

CHAPITRE III

Réseaux de Neurones Artificiels et Analyse En Composantes principales

1. Introduction :

Partant d'un souhait initial de mieux comprendre le fonctionnement du cerveau, les réseaux de neurones artificiels sont une classe de modèles ayant eu un impact important dans le domaine de l'analyse et le traitement des données. L'inspiration pour les réseaux de neurones provient cependant de la volonté de créer des systèmes artificiels sophistiqués, voir intelligents, capables d'effectuer des opérations semblables à celle que le cerveau humain effectue de manière routinière. Les réseaux de neurones intéressent actuellement différents domaines tels que : la médecine, l'électronique, l'informatique, l'automatique, la robotique, la classification, le contrôle, le traitement des signaux, le traitement des images, traitement de l'eau, etc.

On s'attardera un petit peu plus sur le RBF (radial basic fonction).

Les techniques statistiques multi-variables comme l'analyse en composantes principales (ACP) ont été utilisées avec succès dans le domaine du traitement d'eau potable. C'est un outil capable de compresser des données et qui permet de réduire leur dimensionnalité de sorte que l'information essentielle soit conservée et plus facile à analyser que dans l'ensemble original des données. Le but principal de l'ACP est de trouver un ensemble de facteurs (composantes) qui ait une dimension inférieure à celle de l'ensemble original de données et qui puisse décrire correctement les tendances principales. Une limitation importante de la surveillance basée sur l'ACP est que la représentation obtenue est invariante dans le temps, alors qu'elle aurait besoin d'être mise à jour périodiquement. Nous retournerons plus en détails sur la description des ACPs

2. Les réseaux de neurone.

Les réseaux de neurones sont composés d'éléments simples (ou neurones) fonctionnant en parallèle. Ces éléments ont été fortement inspirés par le système nerveux biologique. Comme dans la nature, le fonctionnement du réseau (de neurone) est fortement influencé par la connections des éléments entre eux. On peut entraîner un réseau de neurone pour une tâche spécifique (reconnaissance de caractères par exemple) en ajustant les valeurs des connections (ou poids) entre les éléments (neurone).

En général, l'apprentissage des réseaux de neurones est effectué de sorte que pour une entrée particulière présentée au réseau corresponde une cible spécifique. L'ajustement des poids se fait par comparaison entre la réponse du réseau (ou sortie) et la cible, jusqu'à ce que la sortie corresponde (au mieux ?) à la cible. On utilise pour ce type d'apprentissage dit supervisé un nombre conséquent de pair entrée/sortie.

L'apprentissage « par paquet » (batch training) du réseau consiste à ajuster les poids et biais en présentant les vecteurs d'entrée/sortie de tout le jeu de données.

L'apprentissage « pas à pas ou séquentiel » (incremental training) consiste à ajuster les poids et biais en présentant les composantes du vecteur d'entrée/sortie les unes après les autres. Ce type d'apprentissage est souvent qualifié d'apprentissage « en ligne » (« on line » training) ou « adaptatif » (« adaptive » training).

L'apprentissage permet aux réseaux de neurones de réaliser des taches complexes dans différents types d'application (classification, identification, reconnaissance de caractères, de la voix, vision, système de contrôle...). Ces réseaux de neurones peuvent souvent apporter une solution simple à des problèmes encore trop complexes ne pouvant être résolus rapidement par les ordinateurs actuels (puissance de calcul insuffisante) ou par notre manque de connaissances.

La méthode d'apprentissage dite supervisé est souvent utilisée mais des techniques d'apprentissage non supervisé existent pour des réseaux de neurones spécifiques. Ces réseaux peuvent, par exemple, identifier des groupes de données (réseaux de Hopfield).

Les réseaux de neurones ont une histoire relativement jeune (environ 50 ans) et les applications intéressantes des réseaux de neurones n'ont vu le jour qu'il à une vingtaine d'année (développement de l'informatique) [23].

2.1 Les bases biologiques

Les cellules nerveuses, appelées neurones sont des éléments de base du système nerveux central qui se compose d'environ 100 milliards de neurones (10¹¹), le neurone est une cellule composée de trois parties :

- Un corps cellulaire** ou en anglais « Cell body » qui contient le noyau et se charge d'effectuer les transformations biochimiques essentielles à la synthèse des éléments assurant la vie du neurone.

- Les dendrites** qui sont des ramifications du corps cellulaire. Elles permettent au neurone de capter les signaux lui provenant de l'extérieur.

- L'axone** généralement plus long que les dendrites, il se ramifie à son extrémité où il communique avec les autres neurones. Il sert de moyen de transport pour les signaux émis par le neurone.

Les connexions entre neurones sont réalisées au niveau des synapses, lieu de proximité d'axone émetteur et dendrites réceptrices

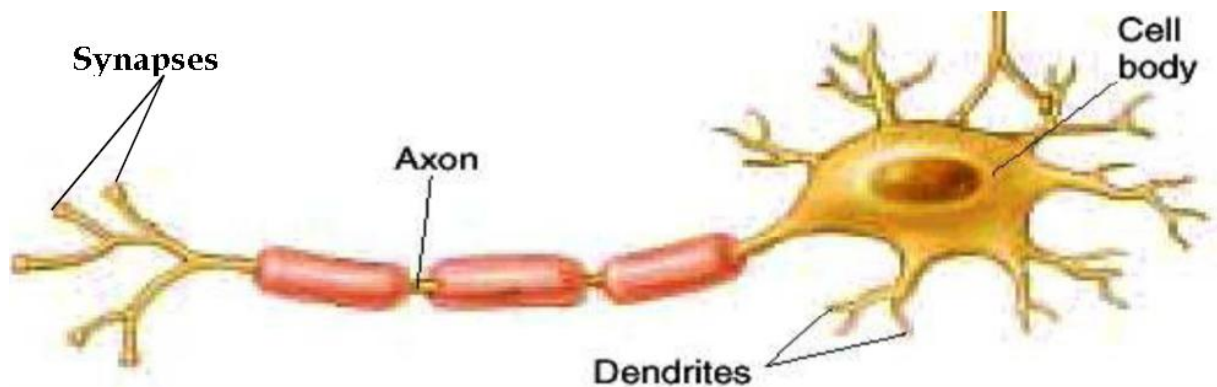


Figure III.1. : Un neurone biologique et ses principaux composants.

2.2 Du neurone biologique au neurone formel

Le neurone formel est une modélisation mathématique qui reprend les principes de fonctionnement du neurone biologique.

Un neurone formel est constitué d'un noyau, d'une liaison synaptique de sortie et des liaisons synaptiques d'entrée. A chaque liaison d'entrées est attachée un poids appelé « poids synaptique ».

Les poids synaptiques : ce sont des coefficients numériques indiquant l'importance de la connexion.

Sur la figure III.2, le neurone a trois connexions en entrée le reliant à trois autres neurones, il reçoit de l'information provenant de chacun de ces trois neurones.

Toutes les connexions n'ont pas d'une importance égale par rapport au neurone.

Certaines sont plus importantes que d'autres. En effet, tout se passe comme si le neurone ne reçoit qu'une entrée E et que celle-ci prend la valeur une fois l'entrée connue, le neurone effectue une opération qui dépend de E. cela revient à dire qu'il applique une fonction f à la valeur E, cette fonction f est appelée fonction d'activation (fonction de traitement des entrées fournissant la sortie du neurone).

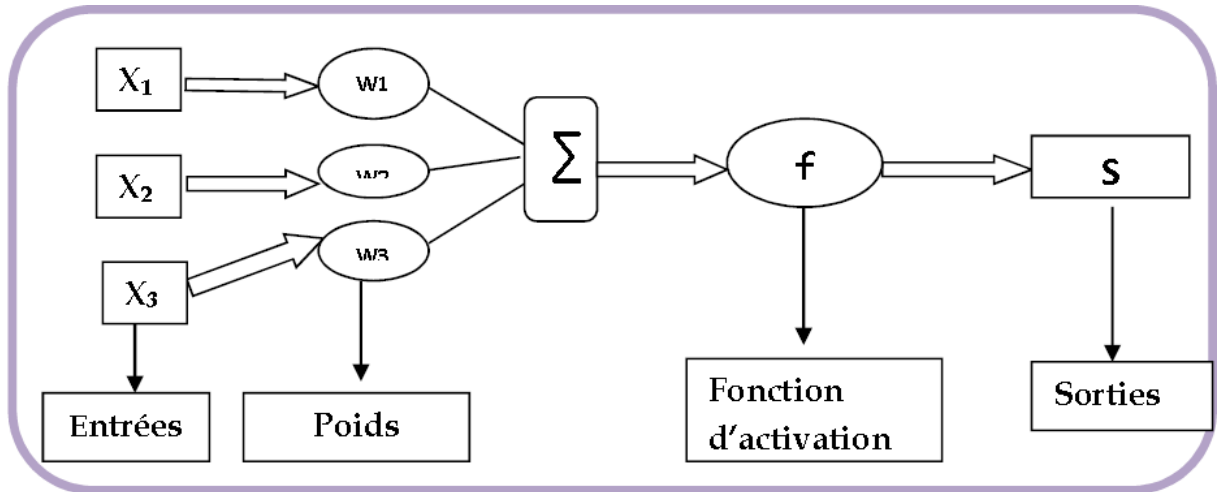


Figure III.2 : Un neurone formel

Entrees (x1, x2, x3) +poids (w1, w2, w3)

$$E = w1 * x1 + w2 * x2 + w3 * x3$$

(6)

$$\text{Sortie } S = f(E)$$

De façon générale, on définit un neurone formel par les cinq paramètres suivants :

- 1-La nature des entrées (booléenne ou réelle)
- 2-La fonction d'entrée totale, définissant le prétraitement effectué sur les entrées
- 3- La fonction d'activation du neurone définissant son état interne en fonction de la somme pondérée de ses entrées.
- 4-La fonction de sortie calculant la sortie du neurone en fonction de son état d'activation
- 5- La nature des sorties du neurone.

3. Définition d'un réseau de neurone :

Un réseau de neurone est un assemblage de neurone formel associé en couches fonctionnant en parallèle.

Dans un réseau, chaque sous-groupe fait un traitement indépendant des autres et transmet le résultat de son analyse au sous-groupe suivant. L'information donnée au réseau va donc se propager couche par couche, de la couche d'entrée à la couche de sortie, en passant soit par aucune, une ou plusieurs couches intermédiaires (couches cachées). Excepter par les couches d'entrées et les couches de sorties, chaque neurone de la dans une couche est connecté à tous les neurones de la couche suivante et de la couche précédente.

Contrairement à chacune des fonctions d'activation f , la fonction g qui transforme les valeurs d'entrée en valeurs de sortie à l'échelle du réseau ne peut pas être explicitée facilement. Elle est en effet beaucoup plus compliquée puisqu'elle est constituée de la superposition de toutes les fonctions f de chaque neurone.

Un réseau de neurones peut donc être représenté par les poids w des différents neurones.

Ces poids peuvent varier au cours du temps, en fonction des entrées présentées E . ce qui représente un vrai problème c'est comment modifier les valeurs de ces poids (la phase d'apprentissage).

4. Les réseaux de neurones à base radiale

Les réseaux de neurones à base radiale sont une classe particulière des réseaux de neurones multicouches. [25]

4.1. Présentation des réseaux RBF

L'idée générale des réseaux RBF dérive de la théorie d'approximation des fonctions, ces réseaux sont une architecture Feedforward puissante. Ce type de réseaux a été introduit pour la première fois par Hardy, et la théorie correspondante a été développée par Powell, ensuite, ces réseaux ont pris le terme de réseaux de neurones grâce à Broomhead et Lowe. Sans oublier les oeuvres de MOODY et DARKEN (1989) d'une part, et de POGGIO et GIROSI

(1990) d'autre part. La raison de son application vient du fait que le réseau utilise des fonctions gaussiennes standard qui sont à symétrie radiale. Son apprentissage est basé sur l'algorithme K-means et l'algorithme des moindres carrées [24].

Les réseaux de neurones RBFs, sont principalement utilisés pour résoudre des problèmes d'approximation de fonctions dans des espaces de grandes dimensions.

Ils sont lus adaptés, en raison d'apprentissage local. Ce type d'apprentissage peut rendre le processus d'entraînement bien plus rapide que dans le cas d'un MLP, qui apprend de façon globale.

4.2 Architecture générale d'un réseau RBF:

Pour des raisons de simplicité, on a décidé de faire une petite dualité entre le réseau

RBF et le PMC, en précisant les ressemblances et les différences entre les deux types de réseaux. Ce choix est justifié par la popularité des PMCs et leur vaste utilisation dans les applications industrielles. Un réseau de neurone de type RBF est un PMC spéciale, son architecture est identique à celle d'un PMC à une seule couche cachée donc on peut dire qu'il prend toutes les caractéristiques d'un PMC simple sauf qu'il diffère en quelques points nous citons quelques uns :

Le nombre des couches cachées :

Un réseau RBF ne peut contenir qu'une seule couche cachée, son architecture est fixée pour tous les problèmes à étudier.

La fonction d'activation :

Le réseau RBF utilise toujours une fonction dite à base radiale centrée d'un point et munie d'un rayon.

Les poids synaptiques :

Les poids entre la couche d'entrée et la couche cachée dans les modèles neuronaux de type RBF sont toujours d'une valeur d'unité, c'est-à-dire que l'information inscrite sur la couche d'entrée sera retransmise sans distorsion vers les neurones de la couche cachée.

En ce qui concerne les ressemblances entre un réseau RBF et un PMC, on peut mentionner quelques points :

La fonction de sortie :

Généralement une simple fonction linéaire qui renvoie une sommation pondérée des valeurs calculées par les neurones de la couche cachée. Bien sûr, ce n'est pas toujours le cas, parfois l'utilisation d'autres fonctions pourrait être plus adéquate dans un problème donné.

Le sens des connexions :

Les connexions entre les couches suivent le même sens, on peut dire qu'elles ne sont pas récurrentes, et chaque neurone est entièrement connecté vers les neurones de la couche suivante.

L'apprentissage :

Pour calculer les poids de la couche de sortie, on utilise un apprentissage supervisé pour les deux types de réseaux.

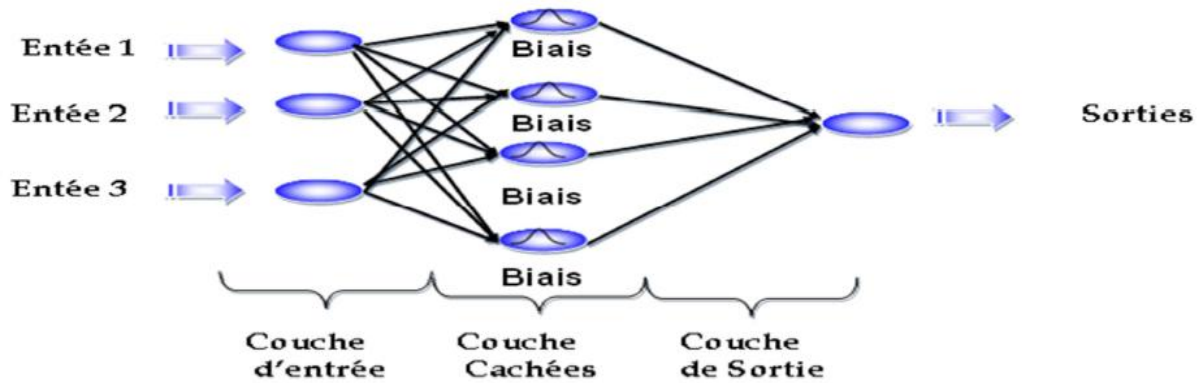


Figure III.3. Architecture d'un Réseau de Neurone RBF

4.3 La fonction à base radiale :

Les fonctions à base radiale (ou RBF), sont apparus à la fin des années 80 comme de nouvelles variantes des réseaux de neurones. Divers types de fonctions peuvent être utilisées comme noyau ou fonction de base, la fonction gaussienne reste cependant la plus utilisée [2].

La fonction **gaussienne** ne répond qu'à une petite région de l'espace d'entrée, région sur laquelle elle est centrée. Sa forme générale est donnée par :

$$H(y) = e^{\left(\frac{-|y-c|^2}{2\mu^2}\right)} \quad (7)$$

Avec y est une entrée scalaire, et $\mu > 0$.

Les paramètres de cette fonction sont le centre C et le rayon noté par μ . La raison essentielle du choix de la gaussienne comme fonction de base des RBFs est que cette fonction est factorisable. En effet, on peut facilement démontrer que parmi toutes les fonctions de base, le gaussien est la seule fonction qui peut être décomposée en produit de fonctions gaussiennes unidimensionnelles.

$$H(y) = \pi i e^{\left(\frac{-|y-c|^2}{2\mu^2}\right)} \quad (8)$$

Avec $y = [y_i]$ et $C = [C_i]$. Cette particularité devient intéressante pour l'adéquation biologique des réseaux de neurones artificiels.

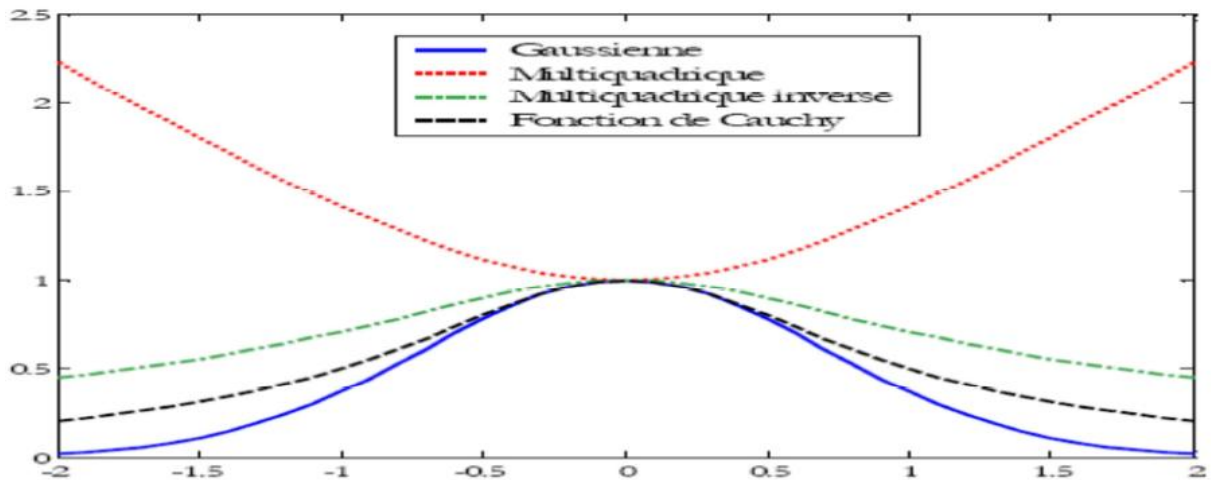


Figure III.4 Quelques fonctions radiales

4.4 La phase d'apprentissage :

Plusieurs méthodes d'apprentissage ont été développées pour ces types de réseaux, on dit souvent que l'apprentissage du réseau RBF est hybride : non supervisé pour l'étape de construction du réseau, et supervisé pour la détermination des poids de la couche de sortie. Lors de l'étape de construction du réseau, il y a quatre paramètres principaux à régler :

Le **nombre** de neurones RBF.

La **position** des centres des gaussiennes de chacun des neurones.

La **largeur** de ces gaussiennes.

Le **poids** des connexions entre les neurones RBF et le(s) neurone(s) de sortie.

- L'étape de construction du RBF :(non supervisé)

Toute modification d'un de ces paramètres entraîne directement un changement du comportement du réseau.

Le **nombre** de neurones RBF (N) et la **position** des gaussiennes sont deux paramètres intimement liés.

Nombre et position avec $N < I$:

N : le nombre de neurone.

I : le nombre d'exemple soumis au réseau.

Dans ce cas-ci, le nombre de neurones RBF devient un véritable paramètre. Il n'existe pas de méthode pour le déterminer. Il s'agit donc de trouver le nombre de centroïdes adéquat lié au problème donné. Une fois le nombre de centroïdes choisi, il faut déterminer leur position.

Pour ce faire, il existe aussi plusieurs techniques parmi ces techniques on utilise la méthode k-means.

Une fois tous les centres C_j choisis, il faut déterminer la **largeur** (β) des gaussiennes.

Une règle empirique consiste à prendre :

$$\beta = \frac{d}{\sqrt{M}} \quad \text{Avec } M = \text{nombre des centroïdes et } d = \max \|c_i - c_j\|, 1 \leq i, j \leq M$$

Si on choisit un β égal pour toutes les gaussiennes.

Mais rien n'impose de prendre la même valeur de β pour chaque centroïde. Dans ce cas, une

$$\text{autre règle nous dit : } \beta_j = \frac{1}{N \sqrt{8}} \sum_{i=1}^N \|X_i - C_j\| \quad (9)$$

5. Avantages et inconvénients des réseaux RBF :

Alors que l'on pouvait redouter d'avoir inventé un modèle aux belles propriétés théoriques, mais inutilisable en raison de durées d'apprentissage prohibitives, c'est tout le contraire qui se produit : la construction d'un réseau RBF est rapide et facile, et c'est là le principal avantage de la technique [24], [26].

Mais cet avantage se paye par des performances qui ne peuvent être aussi bonnes que celles de techniques plus sophistiquées (comme le Perceptron Multicouches). En particulier, les réseaux RBF sont peu performants :

- Sur les données dans des espaces de grande dimension (beaucoup de variables d'entrée). Cette faiblesse est propre à toutes les techniques locales.
- Sur des données très bruitées. La reconstruction locale de la fonction empêche le réseau de "moyenner" le bruit sur tout l'espace (comparez avec la Régression Linéaire, dont l'objectif est justement de moyennner le bruit sur les données).

6. L'Analyse en composant principal (ACP) :

L'ACP est une méthode d'analyse multivariable qui a été souvent utilisée pour le traitement statistique de base de données multidimensionnelle. L'analyse factorielle en composantes principales est un traitement statistique de données dont le but est de représenter et d'expliquer les liaisons statistiques entre les phénomènes. Elle permet d'identifier des variables sous-jacentes, ou facteurs qui expliquent les corrélations à l'intérieur d'un ensemble de variables observées. Elle est souvent utilisée pour réduire un ensemble de données, et dans l'agrégation de l'information, en identifiant un petit nombre de facteurs qui expliquent la plupart des variances observées dans le plus grand nombre de variables manifestes. On peut également utiliser l'analyse factorielle pour résumer, synthétiser, et hiérarchiser l'information

contenue dans un tableau de n lignes (les individus) et p colonnes (les variables). Les n individus sont décrits par un nuage de p variables. L'information représentée par ce nuage revient à la dispersion des n points. Produire un résumé de cette information c'est projeter ces points dans un espace de dimension inférieure à p le nombre de variables initiales. Les axes de ce sous-espace sont dits « axes factoriels » ou « facteurs ». Chaque variable p porte en elle, une part d'information originale ou part d'inertie et une part d'information originale redondante avec les autres, venant des corrélations entre variables. C'est cette part d'information redondante qui va être regroupée dans le résumé factoriel.

Les facteurs sont hiérarchisés de la manière suivante :

Le 1er axe concentre le maximum de l'information : c'est l'axe de la plus grande dimension du nuage de points et il fournit le meilleur résumé dans un espace à une dimension, mais il laisse des résidus d'information.

Le 2e axe concentre le maximum de l'information restante, il est orthogonal au premier et c'est le meilleur résumé dans un espace à deux dimensions. Mais, de même il laisse aussi des résidus.

Le 3e axe prend encore une part d'information moindre, il est orthogonal aux deux premiers. Et ainsi de suite, pour les axes suivants tant que l'on pense qu'ils apportent encore de l'information.

Le nombre de composantes en théorie est égal au nombre de variables originelles. Mais, en pratique, les premières directions permettent de couvrir un pourcentage élevé (80%, 90%) de toutes les données originelles et sont donc utilisées pour restreindre l'espace d'observation.

Concepts de base de l'ACP

Les composantes principales sont déterminées grâce au calcul des vecteurs propres de la matrice de variance. Les vecteurs propres avec les plus grandes valeurs propres seront utilisés comme les vecteurs de la base sur lesquels les données seront projetées [27 ; 28].

Pour un groupe de données, $X = (x_1, x_2, \dots, x_N)$, l'approche par ACP s'effectue par le calcul de la matrice de covariance :

$$CO = \frac{1}{N} \sum_{i=1}^N X_i X_i^T \quad (10)$$

On utilise ensuite n'importe quel algorithme de détermination des vecteurs propres pour trouver les valeurs propres de la matrice de covariance des données :

$$COU = \lambda U$$

Où λ est la valeur propre, et U est le vecteur propre correspondant. Les m composantes principales des données n sont les directions orthogonales m dans les espaces de n qui capturent la plus grande variation des données [10].

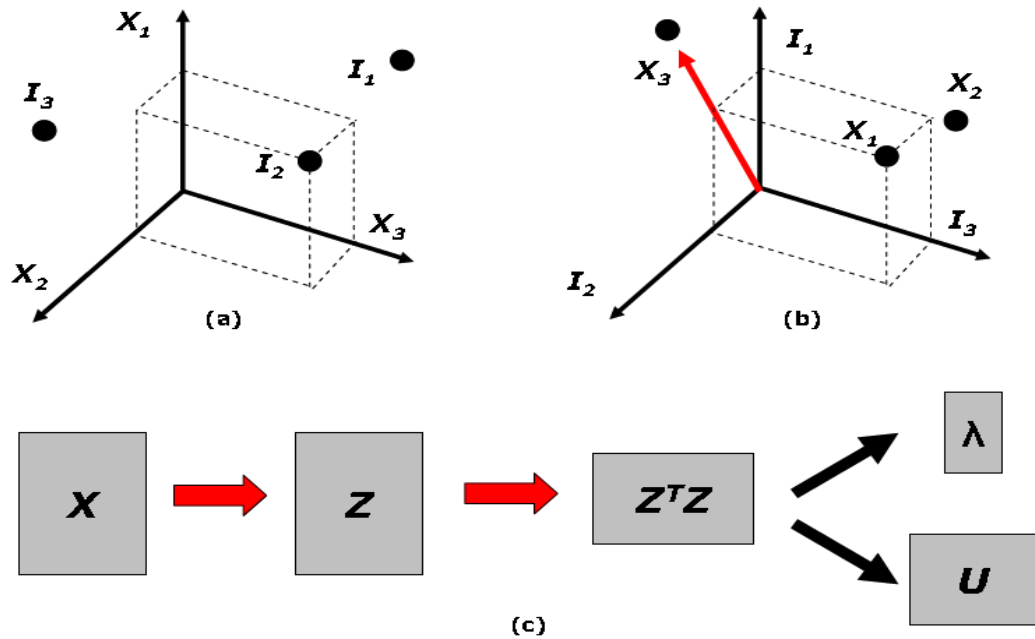


Figure I.5 : Projection (a) dans l'espace des variables, (b) dans l'espace des individus et (c) décomposition de l'ACP [10]

7. Conclusion

Nous venons de voir dans ce chapitre des différents modèles de réseaux de neurones, c'est-à-dire une vue globale sur les réseaux de neurones.

Nous nous sommes intéressés ensuite à la description des réseaux de neurones à base radiale RBF.

Contrairement aux méthodes classiques qui ont montré leurs limites, les réseaux de neurones ont montré leurs tendances à s'adapter à des problèmes complexes grâce à leur grande capacité de calcul et d'apprentissage. Ils sont l'objet d'utilisation dans les différents domaines tels que : La reconnaissance des formes et le traitement de l'eau.

L'étude et l'analyse des performances des ces méthodes choisies appliquées au domaine de contrôle et de surveillance des eaux potables, constituent notre principal objectif. Une étude en simulation ayant pour but d'évaluation des performances de ces techniques, fera l'objet du chapitre suivant. L'évaluation des résultats, reflétant les performances obtenues, nous conduira au meilleur choix de la méthode la mieux adaptée à l'application. L'architecture du système de contrôle et de surveillance proposé est aussi présentée.