

# Régressions de variables qualitatives, analyse des correspondances et réseaux neuronaux

C.Mullon<sup>a</sup> L. Bochereau,<sup>b</sup> B. Palagos<sup>b</sup> G. Pichon<sup>a</sup>

## 1 Introduction : L'apprentissage de règles de discrimination.

Les méthodes de discrimination visent à établir l'appartenance d'un individu à une catégorie donnée à partir de certaines de ses caractéristiques. Elles donnent lieu à toute une série de techniques spécialisées relevant selon les cas de l'inférence statistique (analyse discriminante de Fisher), de l'analyse des données (régressions par boules), de l'intelligence artificielle (apprentissage supervisé, cases based reasoning), etc.

Le cadre commun à toutes ces techniques est le suivant :

- On dispose d'un premier échantillon d'individus (échantillon d'apprentissage) dont on connaît les caractéristiques explicatives et la catégorie d'appartenance ;
- A partir de cet échantillon, on fait apparaître des règles fixant l'appartenance à une catégorie en fonction des caractéristiques ; ces règles sont de forme très diverses : règles logiques, probabilités *a posteriori*, minimisation d'une distance, etc.
- Enfin, la plupart du temps par application de ces règles sur un deuxième échantillon (échantillon test), on propose différents critères de validation.

Nous souhaitons examiner ici deux techniques particulières de discrimination, l'une basée sur les principes de la régression sur facteurs, l'autre sur l'apprentissage par réseau neuronal, et en montrer d'étroites connexions.

## 2 Le problème de la régression dans le cadre de l'analyse factorielle

### 2.1 Introduction

Dans le cadre de l'analyse factorielle, le problème de la discrimination se pose de la façon suivante :

---

<sup>a</sup>Laboratoire d'Informatique appliquée, ORSTOM, 70-74 route d'Aulnay, 93140 Bondy

<sup>b</sup>Laboratoire d'intelligence artificielle, CEMAGREF, Parc de Tourvoie, B.P. 121, 92185 Antony

- Nous disposons d'un tableau  $K_{ij}$ , constitué des variables explicatives, et dont on sait justifier la constitution et le traitement par analyse de correspondances.
- Nous disposons également d'un second tableau  $K_{is}$  constitué de plusieurs colonnes correspondant aux mêmes lignes et "codant" l'appartenance à une catégorie.
- Nous voulons savoir s'il existe entre le tableau  $K_{ij}$  et chaque colonne supplémentaire  $K_{is}$  des relations quantifiables.

Cette situation nous autorise à parler de modèle et d'apprentissage : les données de départ permettent d'estimer, d'apprendre le modèle, son extrapolation à de nouvelles données permet d'effectuer des prévisions. Bien entendu l'essentiel se trouve dans l'identification du modèle, c'est-à-dire dans la forme des relations quantifiables en question et dans l'explication causale qu'elles permettent d'élaborer.

## 2.2 Les tableaux traités par analyse factorielle des correspondances

Les tableaux dont est justifié le traitement par analyse factorielle des correspondances, c'est-à-dire les tableaux de contingence (fréquence) ou les tableaux disjonctifs complets (tableau logique associé à un questionnaire, comprenant une colonne par modalité de réponse à une question) sont caractérisés par le fait qu'on sait significativement calculer les marges en lignes ou en colonnes

$$K_i = \sum_j K_{ij}, K_j = \sum_i K_{ij} \text{ et } K = \sum_{i,j} K_{ij}$$

les fréquences absolues

$$F_{ij} = \frac{K_{ij}}{K}, F_i = \frac{K_i}{K} \text{ et } F_j = \frac{K_j}{K}$$

les profils des différentes colonnes

$$F_i^j = \frac{K_{ij}}{K_j}, F_j^i = \frac{K_{ij}}{K_i} \text{ et } F_i^s = \frac{K_{is}}{K_s}$$

## 2.3 Le modèle de transition

Compte tenu de la forme particulière des tableaux de contingence, on peut envisager que les profils  $F_i^j$  et  $F_i^s$  sont liés par une relation de la forme :

$$F_i^s = \sum_j F_j^s F_i^j \quad (1)$$

On y interprète le coefficient  $F_j^s$  de la façon suivante :

- si  $F_i^j$  est la probabilité d'avoir le critère  $i$  sachant que l'on a le critère  $j$
- si  $F_i^s$  est la probabilité d'avoir le critère  $i$  sachant que l'on a le critère  $s$

- alors  $F_j^s$  est la probabilité d'avoir le critère  $j$  sachant que l'on a le critère  $s$ .

On appelle une formule comme (1) formule de transition. En général, une relation exacte comme

$$F_i^s = \sum_j F_j^s F_i^j$$

n'existe pas ; on se pose alors le problème suivant : quels sont les coefficients  $F_j^s$  pour lesquels les quantités

$$E_i = F_i^s - \sum_j F_j^s F_i^j$$

sont les plus petites possibles ?

## 2.4 Le cadre géométrique

Rappelons le cadre géométrique que l'analyse des correspondances nous permet de construire pour un tel ensemble de données.

Nous considérons dans l'espace  $R^{NI}$ , muni du produit scalaire et de la distance du  $\chi^2$  :

$$\langle X, Y \rangle = \sum_i \frac{X_i Y_i}{F_i}$$

le nuage des points  $X_j$  correspondants aux colonnes  $j$  ; la coordonnée d'un point  $X_j$  dans la direction  $i$  est donnée par  $F_i^j$ .

Cet ensemble de points supporte un espace  $E^J$  de plus petite dimension.

## 2.5 Solution dans le cadre géométrique

La colonne  $s$  a un profil :

$$F_i^s = \frac{K_{is}}{K^s}$$

il donne lieu à un point  $X^s$  et ce point peut être projeté sur l'espace  $E^J$ . Cette projection est notée  $P^s$ .

Le point  $P^s$  est dans l'espace  $E^J$  ; on a donc une relation de la forme :

$$\vec{G}P^s = \sum_i F_j^s \vec{G}X^j$$

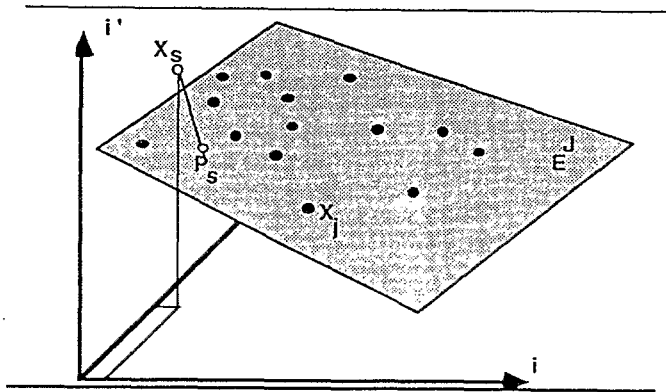


figure 1 : le nuage et le point correspondant à la colonne supplémentaire.

Puisqu'il s'agit d'une projection, les coefficients  $F_j^s$  sont tels que  $\|X^s - P^s\|^2$  est le plus petit possible.

★ **Proposition :**

Les coefficients  $F_j^s$  sont la solution du problème qui consiste à minimiser la quantité

$$\sum_i \frac{(F_i^s - \sum_j F_j^s F_i^j)^2}{F_i}$$

Ils constituent la solution que nous retiendrons pour le problème de minimisation posé ci-dessus.

★ **Proposition :**

le problème qui consiste à chercher les coefficients  $F_j^s$  pour lesquels

$$\sum_i \frac{E_i^2}{F_i} \text{ est minimum (où } E_i = F_i^s - \sum_j F_j^s F_i^j \text{ )}$$

a une solution, donnée par :  $F_j^s = \sum_{j'} M_{jj'}^{-1} N_{j'}$ , avec

$$M_{jj'} = \sum_i \frac{F_j^i F_{j'}^i}{F_i} \text{ et } N_{j'} = \sum_i \frac{F_i^s F_{j'}^i}{F_i}$$

## 2.6 Le cadre factoriel

Formuler ce problème dans le cadre factoriel revient à effectuer un changement de repère. Il se trouve que cela conduit à des formules plus simples, s'appliquant à l'ensemble des situations, qui sont interprétables, ce qui, rappelons-le, est fondamental en analyse de données. Nous nous plaçons dans la situation où, à la suite d'une analyse factorielle des correspondances, nous avons obtenu les directions principales du nuage  $N(J)$  ; il s'agit d'un ensemble ordonné de directions orthogonales entre elles, et maximisant l'inertie directionnelle, c'est-à-dire la dispersion du nuage en projection le long d'une de ces directions.

Rappelons brièvement le formulaire de l'analyse des correspondances : Les coordonnées factorielles sur  $I$  sont notées  $F_\alpha(i)$  et celles sur  $J$  sont notées  $G_\alpha(j)$ . Ces coordonnées sont de moyenne nulle :

$$\sum_i F_i F_\alpha(i) = 0 \quad \text{et} \quad \sum_j F_j G_\alpha(j) = 0 \quad (2)$$

Elles sont de variance

$$\lambda_\alpha = \sum_i F_i F_\alpha(i)^2 = \sum_j F_j G_\alpha(j)^2 \quad (3)$$

Elles sont de covariance 0

$$\sum_i F_i F_\alpha(i) F_\beta(i) = \sum_j F_j G_\alpha(j) G_\beta(j) = 0 \quad (4)$$

Les deux ensembles de coordonnées factorielles sont liés par les formules de transition :

$$\begin{cases} F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_j F_j^i G_\alpha(j) \\ G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_i F_i^j F_\alpha(i) \end{cases} \quad (5)$$

On retrouve les coordonnées initiales par la formule de reconstitution des données :

$$F_{ij} = F_i F_j \left( 1 + \sum_\alpha \frac{F_\alpha(i) G_\alpha(j)}{\sqrt{\lambda_\alpha}} \right) \quad (6)$$

## 2.7 Le point supplémentaire dans l'espace factoriel

Nous pouvons donc projeter le point  $P^s$  sur les directions principales, obtenant ainsi les coordonnées factorielles  $G_\alpha(s)$ .

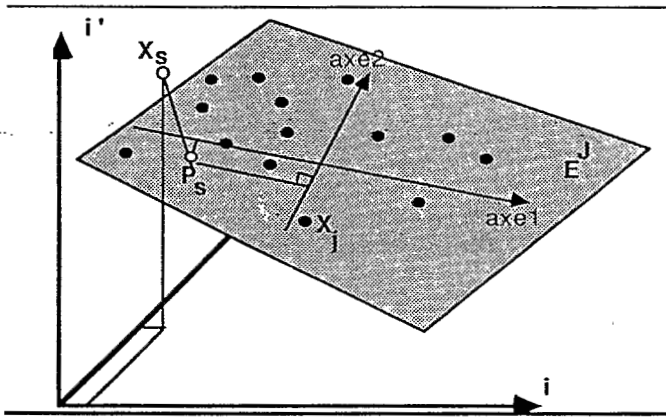


figure 2 : nuage, point supplémentaire et coordonnées factorielles du point supplémentaire.

**\* Proposition :**

$$G_{\alpha}(s) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_i F_i^s F_{\alpha}(i) \quad (7)$$

*Démonstration :*

Cette formule est un simple décalque de la formule de transition reliant les coordonnées factorielles sur l'ensemble  $J$  et celles sur l'ensemble  $I$ . Rappelons que ces formules de transitions sont exactement des formules de changement de base, entre repère initial et repère factoriel.

**\* Proposition :**

Par rapport aux coordonnées factorielles ; les coefficients  $F_j^s$  solution du problème de minimisation sont donnés par :

$$F_j^s = F_j \left( 1 + \sum_{\alpha} \frac{G_{\alpha}(s) G_{\alpha}(j)}{I_{\alpha}} \right) \quad (8)$$

*Démonstration :*

En appliquant le même raisonnement que celui qui a conduit à la formule de reconstitution des données, on trouve que le point  $P^s$  a pour coordonnées :

$$FR_i^s = F_i \left( 1 + \sum_{\alpha} \frac{G_{\alpha}(s) F_{\alpha}(i)}{\sqrt{\lambda_{\alpha}}} \right)$$

En intégrant les formules de transition :

$$F_{\alpha}(i) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_j F_j^i G_{\alpha}(j)$$

dans la formule précédente ; on obtient :

$$\begin{aligned} FR_i^s &= F_i \left( 1 + \sum_{\alpha} \frac{1}{\sqrt{\lambda_{\alpha}}} G_{\alpha}(s) \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_j F_j^i G_{\alpha}(j) \right) \\ &= F_i \left( 1 + \sum_j \sum_{\alpha} \frac{G_{\alpha}(s) G_{\alpha}(j)}{\lambda_{\alpha}} F_j^i \right) \end{aligned}$$

En écrivant que  $1 = \sum_j F_j^i$ , on obtient :

$$\begin{aligned} FR_i^s &= F_i \left( \sum_j F_j^i + \sum_j \sum_{\alpha} \frac{G_{\alpha}(s) G_{\alpha}(j)}{\lambda_{\alpha}} F_j^i \right) \\ &= \sum_j F_i F_j^i \left( 1 + \sum_{\alpha} \frac{G_{\alpha}(s) G_{\alpha}(j)}{\lambda_{\alpha}} \right) \end{aligned}$$

Mais  $F_i F_j^i = F_{ij} = F_j F_i^j$ . Et l'on a :  $FR_i^s = \sum_j F_i^j F_j^s$

$$\text{avec } F_j^s = F_j \left( 1 + \sum_{\alpha} \frac{G_{\alpha}(s) G_{\alpha}(j)}{\lambda_{\alpha}} \right)$$

C.Q.F.D.

## 2.8 L'utilisation de la régression sur facteurs pour la constitution de formules "robustes"

Comparons les deux façons d'écrire les coefficients  $F_j^s$  :

$$F_j^s = \sum_j M_{jj}^{-1} N_j, \quad \text{avec : } M_{jj} = \sum_i \frac{F_j^i F_j^i}{F_i} \quad \text{et } F_j = \sum_i \frac{F_j^i F_j^i}{F_i}$$

$$F_j^s = F_j \left( 1 + \sum_{\alpha} \frac{G_{\alpha}(s) G_{\alpha}(j)}{\lambda_{\alpha}} \right)$$

Une difficulté majeure surgit lorsque la matrice  $M_{jj}$ , n'est pas inversible ; ce qui correspond en fait à l'existence de valeurs propres  $\lambda_{\alpha}$  nulles. Si cette difficulté est insurmontable avec la seconde écriture, la première permet d'envisager une solution. Il suffit d'écrire :

$$F_j^s = F_j \left( 1 + \sum_{\alpha, \lambda_{\alpha} > 0} \frac{G_{\alpha}(s) G_{\alpha}(j)}{\lambda_{\alpha}} \right)$$

Un problème presque aussi délicat surgit lorsque la matrice  $M_{jj}$  est difficilement inversible (son déterminant est presque nul) ; cela conduit à des coefficients "instables" : de petites fluctuations des données de départ conduisent à d'amples variations des coefficients  $F_j^s$ . Ce phénomène correspond à l'existence de petites

valeurs propres  $\lambda_\alpha$ . Et là encore une solution permettant d'obtenir un modèle robuste consiste à calculer les coefficients par une formule du type :

$$F_j^s = F_j \left( 1 + \sum_{\alpha, \lambda_\alpha > \text{seuil}} \frac{G_\alpha(s)G_\alpha(j)}{\lambda_\alpha} \right)$$

Cette façon de calculer les coefficients  $F_j^s$  donne des résultats plus stables, moins sensibles à d'éventuelles perturbations aléatoires ; en effet celles-ci auront un effet sensible seulement sur les axes factoriels de faible inertie qui ne sont pas inclus dans la formule.

**Remarque :**

$$\text{En écrivant que } F_i^s = \frac{K_{is}}{K_j} \text{ et que } F_i^j = \frac{K_{ij}}{K_j}$$

l'équation  $F R_i^s = \sum_j F_i^j F_j^s$  devient :

$$K R_{is} = \sum_j C_{js} K_{ij} \text{ où } C_{js} = \frac{K_s}{K} \left( 1 + \sum_\alpha \frac{G_\alpha(s)G_\alpha(j)}{\lambda_\alpha} \right)$$

Nous appliquerons désormais cette formulation.

## 2.9 Le cas particulier de la régression d'une variable qualitative

Pour comparer deux tableaux en analyse de données, il est recommandé de se placer dans une situation où leurs marges sont égales ou proportionnelles, en effet dans ces conditions, le choix de la distance (au sens du  $\chi^2$ ) ne pose pas de difficulté.

On est donc amené à construire le tableau numérique codant une variable qualitative (avec plusieurs modalités  $s$ )  $K_{is}$ , de la façon suivante :  $K_{is} = K_i$  si l'individu  $i$  présente la modalité  $s$  ;  $K_{is} = 0$  si  $i$  ne présente pas la modalité  $s$  ; Alors, bien évidemment, les tableaux  $K_{ij}$  et  $K_{is}$  ont alors des marges identiques en  $i$  :

$$\sum_j K_{ij} = \sum_s K_{is} = K_i \quad (9)$$

En répétant les opérations vues précédemment, on reconstruit le tableau  $K_{is}$  à partir du tableau  $K_{ij}$  par des formules du type :  $K_{is} = K R_{is} + E_{is}$ , avec

$$K R_{is} = \sum_j C_{js} K_{ij} \text{ où } C_{js} = \frac{K_s}{K} \left( 1 + \sum_\alpha \frac{G_\alpha(s)G_\alpha(j)}{\lambda_\alpha} \right) \quad (10)$$

\* **Proposition :**

Le tableau reconstitué  $K R$  a les mêmes marges que le tableau initial :

$$\sum_s K_{is} = \sum_s K R_{is} = K_i \text{ et } \sum_i K_{is} = \sum_i K R_{is} = K_s$$



Démonstration :

$$\begin{aligned}
 \sum_i K R_{is} &= \sum_i \sum_j C_{js} K_{ij} = \sum_j C_{js} \sum_i K_{ij} = \sum_j C_{js} K_j \\
 &= \sum_j \frac{K_s}{K} \left( 1 + \sum_\alpha \frac{G_\alpha(s) G_\alpha(j)}{\lambda_\alpha} \right) K_j \\
 &= \frac{K_s}{K} \sum_j K_j + \frac{K_s}{K} \sum_i \sum_\alpha \frac{G_\alpha(s) G_\alpha(j)}{\lambda_\alpha} K_j \\
 &= \frac{K_s}{F} K + \frac{K_s}{K} \sum_\alpha \frac{G_\alpha(s)}{\lambda_\alpha} \sum_j K_j G_\alpha(j) = K_s \text{ car} \\
 &\quad \sum_j K_j G_\alpha(j) = 0 \text{ d'après (2)}
 \end{aligned}$$

De la même façon :

$$\begin{aligned}
 \sum_s K R_{is} &= \sum_s \sum_j C_{js} K_{ij} = \sum_j K_{ij} \sum_s C_{js} \\
 \text{Or } \sum_s C_{js} &= \sum_s K_s \left( 1 + \sum_\alpha \frac{G_\alpha(s) G_\alpha(j)}{\lambda_\alpha} \right) K_j \\
 &= K + \sum_\alpha \frac{\sum_s K_s G_\alpha(s) G_\alpha(j)}{\lambda_\alpha} \\
 \text{Mais } \sum_s K_s G_\alpha(s) &= \sum_s K_s \frac{1}{\sqrt{\lambda_\alpha}} \sum_i F_i^s F_\alpha(i) \\
 &= \frac{1}{\sqrt{\lambda_\alpha}} \sum_s \sum_i F_i^s K_s F_\alpha(i) \text{ d'après (7)} \\
 &= \frac{1}{\sqrt{\lambda_\alpha}} \sum_i F_\alpha(i) \sum_s F_i^s K_s = \frac{1}{\sqrt{\lambda_\alpha}} \sum_i F_\alpha(i) \sum_s \frac{K^{is}}{K_s} K_s \\
 &= \frac{1}{\sqrt{\lambda_\alpha}} K \sum_i F_i F_\alpha(i) = 0 \text{ d'après (2) donc}
 \end{aligned}$$

$$\sum_s C_{js} = K \text{ et finalement } \sum_s K R_{is} = \sum_j \frac{K_{ij}}{K} \sum_s C_{js} = \sum_j \frac{K K_{ij}}{K} = K_i$$

C.Q.F.D.

## 2.10 Principe de classification

Compte tenu des propriétés de marge ci dessus, on peut interpréter les quantités :

- $K_s/K$  : comme la probabilité a priori que l'individu  $i$  présente la modalité  $s$
- $K_{i,s}/K_i$  : comme la certitude que l'individu  $i$  présente ou non la modalité  $s$
- $KR_{i,s}/K_i$  : comme la probabilité a posteriori que l'individu  $i$  présente la modalité  $s$

et en déduire le principe de classification suivant : à l'individu  $i$  est associé la modalité  $s$  qui maximise  $KR_{i,s}/K_i$ . Par rapport à la plupart des méthodes de classification supervisées, qui sont assez instables dans le cas de modalités de fréquences disparates, et qui, selon les cas, sur-prédisent les modalités très rares ou très courantes, celle-ci présente l'avantage de prédire les modalités en fonction de leur fréquence effective. Ce fait est une simple conséquence de la formule :

$$\sum_i KR_{i,s} = K_s$$

## 3 Les réseaux de neurones

### 3.1 Introduction

Les réseaux de neurones sont une schématisation très poussée du fonctionnement cérébral. On y distingue des neurones formels, constitué d'un noyau, ayant un état donné (la plupart du temps 0 ou 1) et de synapses, chacune pourvues d'un poids exprimant l'intensité de la connexion qu'elles réalisent.

Un réseau neuronal est alors constitué d'un ensemble fini de neurones (figure 3).

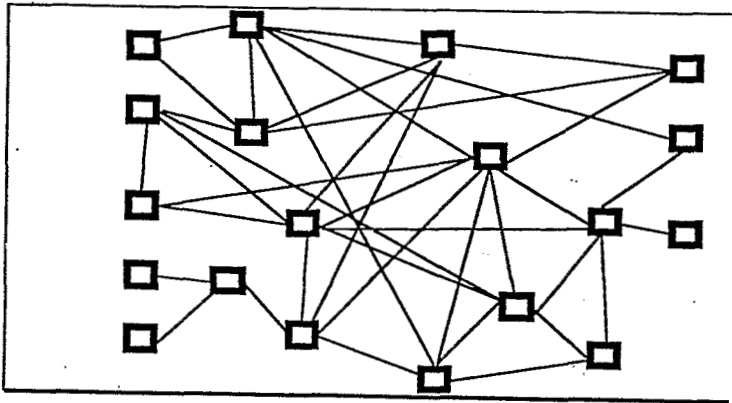


Figure 3 : réseau neuronal.

## 3.2 Le fonctionnement des réseaux neuronaux

Il suit l'algorithme suivant :

à l'instant  $t$  :

- noeud  $i$  :  $N_i(t)$  prend des valeurs déterminées, 0 ou 1
- arc entre deux noeuds :  $W_{ij}(t)$  prend des valeurs quelconques

à l'instant  $t + 1$  :

pour chaque noeud  $i$ ,

- on calcule la quantité  $s = \sum_j W_{ij}(t)N_j(t)$
- on la compare à un seuil donné  $s_0$  ; si  $s > s_0$  alors on pose  $N_{ij}(t + 1) = 1$ , sinon  $N_{ij}(t + 1) = 0$  ;

Le point le plus intéressant de la théorie des réseaux neuronaux est qu'un système aussi simple puisse traiter un grand nombre de problèmes ; et surtout d'établir une série de passerelles entre des disciplines aussi diverses que l'Intelligence artificielle, Mécanique statistique, Neurobiologie, etc...

## 4 L'apprentissage par réseau de neurones

### 4.1 Réseaux à couches avec rétropropagation

Ces réseaux permettent de formuler le problème de l'apprentissage de fonctions discriminantes que nous avons déjà traité dans le cadre factoriel.

Le cas le plus simple est celui où on distingue trois couches : celle de gauche qui contient les inputs, celle de droite qui contient les outputs, et une couche intermédiaire, dite cachée. Dans la couche des inputs, chaque neurone noté  $k$  correspond à un caractère explicatif. Dans la couche des outputs, chaque neurone noté  $j$  correspond à une modalité du phénomène à prédire. Les neurones de la couche intermédiaire sont notés  $i$ .

Alors le réseau se présente comme sur la figure 4.

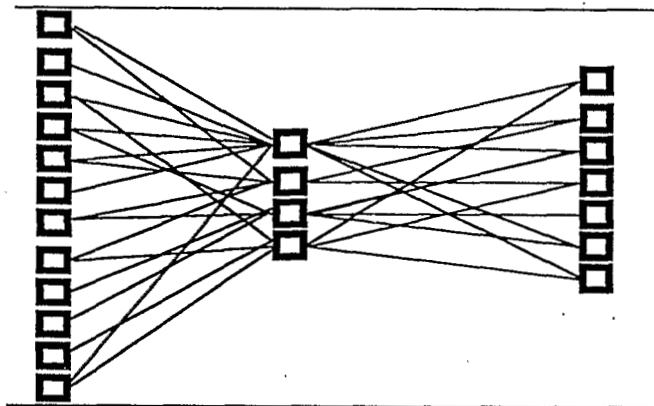


figure 4

Dans ces réseaux, chaque noeud peut prendre une valeur réelle positive (en général comprise entre -1 et 1). Un rôle particulier est joué dans le fonctionnement du réseau par les fonctions de type sigmoïde d'équation

$$f(x) = \frac{1}{1 + e^{-ax}}$$

De telles fonctions permettent de tenir compte très simplement des effets de seuil.

Remarquons que la fonction  $f(x)$  prend ses valeurs entre 0 et 1. On la remplace parfois par la fonction

$$f(x) = \frac{1 - e^{-ax}}{1 + e^{-ax}}$$

qui prend ses valeurs entre -1 et 1.

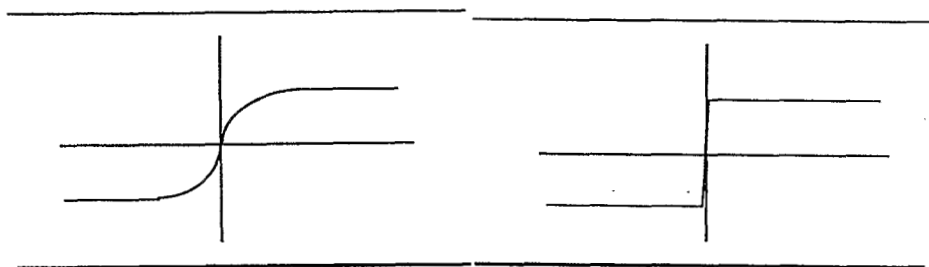


figure 5 et 6 : 5. Fonction sigmoïde entre -1 et 1.

6. Fonction certaine entre -1 et 1.

Lorsque  $a$  devient très grand, la fonction sigmoïde tend vers une fonction logique :  $f(x) = -1$  si  $x < 0$ ,  $f(x) = 1$  si  $x > 0$  :

L'algorithme est le suivant :

### Apprentissage

*Initialisation :*

$$t = 0$$

$$W_{ij}(0) = \text{petit alea}$$

*Répéter :*

pour chaque individu de l'ensemble d'apprentissage,

- début de boucle

$$t = t + 1$$

- initialiser la couche des inputs avec les caractères de l'individu  $n$  :  $x_k$
- calculer la couche intermédiaire avec la formule :  $x_i = f(\sum_k W_{ki}x_k)$
- calculer la couche des outputs avec la formule :  $x_j = f(\sum_i W_{ij}x_i)$
- réajuster alors les poids  $W_{ij}$  entre la couche finale ( $j$ ) et la couche intermédiaire ( $i$ ) ; pour cela, comparer dans la couche finale, la valeur obtenue  $x_j$  avec la valeur souhaitée  $y_j$  ; poser :  $d_j = x_j(1 - x_j)(y_j - x_j)$  puis :  $W_{ij}(t + 1) = W_{ij}(t) + \tau d_j x_i$  (où  $\tau$  est une constante à fixer en fonction de la convergence du système)
- réajuster les poids entre la couche intermédiaire ( $i$ ) et la couche initiale ( $k$ ) :  $d_i = x_i(1 - x_i) \sum_j d_j W_{ij}$  et  $W_{ij}(t + 1) = W_{ij}(t) + n d_j x_i$

- fin de boucle

### Application

- initialiser la couche des inputs avec les caractères de l'individu  $n$  :  $x_k$
- calculer la couche intermédiaire avec la formule :  $x_i = f(\sum_k W_{ki}x_k)$
- calculer la couche des outputs avec la formule :  $x_j = f(\sum_i W_{ij}x_i)$
- attribuer la catégorie  $j$  qui maximise  $x_j$

## 5 Réseaux neuronaux et régression logique

Opérons une élémentaire transformation de formule ; nous savons que la régression sur facteurs nous conduit à une formule du type :

$$KR_{is} = \sum_j C_{js} K_{ij} \text{ où } C_{js} = \frac{K_s}{K} \left( 1 + \sum_{\alpha > 0} \frac{G_\alpha(s) G_\alpha(j)}{\lambda_\alpha} \right)$$

Posons  $G_0(j) = 1, G_0(s) = 1, \lambda_0 = 1$ , et réécrivons cette formule

$$KR_{is} = \sum_j C_{js} K_{ij} \text{ où } C_{js} = \frac{K_s}{K} \left( \sum_{\alpha \geq 0} \frac{G_\alpha(s) G_\alpha(j)}{\lambda_\alpha} \right)$$

Alors si on pose :

$$W_{\alpha j} = \frac{G_\alpha(j)}{\sqrt{\lambda_\alpha}} \text{ et } W_{s\alpha} = \frac{K_s}{K} \frac{G_\alpha(s)}{\sqrt{\lambda_\alpha}}$$

et que l'on construit le réseau neuronal suivant, dans lequel la couche intermédiaire correspond aux facteurs retenus.

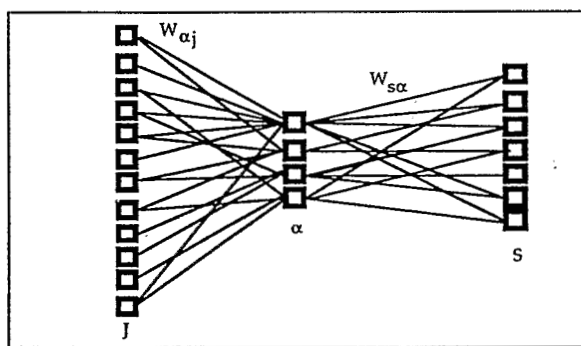


figure 7 : le réseau neuronal correspondant à la régression sur facteurs.  
où l'on utilise une fonction "sigmoïde" telle que :

- $f(x) = x$  si  $-s < x < s$  (où  $s$  est un seuil assez grand)
- $f(x) = -s$  si  $x < -s$
- $f(x) = s$  si  $x > s$

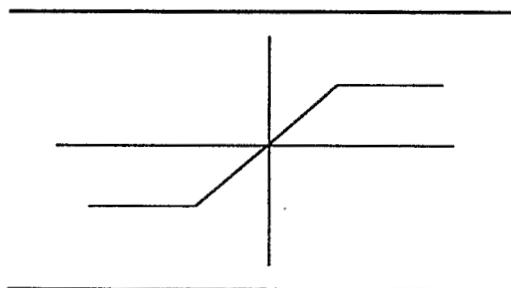


figure 8 : la fonction sigmoïde correspondante.

On voit que l'on obtient un réseau neuronal exactement du même type que celui obtenu après apprentissage selon la technique "perceptron" évoqué plus haut.

## 6 Conclusion

Nous avons montré l'analogie entre deux approches du problème de la discrimination. Pour aller plus loin, on peut envisager plusieurs directions :

- initialiser le réseau neuronal avec les coefficients factoriels pour une convergence plus rapide. En effet les calculs conduisant aux formules de régression sont beaucoup plus rapides que les calculs d'apprentissage du réseau.
- utiliser l'examen de résultats factoriels (% d'inertie des valeurs propres) pour résoudre un problème fondamental des réseaux neuronaux avec rétro-propagation : celui de la définition de la taille de la couche intermédiaire.

## Références bibliographiques

- JP Benzecri, (1977), Analyse discriminante et analyse factorielle, Les cahiers de l'analyse des données, vol II, 1977, n° 4, pages 369-406.
- JP Benzecri, (1973) L'analyse des données, Tome 1 : La taxinomie, Tome 2 : L'analyse des correspondances, Dunod, Paris 1973.
- L Bochereau, P Bourguine, H Epesse Priso, (1990) Generalist vs Specialist Neural Networks, Cognitiva 1990, Madrid, 20-23/11/1990.
- P Cazes, (1976) Régression par boules et par l'analyse des correspondances, Revue de statistique appliquée, Paris, 1976, vol XXIV, n° 4, pages 5-22.
- SE Feldman, DH Ballard, (1982) Connexionist models and their properties, Cognitive Science, 1982, N° 6, pages 205-254.
- P Gallinari, S Thiria, F Badran, F Fogelman-Soulie, (1990), On the relations between discriminant analysis and multi-layer perceptrons, Université Paris-Sud, Laboratoire de recherche en informatique, Rapport de recherche n° 551, 31 pages.
- L. Lebart, A Morineau, N Tabard, (1977) Techniques de la description statistique, Dunod, Paris 1977.
- RP Lippman, (1987), An Introduction to Computing with Neural Nets, IEEE ASSP Magazine, April 1987, pages 4-22.
- C Mullon, (1984), Essais de reconstitution d'un tableau ternaire dont on connaît deux des marges binaires : exemple des reports de votes aux scrutins présidentiels, Les cahiers de l'analyse des données, vol IX, 1984, n° 4, pages 473-487.