui de l'espérance de vie est élevé. Le groupe 6 par exemple, comprend Cuba, l'Albanie et Venezuela. Ces pays ont des niveaux de PIB de 6 000, 5 316 et 6 632 \$ respectivement et des valeurs d'espérance de vie de 77.7, 76.2 et 73.2 ans respectivement.

Figure III.10 : Résultats de la classification des 118 pays selon les deux variables PIB et Espérance de vie.

1 Islande Suisse	9 Irlande Etats-Unis Norvège	17 Luxembourg	25	33 Portugal, R. Tchèque	41 Oman	49 Argentine	57 Pologne
2 Australie, Suède Japon France	10 Finlande, Pays Bas Belgique Canada	18 Danemark Royaume Uni Autriche	26 Quatar	34 Bahrein	42 Bahamas, Hongrie	50 Arabie Saoudite	58 Maurice
3 Espagne Italie	11 Allemagne	19 N.Zélande Israel	27 Koweit Emirate arabes unis	35	43	51	59
4 Grèce Chypre Malte	12	20	28	36	44 Gabon	52	60 Afrique du Sud Botswana
5 Libye Malaisie Mexique Uruguay Chili	13 Roumanie Brésil, Bulgarie Turquie	21 Iran Thaïlande	29 Ukraine	37 Kazakhstan	45	53	61 Namibie
6 Albanie Cuba Venezuela	14 Tunisie Algérie Colombie Chine	22	30 Maldives	38	46 Guinée	54 Tanzanie Ethiopie Kenya B.Faso	62 Ouganda
7 Arménie Jordanie, Liban	15 Égypte Guatemala Indonésie Maroc Pérou Philippines	23	31 Inde , Bolivie Mongolie Pakistan	39 Cam bodge, Ghana Togo	47 Djibouti Mali	55 Cameroun Tchad	63 R.C.Africaine
8 Jamaïque Nicaragua VietNam République arabe syrienne	16 Paraguay	24 Ouzbékistan Tadjikistan,	32 Népal, Bangladesh Yémen Mauritanie Sénégal	40 Haïti Timor Leste	48 Papouasie Nouvelle Soudan Érythrée Bénin Niger	56 Nigeria Rwanda Cote d'Ivoire Congo	64 Angola Sierra Leone, Zambie Mozambique Zimbabwe

L'application de la classification hiérarchique sur les vecteurs codes des groupes créés permet de distinguer entre 5 macro classes qui ressemblent un peu à celles de la dernière classification CAH de l'IDH. La différence réside dans la macro classe comprenant les pays riches qui s'est divisée en 2 macro classes laissant apparaître une nouvelle macro classe regroupant les pays dont le niveau du PIB est très élevé. Les résultats de cette classification sont affichés dans la carte suivante :

Figure III.11 : Macro classes selon la classification hiérarchique

1 Islande Suisse	9 Irlande Etats-Unis Norvège	17 Luxembourg	25	33 Portugal, R. Tchèque	41 Oman	49 Argentine	57 Pologne
2 Australie, Suède Japon France	10 Finlande, Pays Bas Belgique Canada	18 Danemark Royaume Uni Autriche	26 Quatar	34 Bahreïn	42 Bahamas, Hongrie	50 Arabie Saoudite	58 Maurice
3 Espagne Italie	11 Allemagne	19 N.Zélande Israel	27 Koweït Emirate arabes unis	35	43	51	59
4 Grèce Chypre Malte	12	20	28	36	44 Gabon	52	60 Afrique du Sud Botswana
5 Libye Malaisie Mexique Uruguay Chili	13 Roumanie Brésil, Bulgarie Turquie	21 Iran Thaïlande	29 Ukraine	37 Kazakhstan	45	53	61 Namibie
6 Albanie Cuba Venezuela	14 Tunisie Algérie Colombie Chine	22	30 Maldives	38	46 Guinée	54 Tanzanie Ethiopie Kenya B.Faso	62 Ouganda
7 Arménie Jordanie, Liban	15 Égypte Guatemala Indonésie Maroc Pérou Philippines	23	31 Inde , Bolivie Mongolie Pakistan	39 Cam bodge, Ghana Togo	47 Djibouti Mali	55 Cameroun Tchad	63 R.C.Africaine
8 Jamaïque Nicaragua VietNam République arabe syrienne	16 Paraguay	24 Ouzbékistan Tadjikistan,	32 Népal, Bangladesh Yémen Mauritanie Sénégal	40 Haïti Timor Leste	48 Papouasie Nouvelle Soudan Érythrée Bénin Niger	56 Nigeria Rwanda Cote d'Ivoire Congo	64 Angola Sierra Leone, Zambie Mozambique Zimbabwe

■ Macro classe 1 ■ Macro classe 2 ■ Macro classe3
■ Macro classe4 ■ Macro classe 5

Les valeurs moyennes de l'espérance de vie à la naissance et du PIB des macro classes sont résumées dans le tableau suivant :

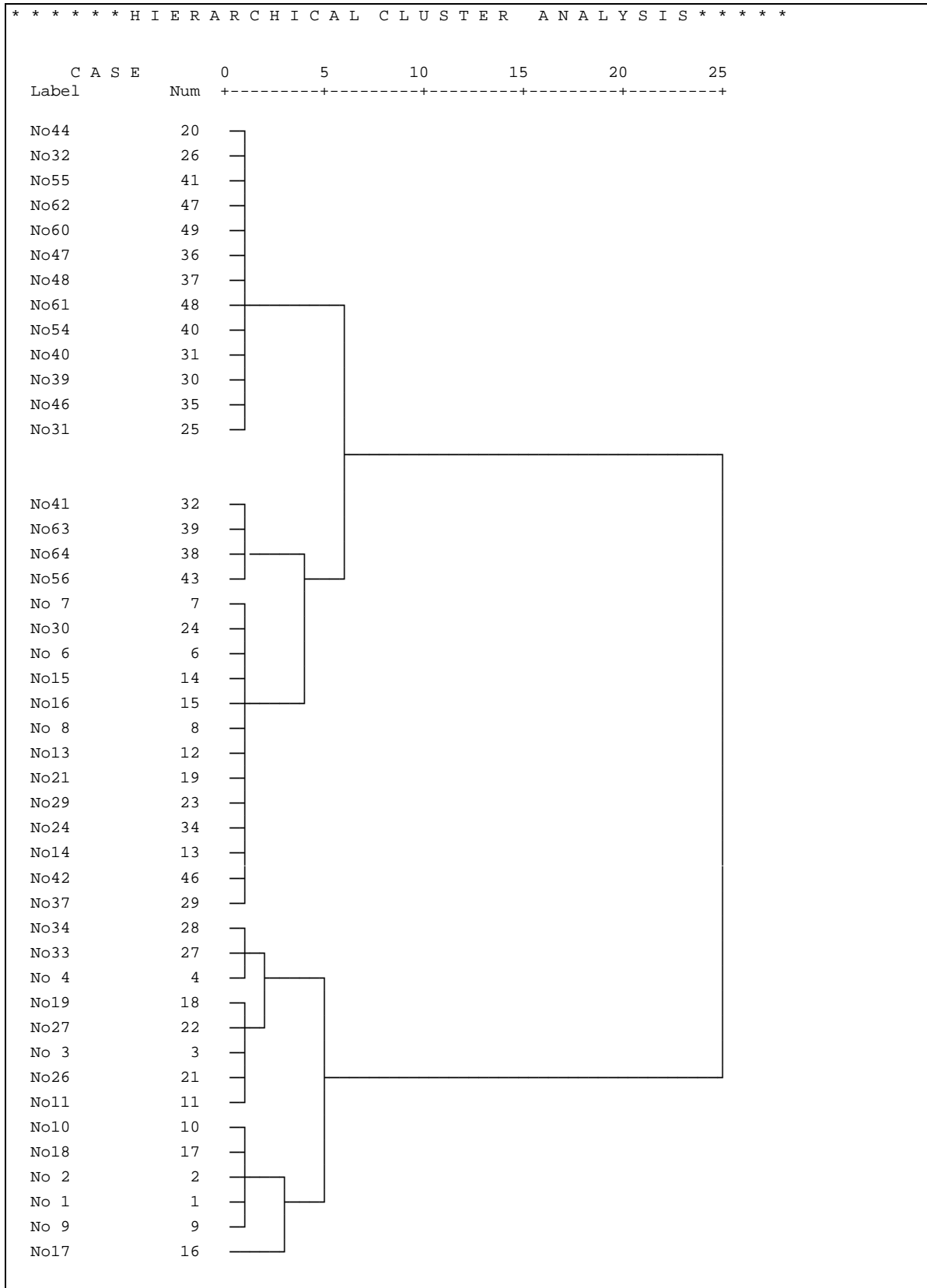
Tableau 4 : Valeurs moyennes de l'espérance de vie à la naissance et du PIB des macro classes

Numéro de la macro classe	1	2	3	4	5
Valeur moyenne De l'espérance de vie	0,80	0,78	0,72	0,69	0,43
Valeur moyenne Du PIB EN \$	35 964,7	24 515,5	7 926,72727	11910,6	1313,7

Les valeurs moyennes de l'espérance de vie à la naissance ne sont pas très différentes en passant de la première macro classe à la deuxième, et de la troisième à la quatrième. Mais l'écart entre l'espérance de vie de la quatrième macro classe et la troisième est assez grand.

Contrairement à l'espérance de vie, le changement dans les valeurs du PIB entre les différentes macro classes est grand. Pour la 1^{ère} macro classe composée de pays très riches, la valeur du PIB est de 35 964,7 \$ lors que pour les pays pauvre le PIB moyen est de 1313,7\$.

Figure III.12 : Dendrogramme de la classification hiérarchique sur les vecteurs code de l'application PIB / espérance de vie à la naissance.

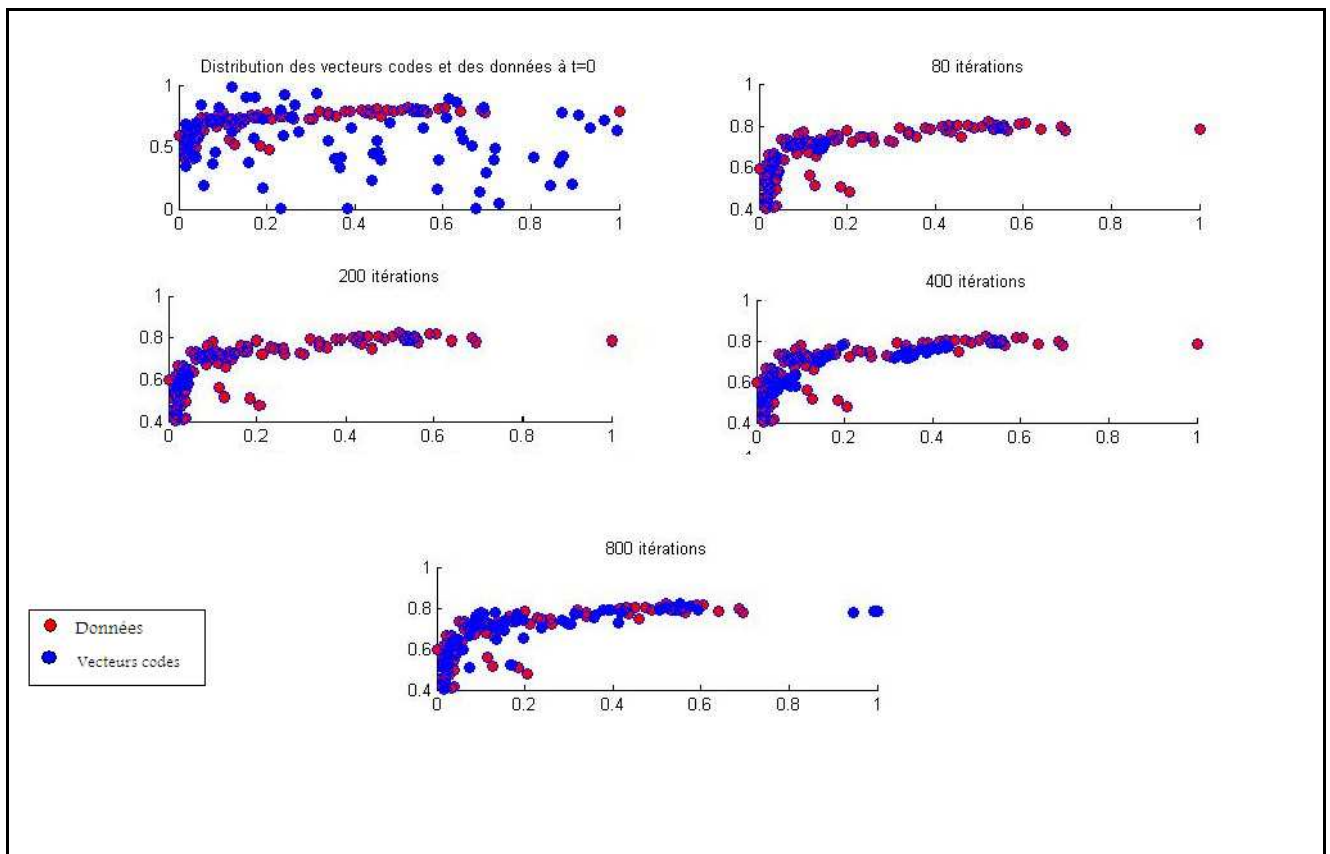


- Déploiement de la carte de Kohonen :

Comme les données utilisées sont de deux dimensions, la présentation du déploiement de la carte de Kohonen se fait facilement sur un plan. La variable espérance de vie est présentée en ordonnée et la variable PIB en abscisse.

Les figures suivantes montrent l'évolution des poids des vecteurs code présentés simultanément avec les valeurs des deux variables :

Figure III.13 : Déploiement de la carte à travers les différentes itérations.

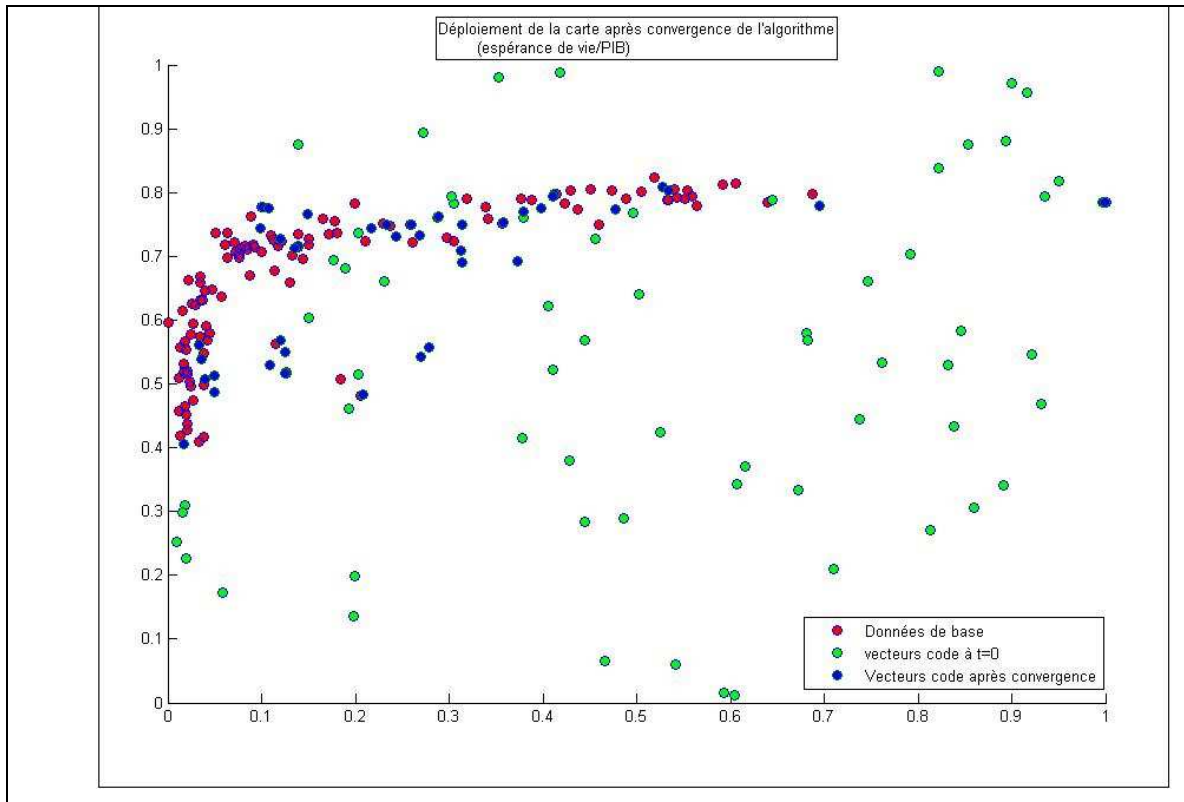


Le premier graphe à gauche montre la répartition des données de base (points rouges) et des vecteurs codes initiaux (points bleus). Comme les vecteurs codes sont choisis aléatoirement, ils sont distribués sur toutes les cartes. Après 80 itérations (graphe en haut à droite), ils ne sont plus aussi dispersés, et ils commencent à prendre la forme des données. Ainsi, plus l'algorithme est réitéré, plus les vecteurs codes se rapprochent de la base des données.

La dernière figure (en bas) montre les vecteurs codes après convergence de l'algorithme. Chaque vecteur code représente une région de la base des données.

La figure suivante montre une représentation simultanée des vecteurs codes initiaux, des données et des vecteurs codes finaux.

Figure III.14 : Représentation simultanée des données et des vecteurs codes après convergence



4. Classification des 48 wilayas de l'Algérie :

Pour cette classification, les variables considérées sont : le taux d'analphabétisme féminin, le pourcentage de chefs de ménages sans instruction, la part chefs de ménages actifs et le pourcentage de ménages propriétaires de leur logement de chaque wilaya.

La carte utilisée comporte 25 neurones (5 x 5).

Les résultats de la classification sont présentés ci-dessous :

Figure III.15 : Classification des 48 wilayas d'Algérie

1 ADRAR	6 OUARGLA EL OUED GHARDAIA	11 BATNA NAAMA	16 TLEMCEM	21 BEJAIA TIZI OUZOU
2 TEBESSA M SILA KHENCHELA EL BAYADH	7 MASCARA	12 BISKRA BOUIRA MILA	17 SETIF B.B.ARRERIDJ	22 JJJEL
3 CHLEF TISSEMSILT RELIZANE	8 MOSTAGANEM	13 SOUK AHRAS	18 EL TARF BOUMERDES	23 BECHAR SIDI BEL ABBES AINTIMOUCHENT
4 AIN DEFLA	9 LAGHOUAT MEDEA	14 OUM EL BOUAGHI SAIDA TINDOUF	19 SKIKDA	24
5 TAMANRASSET ILLIZI DJELFA	10 TIARET	15 GUELMA TIPAZA	20 BLIDA	25 GVT GRAND ALGER CONSTANTINE ANNABA ORAN

La classification a permis de créer 24 groupes dont plusieurs paraissent évidents. A titre d'exemple, le groupe 25 comprend le grand gouvernorat d'Alger, Annaba, Oran et Constantine. Ces wilayas représentent les grandes métropoles.

Les valeurs des différentes variables utilisées de ces villes sont proches : un taux d'analphabétisme tournant autour de 25% (25, 28, 26 et 24% respectivement), un pourcentage de chefs de ménages sans instruction de 36.2% en moyenne et une proportion de chefs de ménages actifs est estimée à 68.3%. Un peu moins du tiers des ménages vivants dans ces villes est propriétaire de son logement (30.2% en moyenne).

Le groupe 17 englobe les wilayas de Sétif et de B B..Arreridj. Les valeurs des variables dans ce groupe sont aussi très proches : respectivement 43 et 42% pour le taux d'analphabétisme féminin, 51.8 et 49.9% pour le pourcentage de chefs de ménages sans instruction, 69.4 et 70.6% pour la proportion de chefs de ménages actifs et enfin, 54.8 et 50.6% pour le pourcentage de ménages possédant leur logement.

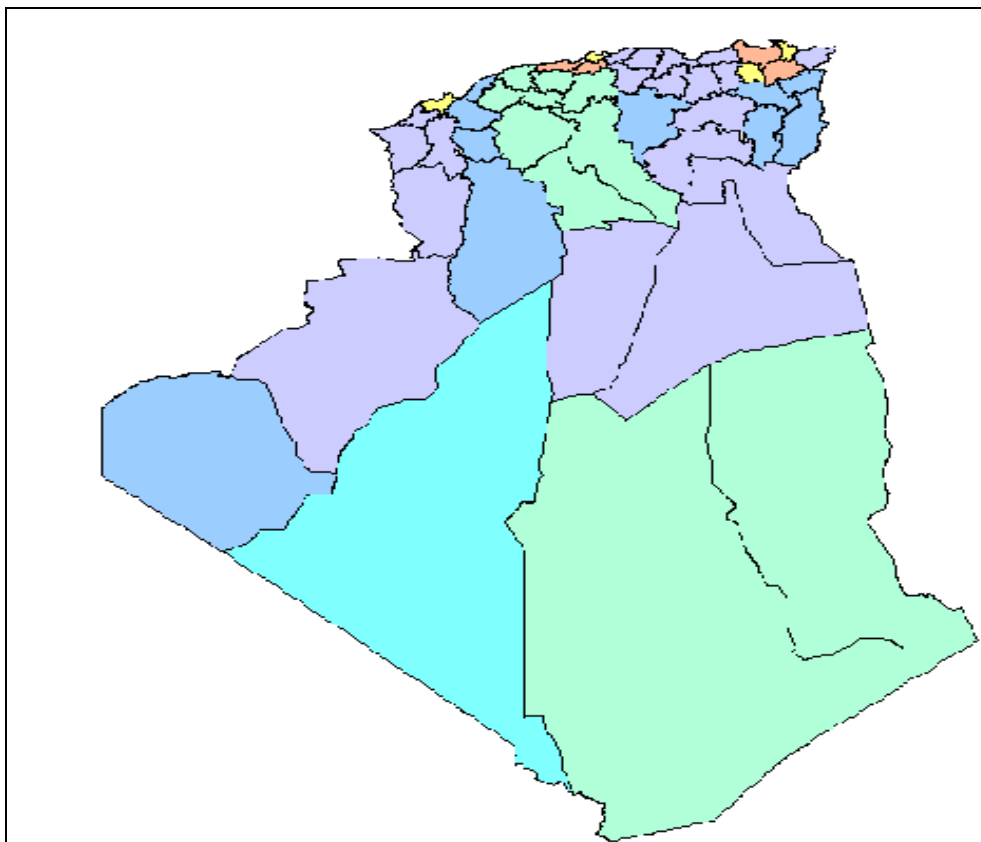
Une classification hiérarchique sur ces données permet de mieux voir les blocs de wilayas se ressemblant. L'application de cette classification a permis de discerner 6 macro classes. Ces macro classes sont affichés sur la grille ci dessous ainsi que sur la carte d'Algérie qui suit.

Figure III.16: Classification hiérarchique des wilayas d'Algérie

1 ADRAR	6 OUARGLA EL OUED GHARDAIA	11 BATNA NAAMA	16 TLEMCEM	21 BEJAIA TIZI OUZOU
2 TEBESSA M SILA KHENCHELA EL BAYADH	7 MASCARA	12 BISKRA BOUIRA MILA	17 SETIF B.B.ARRERIDJ	22 JIJEL
3 CHLEF TISSEMSILT RELIZANE	8 MOSTAGANEM	13 SOUK AHRAS	18 EL TARF BOUMERDES	23 BECHAR SIDI BEL ABBES AINTIMOUCHENT
4 AIN DEFLA	9 LAGHOUAT MEDEA	14 OUM EL BOUAGHI SAIDA TINDOUF	19 SKIKDA	24
5 TAMANRASSET ILLIZI DJELFA	10 TIARET	15 GUELMA TIPAZA	20 BLIDA	25 GVT GRAND ALGER CONSTANTINE ANNABA

Macro classe 1
 Macro classe 2
 Macro classe 3
 Macro classe 4
 Macro classe 5
 Macro classe 6

Figure III.17 : Classification des wilayas d'Algérie en 6 classes.



Les grandes métropoles (GV d'Alger, Constantine, Oran et Annaba) forment à elles seules une macro classe (macro classe 1). Cette macro classe possède en moyenne les plus faibles taux d'analphabétisme féminin, de chefs de ménages sans instruction et de ménages possédant leur logement.

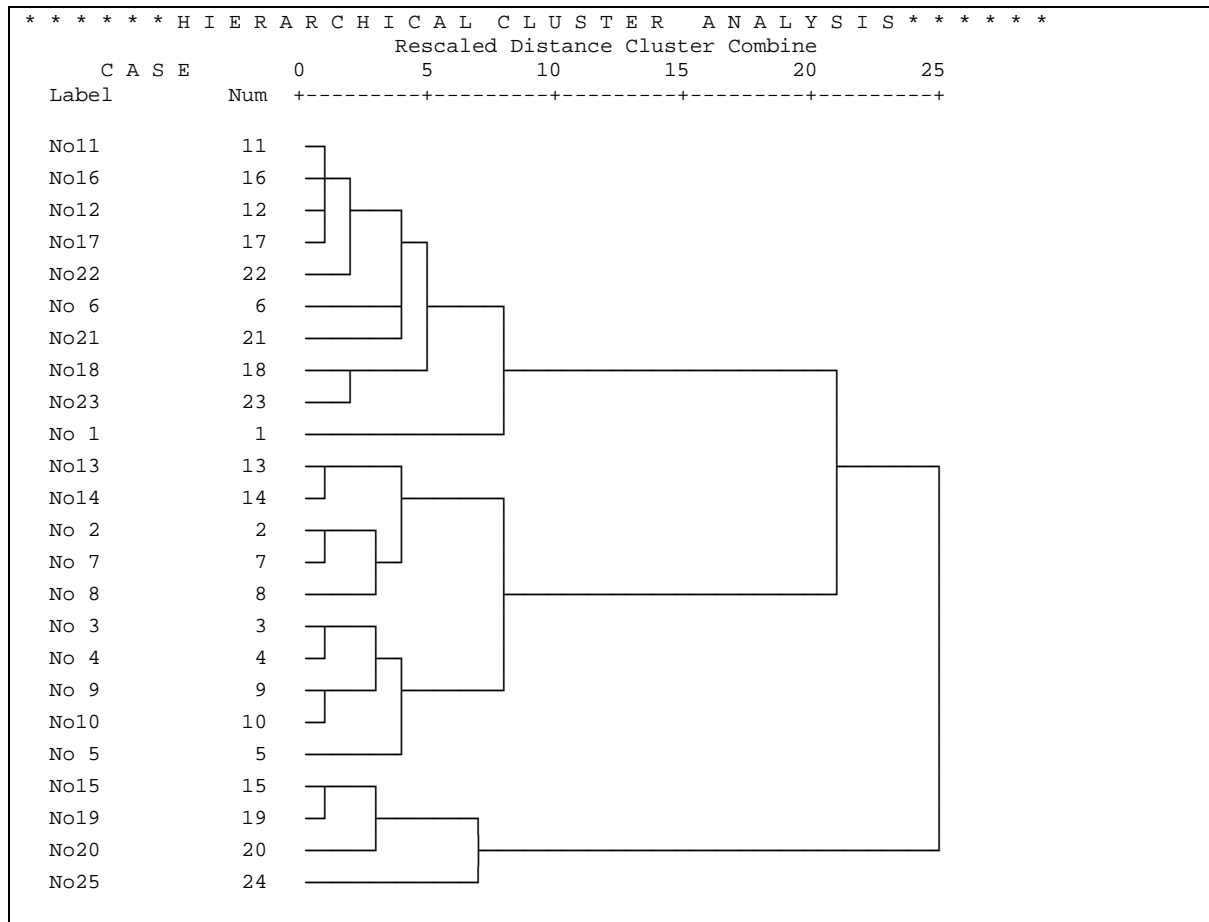
Si on ne considère pas la wilaya d'Adrar, qui forme un singleton, les valeurs des trois variables : taux d'analphabétisme féminin, pourcentage de chefs de ménages sans instruction et actifs et la proportion de ménages possédant leur logement ; sur les différentes macro classes, augmentent selon la diagonale de bas en haut.

Les statistiques descriptives des différentes macro classes sont présentées dans le tableau suivant :

Tableau 5 : valeurs moyennes des différentes variables par macro classe

Numéro de la Macro classe	Analphabétisme féminin	% CM sans instruction	% CM actifs	% Propriétaire du logement
1	25,7	36,2	68,3	30,2
2	36,3	39,7	70,4	32,5
3	53,2	44,7	70,9	35,7
4	46,8	47,7	71,4	38,8
5	50,4	50,5	72,4	39,1
6	38,9	54,6	71,5	42,1

Figure III.18 : Dendrogramme de la classification hiérarchique sur les vecteurs code de l'application les 48 wilayas d'Algérie.



CONCLUSION

Conclusion

Nous avons tenté, dans ce mémoire, de mettre en exergue les principales caractéristiques des cartes de Kohonen. En effet, nous avons vu qu'elles permettaient la projection de grandes bases de données dans un espace de dimension réduite et la visualisation directe des relations présentes entre les données analysées. Leur application est facile et l'apprentissage est rapide. Le réseau détermine automatiquement les groupes auxquels appartiennent les données. Ces derniers (groupes) sont facilement discernables grâce à une représentation graphique simple.

Ces cartes permettent aussi de traiter des problèmes difficilement abordables par des méthodes linéaires ou par des équations algébriques.

Ces propriétés intéressantes font que leur application peut s'opérer dans divers domaines. Ainsi, en médecine, les cartes SOM ont été utilisées pour classer des tumeurs cérébrales. En économie, la prédiction de séries temporelles financières se fait par double carte de Kohonen. Elles peuvent également être très utiles pour la reconnaissance de formes et dans les recherches sur web.

Dans ce mémoire des applications des cartes de Kohonen ont été effectuées. Elles concernent la classification de 118 pays et la classification des 48 wilayas d'Algérie selon des indicateurs socio économiques.

Pour la classification des 118 pays, dans un premier temps une classification d'une dimension a été réalisée en considérant la variable IDH. Dans un deuxième temps, deux variables ont été retenues pour faire une classification à 2 dimensions. Il s'agit du PIB/habitant et de l'espérance de vie à la naissance.

La première classification a permis de distinguer 47 classes, regroupées en 04 macro classes. Ces classes sont homogènes et regroupent des pays dont le niveau d'IDH est proche. Ainsi les pays sont distribués selon le niveau de leur développement humain: niveau élevé, niveau moyen, faible ou très faible. Quant à la deuxième classification, elle a révélé 49 classes et a permis de générer cinq macro classes. Ces résultats semblent intéressants dans la mesure où le PNUD n'utilise que trois grands groupes de pays selon la valeur de l'IDH. Une classe supplémentaire peut fournir une classification plus fine et plus homogène.

La classification des 48 wilayas d'Algérie a permis de distinguer 24 groupes de wilayas qui ont été regroupés en 6 macro classes. Le taux d'analphabétisme féminin est corrélé positivement avec le pourcentage de chefs de ménages sans instruction. Les plus faibles

valeurs de ces deux variables sont enregistrées dans la macro classe regroupant les grandes métropoles.

L'application des cartes faite a permis de comprendre le déroulement de l'algorithme et comment la carte s'auto organisait. Cependant, les bases de données utilisées restent tout de même assez simples par rapport aux possibilités de l'algorithme dont nous n'avons utilisé que l'algorithme standard de Kohonen. Il serait donc intéressant de pouvoir utiliser dans le futur, des bases de données nettement plus grandes et d'essayer de voir comment modifier l'algorithme originel pour l'appliquer à d'autres domaines de la statistique. Par exemple l'ajout de la variable temporelle peut être un élément d'étude. En tout cas, si des perspectives pouvaient s'ouvrir pour une plus grande appropriation de ces méthodes, même à petite échelle, ce travail aura été utile.

Pour terminer, il est important de signaler les difficultés rencontrées lors de la conception du logiciel régissant l'algorithme de Kohonen. L'absence d'une documentation technique suffisante fait que son élaboration n'a pas été une tâche facile d'autant plus que les détails sur les différents paramètres de l'algorithme ne sont jamais exposés dans la documentation disponible.

BIBLIOGRAPHIE

Bibliographie

- [1]. Aaron Catherine, Galanti Sébastien, Tadjeddine Yamina. Article : Les styles des gérants de SICAV actions en France 1999-2002. Une approche dynamique à partir des cartes de Kohonen.
- [2]. Akarçay-Gurbuz and Perraudin. Article : Classification des pays européens 2004.
- [3]. Andronova Vania. Projet de fin d'étude Master 2nd année : Utilisation de données météo et des réseaux de neurones pour la prédiction de vitesses de vent. Juillet 2006.
- [4]. Bélanger Hugues. Maîtrise : Réseau de Kohonen pour la détection des contours d'objets dans une image à niveaux de gris. 1998.
- [5]. Charles Bouveyron. Modélisation et classification des données de grande dimension, application à l'analyse d'images. Septembre 2006.
- [6]. Clergue Manuel. Article : Réseaux de Neurones Artificiels.
- [7]. Cottrell Marie, Fort Jean Claude. Article : Etude d'un processus d'auto organisation. Université de Paris 1987.
- [8]. Cottrell Marie, Ibbou Smaïl, Letrémy Patrick, Rousset Patrick. Article : « Cartes auto-organisées pour l'analyse exploratoire de données et la visualisation » Journal de la Société Française de Statistique, 2003.
- [9]. Cottrel Marie, Letrémy Patrick. Article: How to use the Kohonen algorithm to simultaneously analyze individuals and modalities in a survey Université de Paris I, SAMOS. Août 2004.
- [10]. Cottrell Marie, Ibbou Smaïl, Letrémy Patrick. Article : Traitement des données manquantes au moyen de l'algorithme de Kohonen SAMOS. 2003.
- [11]. Cziho András. Thèse de Doctorat : Quantification vectorielle et compression d'image. Application à l'imagerie médicale. Université de Rennes1. mai 1999.
- [12]. Dablemont Simon, Lendasse Amaury, Article : Prédiction de séries temporelles financières par double carte de Kohonen et modèles RBFNS. Application à la prédiction de l'indice boursier DAX30.
- [13]. Delorme Fabien. Thèse de Doctorat : Évaluation et modélisation automatiques des connaissances des apprenants à l'aide de cartes conceptuelles Institut National des Sciences Appliquées de Rouen. Novembre 2005.
- [14]. Dreyfus Gérard. Reconnaissance de formes par réseaux de neurones. Juillet 2002.
- [15]. Framling Kary. Mémoire de DEA : Les réseaux de neurones comme outils d'aide à la décision floue. Juillet 1992.

- [16]. Fritzke, B. article: *A growing neural gas network learns topologies*, Advances in Neural Information Processing Systems. Press, Cambridge MA. 1995.
- [17]. Gaudier Fabrice. Thèse de Doctorat : Modélisation par réseaux de neurones : Application à la gestion du combustible dans un réacteur. Janvier 1999.
- [18]. Goodhill G J, Finch S et Sejnowski T J. Rapport: Quantifying neighborhood reservation in topographic mapping, 1995.
- [19]. Gosselin Bernard Faculté Polytechnique de Mons. Article : Application de réseaux de neurones artificiels à la reconnaissance automatique de caractères manuscrits 1996.
- [20]. HO Tuong Vinh et PHAM Quang Dung. Mémoire : Réseaux de neurones pour la reconnaissance des formes. Hanoi. Juillet 2005.
- [21]. Kohonen Teuvo, Livre : "Self-Organizing Maps", 3rd Edition, Springer ,2001.
- [22]. Kohonen, T. Article: Self Organizing Maps, volume 30 of Springer series of information sciences. Springer Verlag, 1995.
- [23]. Koikkalainen, P. *Progress with the tree-structured self-organizing map*, in Cohn, A. G., editor, Proceedings of ECAI'94, 11th European Conference on Artificial Intelligence. John Wiley & Sons. 1994.
- [24]. Ladjadj Rachid. Article : Les réseaux de neurones. Informatique et réseaux. 2003.
- [25]. Lambert Tanguy, Xavier Lucas, Manier Xavier. Projet de fin d'études : Méthodes de classification et comparaison de partitions. Janvier 2006.
- [26]. Laviale Martin. Mémoire de DEA : Utilisation des réseaux de neurones dans la détermination de la structure des communautés phytoplanctoniques. 2004.
- [27]. Lebart Ludoviq. Mémoire de DEA : Quelques méthodes de classification. .2004.
- [28]. Lebart Ludoviq, Morineau Alain, Piron Marie. Livre : Statistique exploratoire multidimensionnelle, Dunod, 1995.
- [29]. Lesot Marie-Jeanne, d'Alché-Buc Florence, Siolas Georges. Article : Evaluation des cartes auto-organisatrices et de leur variante à noyaux. Laboratoire d'Informatique de Paris. Université Pierre et Marie Curie.
- [30]. Letrémy Patrick. Article : Traitements de données qualitatives par des algorithmes fondés sur l'algorithme de Kohonen Université de Paris1. SAMOS-MATISSE.
- [31]. Mé Ludovic, Alanou Véronique et Abraham Supélec Jorg. Pré-étude : Utilisation de cartes de Kohonen pour détecter des intrusions dans un système informatique.
- [32]. Miikkulainen, R. Bednar, J. A. Choe, Y. et Sirosh, J. *Self-Organization, Plasticity, and Low-level Visual Phenomena in a Laterally Connected Map Model of the Primary Visual Cortex*, in Goldstone. Psychology of Learning and Motivation. Academic Press, San Diego CA. 1997

- [33]. Moutarde Fabien. Communication : Ecole des Mines de Paris. Novembre 2005.
- [34]. Nakache J.P., Confais J. Livre : Approche pragmatique de la classification, arbres hiérarchiques, partitionnement. Edition Technip.
- [35]. *Pilato Vincent*. Rapport scientifique et technique : *Application des réseaux de neurones aux méthodes de mesure basées sur l'interaction rayonnement matière*. Direction des technologies avancées. Département des applications et de la météorologie des rayonnements ionisants. 1999.
- [36]. Polani D. et Uthmann T. Article: *Training kohonen feature maps in different topologies: an analysis using genetic algorithms*, In Proceedings of the 5th International Conference on Genetic Algorithms. 1993.
- [37]. Robaye H. Article : Investigation et amélioration de méthodes d'approximation par réseaux de fonctions à base radiale généralisées. 2006.
- [38]. Rousset Patrick. Thèse de Doctorat : Applications des algorithmes d'auto-organisation à la classification et à la prévision. Université Paris1. Département de mathématiques et d'informatique 2000.
- [39]. Saalbach, A. Twellmann, T. Wismüller, A. Ontrup, J. Ritter, H. et Nattkemper, T.W. Article: *A Hyperbolic Topographic Mapping for Proximity Data*, Proc. of the IASTED International Conference on Artificial Intelligence and Applications. 2005.
- [40]. Sapota. G, *Probabilités, Analyse des Données et Statistique*, Technip, 1990
- [41]. Touzet Claude. Thèse de Doctorat : Les réseaux de neurones artificiels : introduction au connexionisme. Juillet 1992.
- [42]. Turenne Nicolas. Thèse de Doctorat : Apprentissage statistique pour l'extraction de concepts à partir de textes. Application au filtrage d'informations textuelles.. Novembre 2000.
- [43]. Valentin Nicolas. Thèse de Doctorat : Construction d'un capteur logiciel pour le contrôle automatique du procédé de coagulation en traitement d'eau potable. Décembre 2000.
- [44]. Vesanto Juha. Mémoire : Using SOM in Data Mining. Département d'informatique de l'Université des technologies de Helsinki. Finlande. Avril 2000.
- [45]. Villnamm T Der R, Herrmann M et Martinez T. Topology perservation in self-Organizing feature maps, IEEE Transaction on Neural Networks, 1997.
- [46]. Vladan M Babovic and Gopakumar R. Stork, Article: "Seabed Recognition Using Neural Networks", *D2K Technical Report* , March, 1999.
- [47]. Wira Patrice. Thèse de Doctorat : Réseaux neuroméitiques, modularité et statistiques : estimation du mouvement pour l'asservissement visuel de robots.. Université de Haute Alsace des Sciences et Techniques Janvier 2002

- [48]. Wynsa B. Boullarta L., Setteb S., Baetenc D., Hoffmanc I., De Keyserc F. Article: Prediction of arthritis using a modified Kohonen mapping and case based reasoning. Elsevier 2004.
- [49]. Zrehen S et Blayo F. Article: A geometric organization measure for Self-Organization Kohonen Maps. 1992.
- [50]. Article: Clustering With GIS: An Attempt to Classify Turkish District Data Ece AKSOY, Turkey Shaping the Change XXIII FIG Congress Munich, Germany, Octobre 2006.
- [51]. Apprentissage, Réseaux de Neurones et Applications. Université Lyon 2 Discipline: Informatique.
- [52]. Cours sur l'Apprentissage automatique : les réseaux de neurones : <http://www.grappa.univ-lille3.fr/polys/apprentissage/sortie005.html>. 2002.
- [53]. Site officiel du PNUD : www.dz.undp.org

ANNEXE

Table de Burt et tableau disjonctif complet

Table de Burt et tableau disjonctif complet

1) Tableau disjonctif complet :

On considère une base de données où toutes les observations sont décrites par K variables qualitatives, chacune ayant un certain nombre de modalités.

Soit un ensemble de N individus et K variables ou questions. Chaque question k ($1 \leq k \leq K$) possède m_k modalités. Les individus répondent à chaque question k en choisissant seulement une modalité parmi les m_k modalités.

On pose $M = \sum_{k=1}^K m_k$ le nombre total de modalités, chaque individu est représenté par un vecteur ligne de M composantes à valeurs dans $\{0, 1\}$. Il n'y a qu'un 1 parmi les m_1 premières composantes, seulement un 1 entre la $(m_1 + 1)^{\text{ième}}$ et la $(m_1+m_2)^{\text{ième}}$, etc.

Par simplification, on numérote toutes les modalités de 1 à M et on note Z_j ($1 \leq j \leq M$) le vecteur colonne constitué des N réponses de la $j^{\text{ième}}$ modalité. Le $i^{\text{ème}}$ élément du vecteur Z_j vaut 1 ou 0 (c'est 1 si et seulement si l'individu i a choisi la modalité j).

Ainsi, on peut définir la matrice D ($N \times M$) composée de 0 et de 1 et dont les colonnes sont les vecteurs Z_j . Cette matrice est composée de K blocs où chaque $(N \times m_k)$ bloc contient les N réponses de la question k :

$$D = (Z_1, \dots, Z_{m_1}, \dots, Z_j, \dots, Z_M)$$

La matrice D est appelée tableau disjonctif complet et on note :

$$D = (d_{ij}), \quad i = 1, \dots, N; \quad j = 1, \dots, M.$$

Ce tableau D contient toute l'information concernant les individus. Il est le résultat brut de toute enquête.

Figure 1: exemple de tableau disjonctif complet.

	m_1			m_2		m_3		
<i>Ind</i>	1	2	3	1	2	1	2	3
1	0	1	0	0	1	1	0	0
2	1	0	0	1	0	0	1	0
...								
...								
<i>i</i>	0	0	1	0	1	0	0	1
<i>N</i>	0	0	1	1	0	0	0	1

- Tableau disjonctif complet corrigé :

Notons D^c la matrice corrigée de D , dont les entrées d_{ij}^c sont données par :

$$d_{ij}^c = \frac{d_{ij}}{\sqrt{d_{.i}} \sqrt{d_{.j}}}$$

La matrice D^c n'est pas symétrique.

2) Table de Burt :

La table de Burt est définie par : $B = D' D$. Où D' est la transposée de la matrice D . La matrice B de format $(M \times M)$ est composée de $K \times K$ blocs, parmi lesquels le (k, l) bloc B_{kl} (pour $1 \leq k \neq l \leq K$) est le tableau de contingence qui croise la question k et la question l . le bloc B_{kk} est une matrice diagonale dont les termes sont les effectifs des individus qui ont choisi respectivement les modalités $1, \dots, m_k$ pour la question k .

La table de Burt peut être vue comme un tableau de contingence généralisé à plus de deux variables.

Soient b_{jl} les entrées de la matrice B . L'entrée représente le nombre d'individus qui choisissent à la fois les modalités j et l .

Si j et l sont deux modalités différentes de la même question, $b_{jl} = 0$.

Si $j = l$, l'entrée b_{jj} est le nombre d'individus qui choisissent la modalité j . Dans ce cas, on n'utilise qu'un seul indice et l'on écrit b_j à la place de b_{jj} . Ce nombre n'est rien d'autre que la somme des éléments du vecteur Z_j .

Chaque ligne ou colonne de la matrice B caractérise une modalité d'une question (aussi appelée variable). Nous représentons ci-dessous la table de Burt associée au tableau disjonctif complet ($K=3, m_1=3, m_2=2$ et $m_3=3$) reproduit dans la figure 2

Figure 2: Exemple de Table de Burt.

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
Z_1	b_1	0	0	b_{14}	b_{15}	b_{16}	b_{17}	b_{18}
Z_2	0	b_2	0	b_{24}	b_{25}	b_{26}	b_{27}	b_{28}
Z_3	0	0	b_3	b_{34}	b_{35}	b_{36}	b_{37}	b_{38}
Z_4	b_{41}	b_{42}	b_{43}	b_4	0	b_{46}	b_{47}	b_{48}
Z_5	b_{51}	b_{52}	b_{53}	0	b_5	b_{56}	b_{57}	b_{58}
Z_6	b_{61}	b_{62}	b_{63}	b_{64}	b_{65}	b_6	0	0
Z_7	b_{71}	b_{72}	b_{73}	b_{74}	b_{75}	0	b_7	0
Z_8	b_{81}	b_{82}	b_{83}	b_{84}	b_{85}	0	0	b_8

La somme totale des entrées de B est $b = \sum_{j,l} b_{jl} = K^2N$

Sachant que pour chaque j on a $\sum_l b_{jl} = Kb_j$

et que :

$$\sum_j b_j = \sum_{k=1}^K \sum_{l=1}^{m_k} b_l = \sum_{k=1}^K N = KN$$

On en déduit que :

$$b = \sum_{j,l} b_{jl} = \sum_j \sum_l b_{jl} = \sum_j Kb_j = K^2N$$

- *Table de Burt corrigé :*

Notons B^c la matrice corrigée de B, dont les entrées b^c_{ij} sont définies par :

$$b^c_{jl} = \frac{b_{jl}}{\sqrt{b_j} \cdot \sqrt{b_l}} = \frac{b_{jl}}{K \sqrt{b_j} \sqrt{b_l}}$$

Puisque $b_j = Kb_j$ et $b_l = Kb_l$

La matrice de Burt corrigée est symétrique.