

UNIWIN VERSION 9.7.0

RESEAUX DE NEURONES ARTIFICIELS

Révision : 02/09/2023

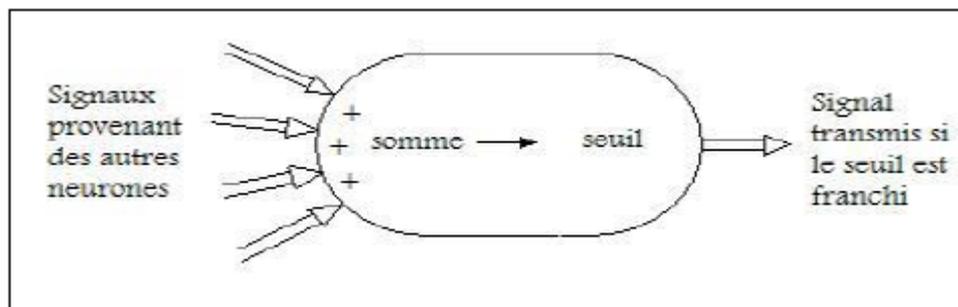
Définition.....	1
Entrée des données	3
Données manquantes	5
Exemple 1 : Fichier Ann (Classement)	5
L'option Rapports	9
L'option Graphiques	12
Exemple 2 : Fichier Iris3 (Classement)	15
L'option Rapports	18
L'option Graphiques	20
Exemple 3 : Fichier Wines2 (Classement).....	21
L'option Rapports	23
L'option Graphiques	23
Exemple 4 : Fichier Boston (Régression)	24
L'option Rapports	26
L'option Graphiques	27
Exemple 5 : Fichier Abalone (Régression)	30
L'option Rapports	32
L'option Graphiques	33
Exemple 6 : Fichier Titanic (Classement)	36
Les variables internes créées par la procédure	39
Références	39

Définition

Inspirés des neurones du cerveau, les réseaux de neurones artificiels n'ont rien de biologique. Ce sont des fonctions mathématiques à plusieurs paramètres. L'analogie date des premiers automates proposés en 1943 par Warren McCulloch et Walter Pitts.

Comme dans les neurones du cerveau où des connexions se créent, disparaissent ou se renforcent en fonction de différents stimuli et produisent une action, les réseaux de neurones artificiels ajustent des paramètres (appelés poids synaptiques en référence au fonctionnement biologique du cerveau) en fonction de données d'entrée afin de fournir la meilleure réponse possible.

En apprentissage machine, un neurone fait une combinaison linéaire des entrées qu'il reçoit, à laquelle il ajoute une valeur appelée biais. Une fonction non linéaire, dite d'activation (logistique ou tangente hyperbolique) est alors appliquée à la valeur de sortie. Cette valeur est ensuite transmise à la couche de neurones suivante si le seuil est franchi.



Chaque neurone effectue ainsi un calcul très rudimentaire et c'est la succession des couches de neurones qui permet d'obtenir des réseaux complexes. Durant cette phase dite d'apprentissage, le réseau va ajuster automatiquement les paramètres de chaque neurone, c'est-à-dire les valeurs des poids et du biais afin de minimiser l'erreur moyenne calculée sur l'ensemble des observations entre la sortie attendue et celle observée. L'hypothèse est qu'après cette phase d'apprentissage, le réseau sera capable de traiter de manière satisfaisante de nouvelles observations, dont la sortie est inconnue, en fonction de ce qu'il a appris. Dans un réseau de neurones à deux couches, la première couche est constituée d'un ensemble de neurones connectés en parallèle et fournissant un ensemble de sorties, elles-mêmes combinées pour devenir les entrées d'un nouvel ensemble de neurones formant une seconde couche.

Cette procédure est basée sur le package R 'neuralnet'.

Entrée des données

Cliquons sur l'icône ANN dans le ruban Expliquer.

La boîte de dialogue montrée ci-dessous s'affiche.

Réseaux de neurones artificiels

Variable à expliquer :

Variables explicatives :

(Libellé de la variable à expliquer :)

(Libellés des variables explicatives :)

(Libellés des observations :)

Neurones par couche cachée : 1

Seuil d'arrêt : 0,01

Nb. max. d'étapes : 100000

Nb. de répétitions : 1

Méthode

Classement Régression

Prétraitement des données

Aucun Centrage et réduction Min-Max

Algorithme : rprop+

Fonction d'erreur : Erreur quadratique

Fonction d'activation : Logistique

Sortie linéaire

Racine aléatoire : 280090412

Ok Annuler Sélection Supprimer Aide

Cette boîte de dialogue permet de préciser la variable à expliquer (quantitative ou qualitative alphanumérique), les variables explicatives quantitatives, les libellés optionnels de la variable à expliquer, des variables explicatives et des observations.

Elle permet également de préciser :

- Nombre de neurones pour chacune des couches cachées (ne pas renseigner s'il n'y a pas de couche cachée)
- Seuil d'arrêt pour les valeurs absolues des dérivées partielles de la fonction d'erreur
- Nombre maximal d'étapes pour l'apprentissage du réseau
- Nombre de répétitions pour l'apprentissage du réseau
- Méthode : classement ou régression
- Normalisation des données : aucune, centrage et réduction, min-max
- Algorithmes mis en œuvre :
 - rprop+ (rétropropagation résiliente avec retour arrière du poids)
 - rprop- (rétropropagation résiliente sans retour arrière du poids)
 - backprop (rétropropagation classique)
 - sag (plus petite dérivée en valeur absolue)
 - slr (plus petite vitesse d'apprentissage)
- Fonction d'erreur : erreur quadratique (si régression) ou entropie croisée (si classement ou régression)

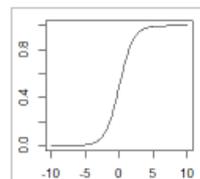
$$E = \frac{1}{2} \sum_{l=1}^L \sum_{h=1}^H (o_{lh} - y_{lh})^2$$

$$E = - \sum_{l=1}^L \sum_{h=1}^H (y_{lh} \log(o_{lh}) + (1 - y_{lh}) \log(1 - o_{lh}))$$

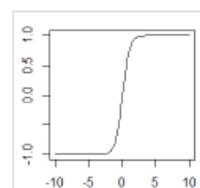
où o et y indiquent les sorties observées et prévues, $l = 1, \dots, L$ est l'indice des observations et $h = 1, \dots, H$ est l'indice des nœuds de la couche de sortie

- Fonction d'activation (sigmoïde) : logistique ou tangente hyperbolique

$$u = \frac{1}{1 + e^{-x}}$$



$$u = \frac{e^{2x} - 1}{e^{2x} + 1}$$



- Sortie linéaire : si coché la fonction d'activation n'est pas utilisée pour les neurones de sortie, si non coché, elle l'est
- Racine aléatoire pour déterminer les poids initiaux du réseau

Données manquantes

Les données manquantes ne sont pas autorisées pour les variables explicatives. Elles le sont pour la variable à expliquer. Les lignes pour lesquelles cette variable est manquante sont utilisées comme jeu de prévision.

Exemple 1 : Fichier Ann (Classement)

Pour ce premier exemple « académique », nous utiliserons le fichier Ann. Ce fichier contient 2000 observations décrites par trois variables X1, X2 et Y.

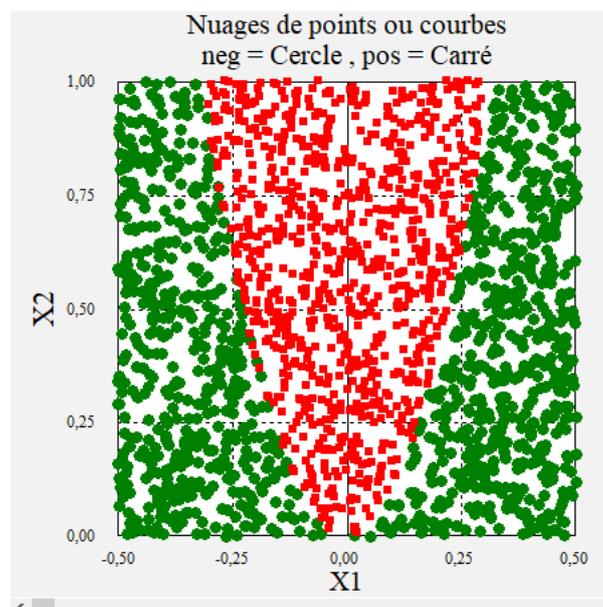
X1 et X2 sont deux variables aléatoires uniformes définies respectivement sur les intervalles $[-0,5 ; 0,5]$ et $[0 ; 1]$.

Y est une variable à deux modalités prenant les valeurs 'négatif' et 'positif' selon la formule bien évidemment inconnue :

si $(0,1 * X2 > X1^2)$ alors Y = 'positif' sinon Y = 'négatif'.

La variable TYPE précise le jeu d'appartenance de chaque observation : 'A' pour le jeu d'apprentissage, 'V' pour le jeu de validation et 'P' pour le jeu de prévision.

Visualisons les données dans un nuage de points codifiés.



Cliquons sur l'icône ANN dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après :

Réseaux de neurones artificiels

X1
X2
Y
Yn
TYPE

Variable à expliquer : Y

Variables explicatives : X1
X2

(Libellé de la variable à expliquer :)

(Libellés des variables explicatives :)

(Libellés des observations :)

Neurons par couche cachée : 2

Seuil d'arrêt : 0,01

Nb. max. d'étapes : 100000

Nb. de répétitions : 1

Méthode
 Classement Régression

Prétraitement des données
 Aucun
 Centrage et réduction
 Min-Max

Algorithme : rprop+

Fonction d'erreur : Erreur quadratique

Fonction d'activation : Logistique

Sortie linéaire

Racine aléatoire : 12345

Ok Annuler Sélection Supprimer Aide

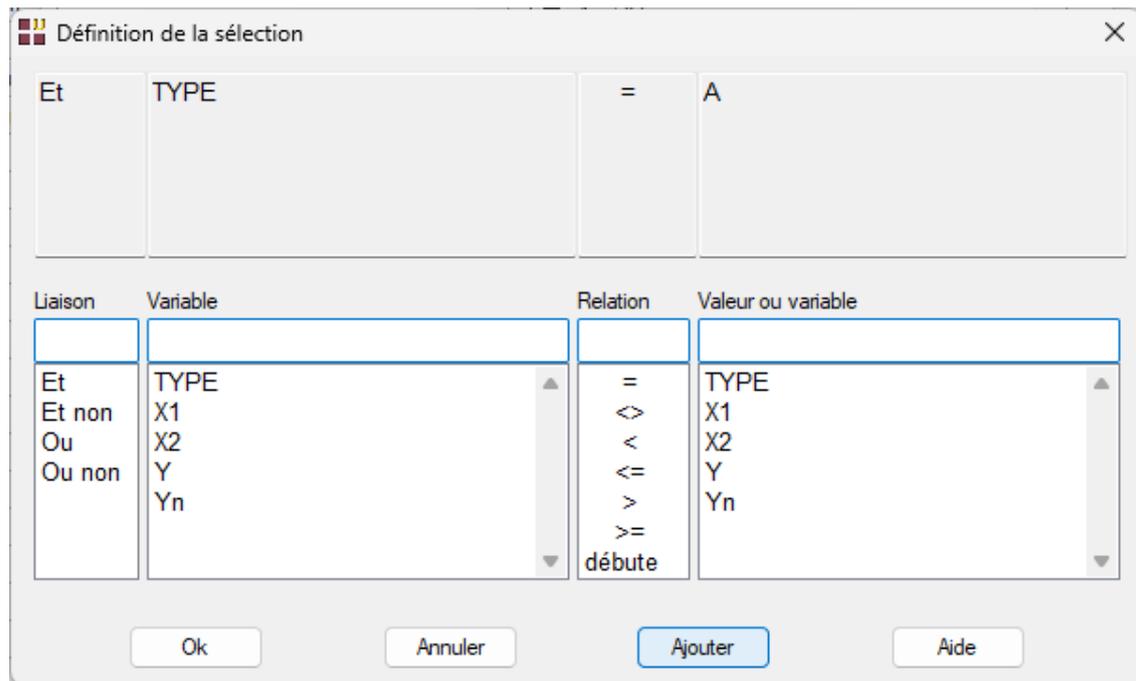
Cliquons sur le bouton 'Sélection' pour définir les données du jeu d'apprentissage et utilisons la colonne 'TYPE' du fichier des données pour cela.

Cette sélection comporte 1596 des 2000 observations :

Définition de la sélection

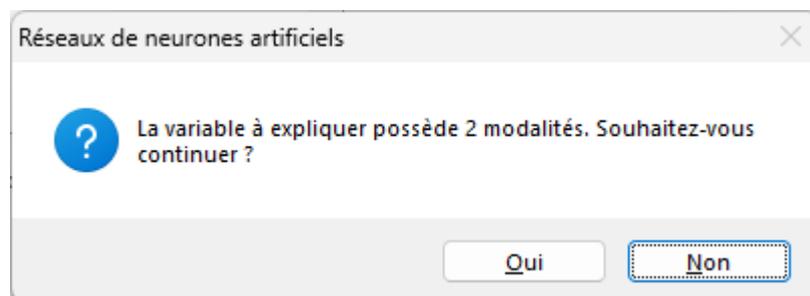
La sélection définie comporte 1596 éléments.

OK

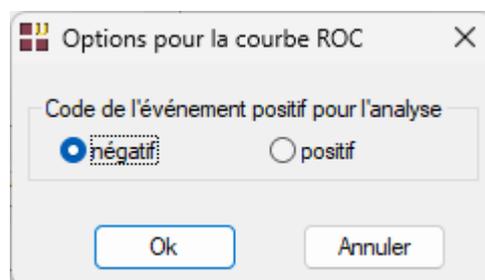


Ces 1596 observations définissent le jeu d'apprentissage, 399 autres observations seront utilisées comme jeu de validation et 5 comme jeu de prévision.

Cliquons sur Ok. Le message informatif suivant s'affiche alors :



Puisque la variable à expliquer possède 2 modalités, il est ensuite demandé de préciser le code de l'événement positif pour l'analyse.

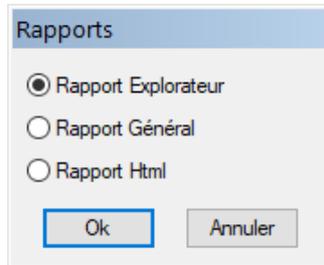


Cliquons sur Ok. UNIWIN débute les calculs.

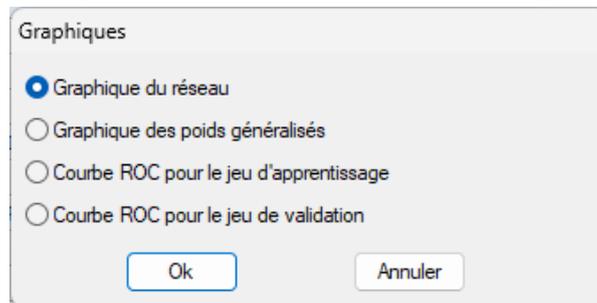
Après quelques instants, la fenêtre Rapports et Graphiques s'affiche.

La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

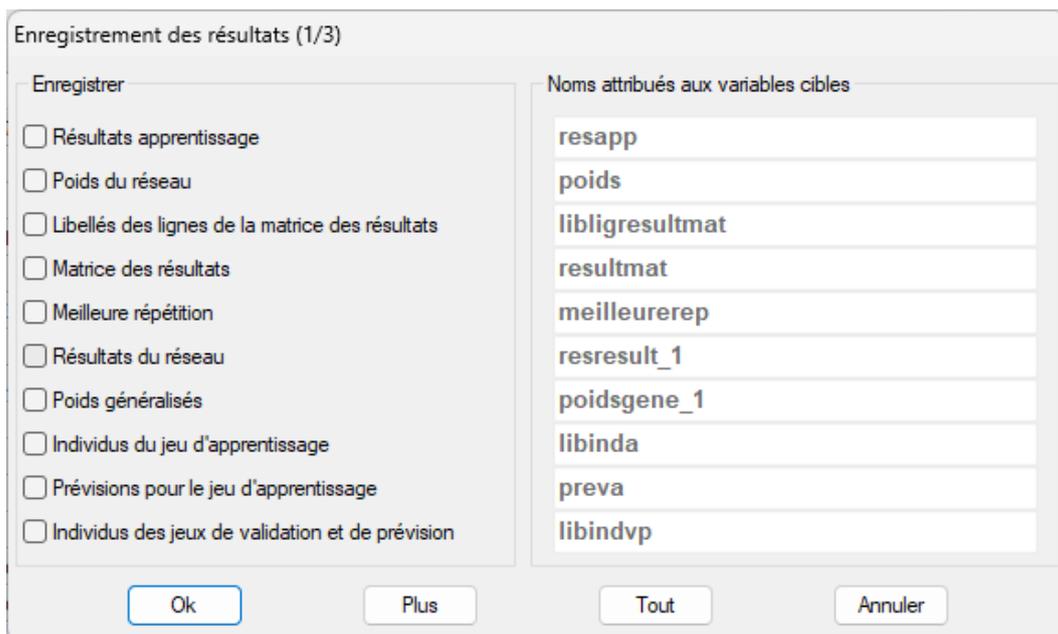
L'icône « Rapports »  affiche la boîte de dialogue des options pour les rapports.



L'icône « Graphiques »  affiche la boîte de dialogue des options pour les graphiques.



L'icône 'Enregistrer'  permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.

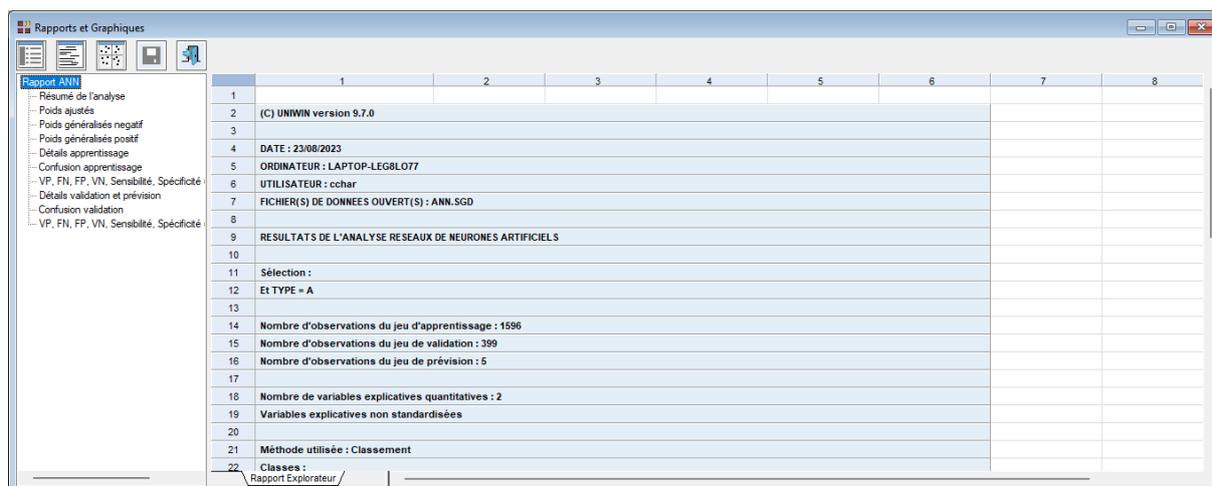


L'icône « Quitter »  permet de quitter l'analyse.

L'option Rapports

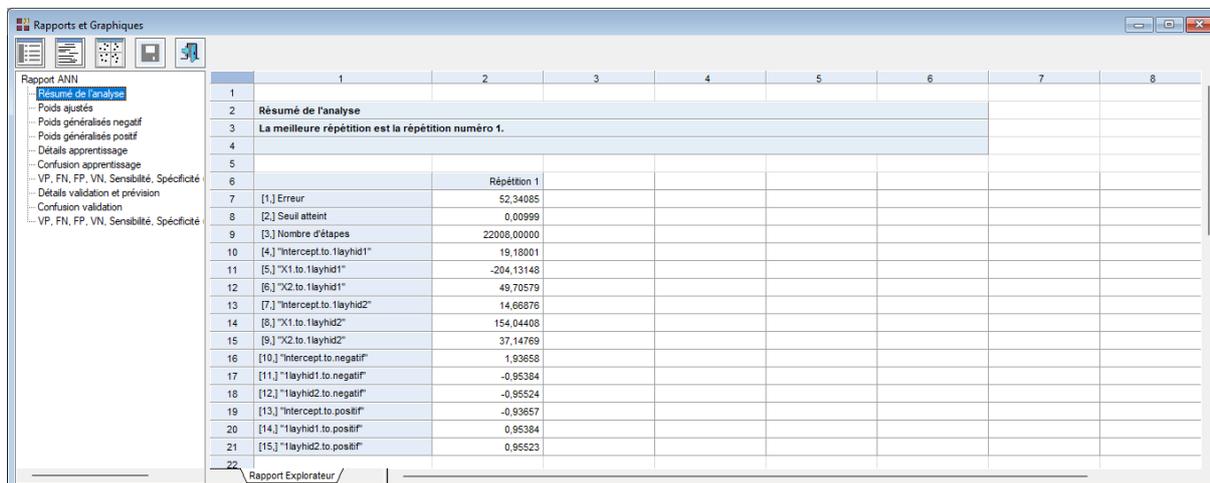
Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableau ou au format HTML.

Le premier tableau rappelle les paramètres et options de l'analyse :



	1	2	3	4	5	6	7	8
1								
2	(C) UNIWIN version 9.7.0							
3								
4	DATE : 23/08/2023							
5	ORDINATEUR : LAPTOP-LEG8LO77							
6	UTILISATEUR : cchar							
7	FICHIER(S) DE DONNEES OUVERT(S) : ANN.SGD							
8								
9	RESULTATS DE L'ANALYSE RESEAUX DE NEURONES ARTIFICIELS							
10								
11	Sélection :							
12	ET TYPE = A							
13								
14	Nombre d'observations du jeu d'apprentissage : 1596							
15	Nombre d'observations du jeu de validation : 399							
16	Nombre d'observations du jeu de prévision : 5							
17								
18	Nombre de variables explicatives quantitatives : 2							
19	Variables explicatives non standardisées							
20								
21	Méthode utilisée : Classement							
22	Classes : Rapport Explorateur /							

Le deuxième tableau affiche un résumé de l'analyse :



	1	2	3	4	5	6	7	8
1								
2	Résumé de l'analyse							
3	La meilleure répétition est la répétition numéro 1.							
4								
5								
6		Répétition 1						
7	[1.]	Erreur	52,34085					
8	[2.]	Seuil atteint	0,00999					
9	[3.]	Nombre d'étapes	22008,00000					
10	[4.]	"Intercept.to.1layhid1"	19,18001					
11	[5.]	"X1.to.1layhid1"	-204,13148					
12	[6.]	"X2.to.1layhid1"	49,70579					
13	[7.]	"Intercept.to.1layhid2"	14,66876					
14	[8.]	"X1.to.1layhid2"	154,04408					
15	[9.]	"X2.to.1layhid2"	37,14769					
16	[10.]	"Intercept.to.négatif"	1,93658					
17	[11.]	"1layhid1.to.négatif"	-0,95384					
18	[12.]	"1layhid2.to.négatif"	-0,95524					
19	[13.]	"Intercept.to.postif"	-0,93657					
20	[14.]	"1layhid1.to.postif"	0,95384					
21	[15.]	"1layhid2.to.postif"	0,95523					
22	Rapport Explorateur /							

Il indique pour chaque répétition :

- l'erreur du réseau
- le seuil d'arrêt atteint
- le nombre d'étapes
- les poids synaptiques ajustés du réseau

Le troisième tableau affiche les poids synaptiques ajustés pour chaque répétition.

	1	2	3	4	5	6	7	8
1								
2	Poids ajustés							
3	Cette liste affiche, pour chaque répétition, les poids synaptiques ajustés							
4	du réseau pour chaque couche et chaque noeud.							
5	Ces poids sont également affichés dans le graphique du réseau :							
6	- les lignes noires affichent les poids des liens entre les noeuds.							
7	- les lignes bleues affichent les biais ajoutés à chaque étape.							
8								
9	[[1]]							
10	[[1]][1]]							
11	[1] [2]							
12	[1]	19.18001	14.68876					
13	[2]	-204.13148	154.04408					
14	[3]	49.70579	37.14769					
15	[[1]][2]]							
16	[1] [2]							
17	[1]	1.9365784	-0.9365714					
18	[2]	-0.9536402	0.9536446					
19	[3]	-0.9552446	0.9552315					
20								
21								
22								

Les quatrième et cinquième tableaux affichent les poids généralisés de la meilleure répétition pour chacune des classes, dans notre exemple 'négalif' et 'positif'. Ces poids indiquent l'effet de chacune des variables explicatives sur le réseau.

	1	2	3	4	5	6	7	8
1								
2	Poids généralisés							
3	Ce tableau affiche pour chaque variable explicative les poids généralisés pour la meilleure répétition (1).							
4	Ces poids indiquent l'effet de chacune des variables explicatives sur le réseau.							
5	Variable à expliquer : négatif							
6								
7								
8								
9			X1	X2				
10	o6		0,00000	0,00000				
11	o7		0,00000	0,00000				
12	o8		0,00000	0,00000				
13	o10		0,00000	0,00000				
14	o12		0,00000	0,00000				
15	o13		0,00000	0,00000				
16	o14		0,00000	0,00000				
17	o15		-67,46793	-16,26988				
18	o16		0,00000	0,00000				
19	o17		-0,22847	-0,05461				
20	o18		-0,00154	-0,00037				
21	o19		0,00000	0,00000				
22	o21		0,00000	0,00000				

Le tableau 'Détails apprentissage' affiche pour la meilleure répétition et pour le jeu d'apprentissage les affectations aux classes de la variable à expliquer.

	1	2	3	4	5	6	7	8
1								
2	Résultats pour le jeu d'apprentissage							
3	Ce tableau affiche, pour la meilleure répétition (1) et pour le jeu d'apprentissage,							
4	les affectations aux classes de la variable à expliquer.							
5	(*) = observation mal classée.							
6								
7								
8			negatif	positif				
9	o6	- Observé : négatif - Prévu : négatif	0,98274	0,01727				
10	o7	- Observé : négatif - Prévu : négatif	0,98133	0,01866				
11	o8	- Observé : négatif - Prévu : négatif	0,98274	0,01727				
12	o10	- Observé : négatif - Prévu : négatif	0,98133	0,01866				
13	o12	- Observé : positif - Prévu : positif	0,02749	0,97250				
14	o13	- Observé : positif - Prévu : positif	0,02749	0,97250				
15	o14	- Observé : positif - Prévu : positif	0,02749	0,97250				
16	o15	- Observé : négatif - Prévu : négatif	0,96969	0,03033				
17	o16	- Observé : positif - Prévu : positif	0,02749	0,97250				
18	o17	- Observé : négatif - Prévu : négatif	0,98271	0,01730				
19	o18	- Observé : négatif - Prévu : négatif	0,98274	0,01727				
20	o19	- Observé : négatif - Prévu : négatif	0,98133	0,01866				
21	o21	- Observé : négatif - Prévu : négatif	0,98274	0,01727				
22	o23	- Observé : positif - Prévu : positif	0,02749	0,97250				

Le tableau 'Confusion apprentissage' affiche la matrice de confusion pour le jeu d'apprentissage pour la meilleure répétition. Le réseau classe correctement 96% des données d'apprentissage.

	1	2	3	4	5	6	7	8
1								
2	Matrice de confusion pour le jeu d'apprentissage							
3	Ce tableau affiche pour la meilleure répétition (1) les nombres d'observations observées							
4	par classe en lignes et les nombres d'observations prévues par classe en colonnes.							
5								
6	Pourcentage de mal classés : 3,759 %							
7	Pourcentage de bien classés : 96,241 %							
8								
9								
10	Observé \ Prévu	negatif	positif	Total				
11	negatif	848	32	880				
12	positif	28	688	716				
13	Total	876	720	1596				

Le tableau 'VP, FN, FP, VN, Sensibilité, Spécificité' affiche les informations permettant le calcul de la courbe ROC pour le jeu d'apprentissage.

	1	2	3	4	5	6	7	8
1								
2	VP, FN, FP, VN, SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE							
3								
4	Mesures = Mesures uniques							
5	VP = Nombres de vrais positifs							
6	FN = Nombres de faux négatifs							
7	FP = Nombres de faux positifs							
8	VN = Nombres de vrais négatifs							
9	Sensibilité en %							
10	Spécificité en %							
11								
12	Code de l'événement positif : negatif							
13	Aire sous la courbe (AUC) = 0,979							
14								
15								
16		Mesures	VP	FN	FP	VN	Sensibilité	Spécificité
17	1	Infini	880	0	716	0	100,0000	0,0000
18	2	0,02749	880	0	716	0	100,0000	0,0000
19	3	0,02749	874	6	546	170	99,31818	23,7430
20	4	0,02749	874	6	544	172	99,31818	24,0223
21	5	0,02749	874	6	542	174	99,31818	24,3016
22	6	0,02749	874	6	538	178	99,31818	24,8603

Le tableau 'Détails validation et prévision' affiche pour les jeux de validation et de prévision les classes observées et prévues de la variable à expliquer.

	1	2	3	4	5	6	7	8
1								
2	Résultats pour les jeux de validation et de prévision							
3	Ce tableau affiche, pour la meilleure répétition (1) et pour les jeux de validation et de prévision,							
4	les affectations aux classes de la variable à expliquer.							
5	(*) = observation mal classée.							
6								
7								
8		negatif	positif					
9	o1 - Observé : ? - Prévu : negatif	0,98274	0,01727					
10	o2 - Observé : ? - Prévu : negatif	0,98133	0,01866					
11	o3 - Observé : ? - Prévu : negatif	0,98133	0,01866					
12	o4 - Observé : ? - Prévu : positif	0,02751	0,97248					
13	o5 - Observé : ? - Prévu : positif	0,02749	0,97250					
14	o9 - Observé : negatif - Prévu : negatif	0,98133	0,01866					
15	o11 - Observé : negatif - Prévu : negatif	0,97948	0,02053					
16	o20 - Observé : positif - Prévu : positif	0,02749	0,97250					
17	o22 - Observé : negatif - Prévu : negatif	0,98274	0,01727					
18	o25 - Observé : positif - Prévu : positif	0,02749	0,97250					
19	o27 - Observé : negatif - Prévu : negatif	0,98133	0,01866					
20	o28 - Observé : positif - Prévu : positif	0,02749	0,97250					
21	o38 - Observé : negatif - Prévu : negatif	0,98133	0,01866					
22	o40 - Observé : positif - Prévu : positif	0,02749	0,97250					

Le tableau 'Confusion validation' affiche pour la meilleure répétition la matrice de confusion pour le jeu de validation. Le réseau classe correctement 95% des données validation.

Observé \ Prévu	negatif	positif	Total
negatif	219	9	228
positif	10	161	171
Total	229	170	399

Pourcentage de mal classés : 4,762 %
 Pourcentage de bien classés : 95,238 %

Le tableau 'VP, FN, FP, VN, Sensibilité, Spécificité' affiche les informations permettant le calcul de la courbe ROC pour le jeu de validation.

Mesures	VP	FN	FP	VN	Sensibilité	Spécificité
1	Infini	228	0	171	0	100,00000
2	0,02749	228	0	171	0	100,00000
3	0,02749	225	3	141	30	98,68421
4	0,02749	225	3	138	33	98,68421
5	0,02749	225	3	137	34	98,68421
6	0,02749	225	3	136	35	98,68421

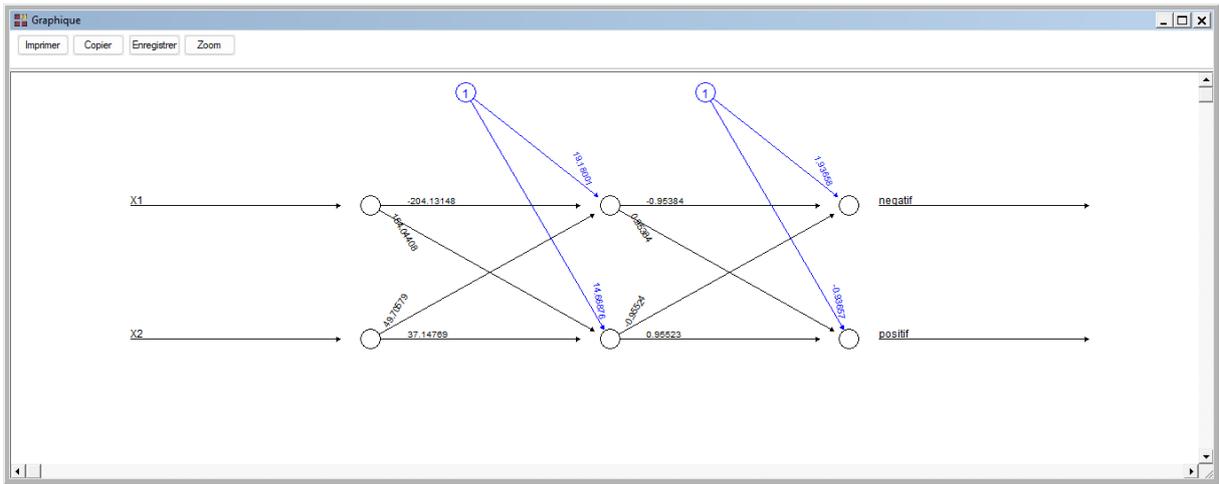
Aire sous la courbe (AUC) = 0,962

L'option Graphiques

Graphique du réseau

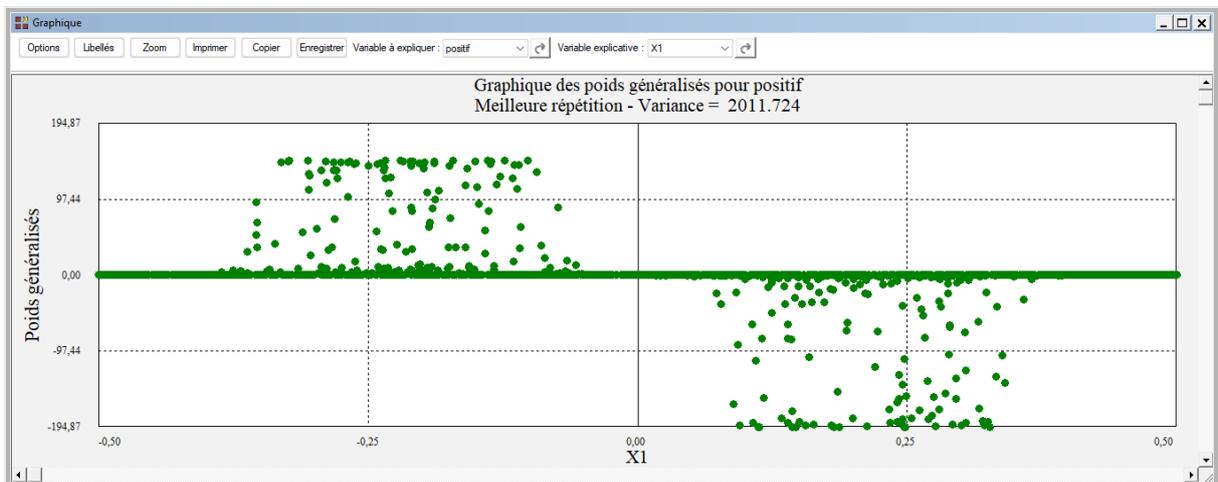
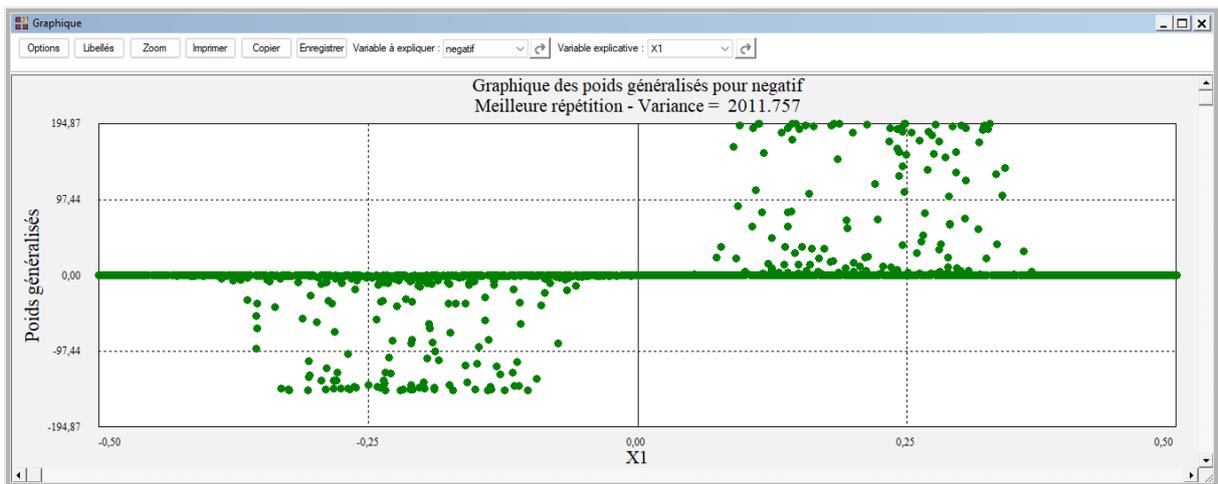
Ce graphique affiche la structure du réseau, les poids synaptiques (lignes noires) et les biais (lignes bleues) :

- Les 2 nœuds en entrée associés aux 2 variables explicatives X1 et X2
- Les 2 nœuds pour l'unique couche cachée demandée dans cette analyse
- Les 2 nœuds de sortie associés à chacune des 2 modalités de Y



Graphique des poids généralisés

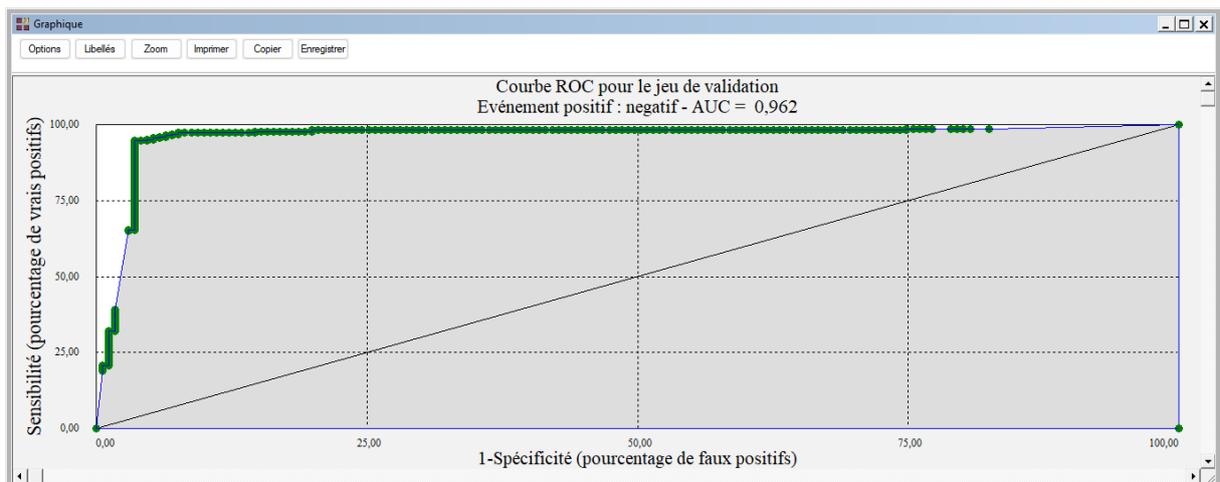
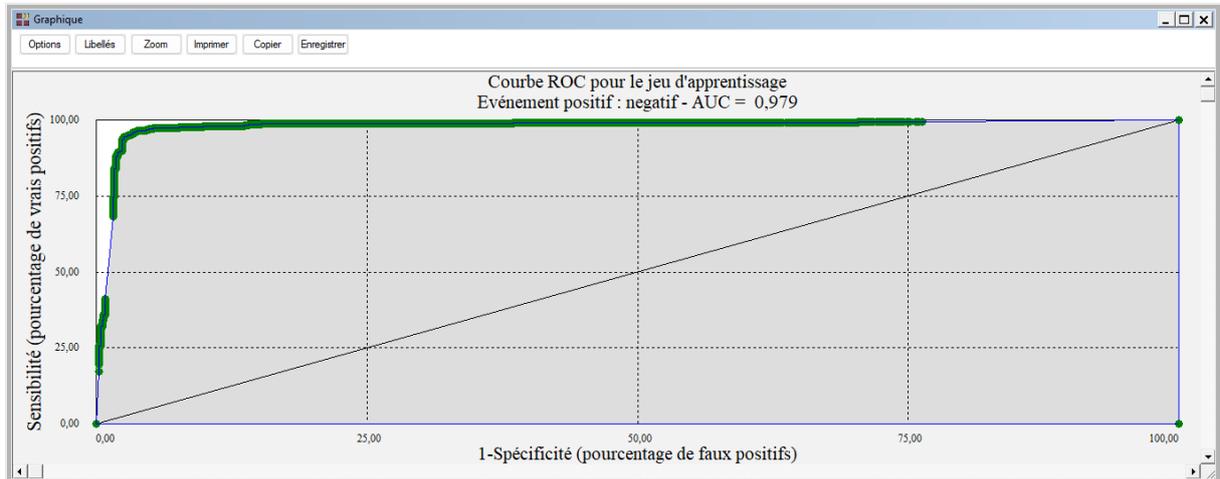
Les poids généralisés (Intrator et Intrator, 2001) indiquent l'effet de la variable explicative sur la variable à expliquer : effet linéaire lorsque la variance des poids est petite, effet non linéaire lorsque cette variance est grande et aucun effet si les poids se rassemblent autour de 0.



Les variables à représenter dans ce graphique peuvent être sélectionnées à partir de la barre d'outils. Tous les graphiques des poids généralisés utilisent la même échelle pour l'axe vertical des poids, ce qui permet une analyse aisée.

Courbes ROC

Les courbes ROC peuvent être affichées pour les jeux d'apprentissage et de validation.



Les aires sous les courbes indiquent que le modèle ajusté est performant.

En conclusion

Reprenons le graphique du réseau et utilisons les poids synaptiques reliant la couche d'entrée à la couche cachée :

	Neurone 1	Neurone 2
X1	-245,20	210,86
X2	58,59	50,02
Biais	23,75	20,62

On obtient ainsi les équations de deux droites :

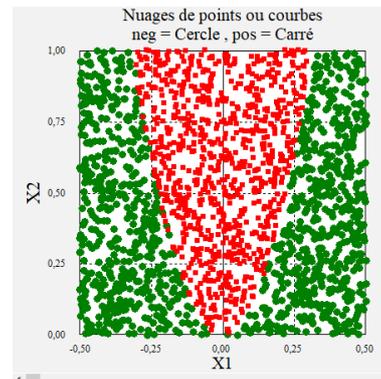
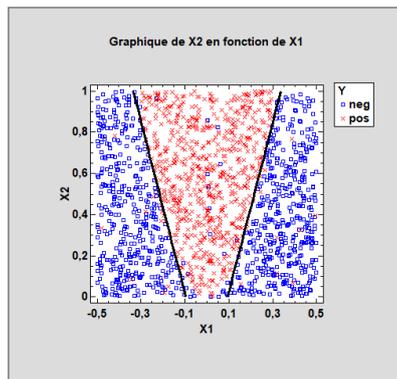
$$-245,20 X_1 + 58,59 X_2 + 23,75 = 0$$

$$X_2 = - 4,185 X_1 - 0,405$$

$$210,86 X_1 + 50,02 X_2 + 20,62 = 0$$

$$X_2 = 4,216 X_1 - 0,412$$

Affichons ces droites dans le graphique des données en utilisant STATGRAPHICS Centurion (à gauche) et le graphique des données (à droite).



Clairement, l'utilisation d'une couche cachée à deux neurones définissant les frontières entre 'négatif' et 'positif' a été efficace.

Exemple 2 : Fichier Iris3 (Classement)

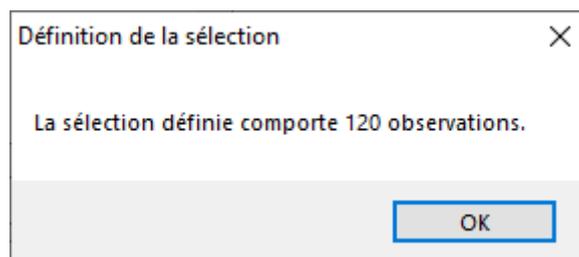
Pour ce deuxième exemple, ouvrons le fichier Iris3.

Ce fichier contient les données relatives à 150 iris de trois espèces: Iris Setosa, Iris Versicolor et Iris Virginica. Les mesures effectuées sont: longueur du sépale (lonsepal), longueur du pétale (lonpetal), largeur du sépale (larsepal), largeur du pétale (larpetal).

Cliquons sur l'icône ANN dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après.

Cliquons sur le bouton 'Sélection' pour définir les données du jeu d'apprentissage et utilisons la colonne 'type' du fichier des données pour cela.

Cette sélection comporte 120 des 150 observations :



Réseaux de neurones artificiels

type
lonsepal
larsepal
lonpetal
larpetal
codesp1
codesp2
numiris
mesures
nomesp

Variable à expliquer :
codesp2

Variables explicatives :
lonsepal
larsepal
lonpetal
larpetal

(Libellé de la variable à expliquer :)

(Libellés des variables explicatives :)

(Libellés des observations :)

Neurones par couche cachée : 3

Seuil d'arrêt : 0,01

Nb. max. d'étapes : 100000

Nb. de répétitions : 1

Méthode
 Classement Régression

Prétraitement des données
 Aucun
 Centrage et réduction
 Min-Max

Algorithme : prop+

Fonction d'erreur : Erreur quadratique

Fonction d'activation : Logistique

Sortie linéaire

Racine aléatoire : 321

Ok Annuler Sélection Supprimer Aide

Définition de la sélection

Et type = A

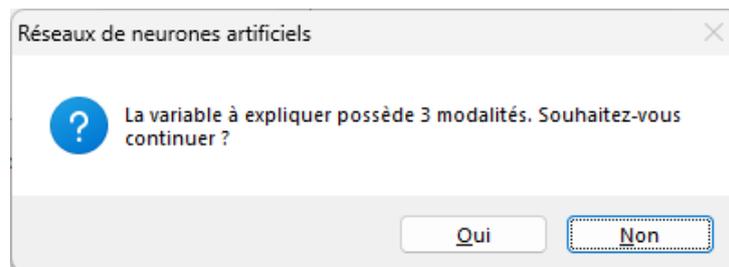
Liaison	Variable	Relation	Valeur ou variable
Et	larsepal	=	codesp1
Et non	lonsepal	<>	codesp2
Ou	lonsepal	<	larpetal
Ou non	mesures	<=	larsepal
	nomesp	>	lonpetal
	numiris	>=	lonsepal
	type	début	mesures

Ok Annuler Ajouter Aide

Ces 120 observations définissent le jeu d'apprentissage, les 30 autres observations seront utilisées comme jeu de validation.

S'il y a des données manquantes pour la variable à expliquer, ces observations définissent le jeu de prévision. Dans cet exemple, il y a 6 valeurs manquantes dans la colonne codesp2. Le jeu de validation est donc réduit à 24 observations.

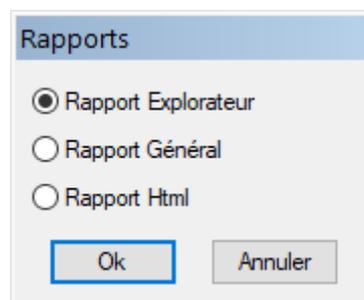
Cliquons sur OK. Le message informatif suivant s'affiche alors :



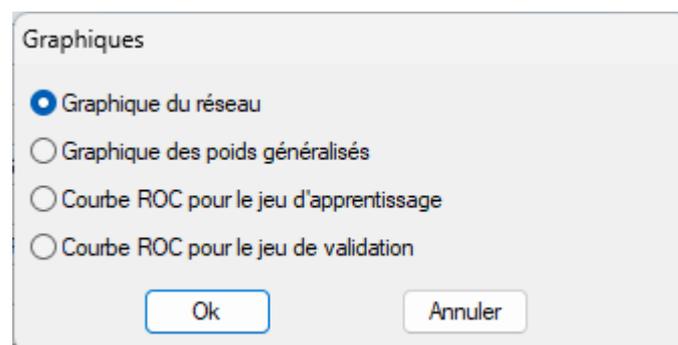
Cliquons sur Oui. UNIWIN débute les calculs. Après quelques instants, la fenêtre Rapports et Graphiques s'affiche.

La barre d'outils 'Rapports et Graphiques' permet par l'icône 'Données'  de rappeler la boîte de dialogue d'entrée des données.

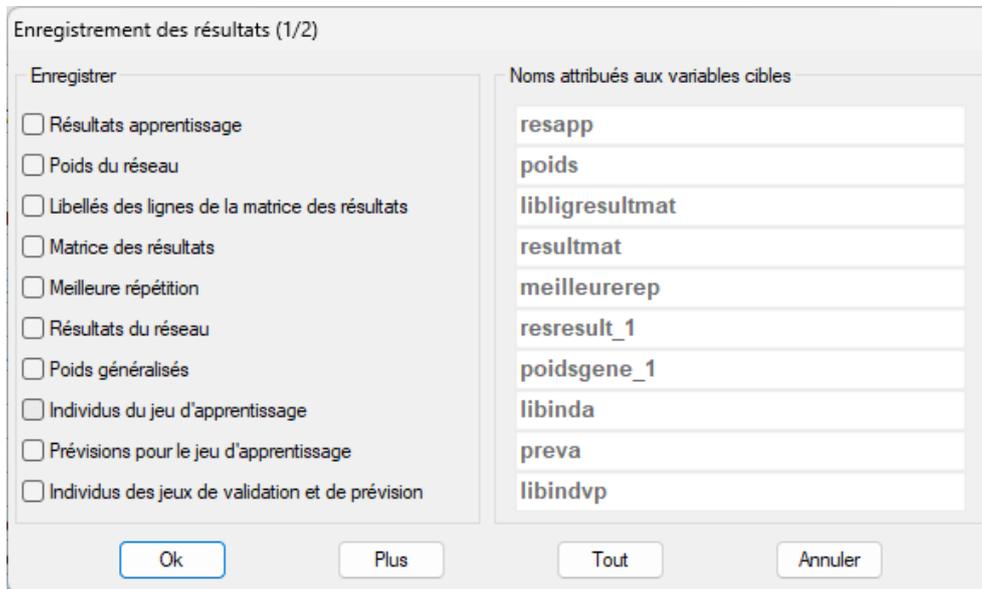
L'icône « Rapports »  affiche la boîte de dialogue des options pour les rapports.



L'icône « Graphiques »  affiche la boîte de dialogue des options pour les graphiques.



L'icône 'Enregistrer'  permet de sélectionner les résultats de l'analyse à enregistrer dans un fichier.



L'icône « Quitter »  permet de quitter l'analyse.

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML.

Voici trois exemples du rapport pour l'analyse ANN : Explorateur, Général, HTML.

Rapports et Graphiques

Rapport ANN

- Résumé de l'analyse
- Poids ajustés
- Poids généralisés Setosa
- Poids généralisés Versicolor
- Poids généralisés Virginica
- Détails apprentissage
- Confusion apprentissage**
- Détails validation et prévision
- Confusion validation

	1	2	3	4	5	6	7	8
1								
2	Matrice de confusion pour le jeu d'apprentissage							
3	Ce tableau affiche pour la meilleure répétition (1) les nombres d'observations observées							
4	par classe en lignes et les nombres d'observations prévues par classe en colonnes.							
5								
6	Pourcentage de mal classés : 0,000 %							
7	Pourcentage de bien classés : 100,000 %							
8								
9								
10	Observé \ Prévu	Setosa	Versicolor	Virginica	Total			
11	Setosa	43	0	0	43			
12	Versicolor	0	38	0	38			
13	Virginica	0	0	39	39			
14	Total	43	38	39	120			
15								
16								
17								
18								
19								
20								
21								
22								

Rapport Explorateur

Rapports et Graphiques

Résultats pour le jeu d'apprentissage

Ce tableau affiche, pour la meilleure répétition (1) et pour le jeu d'apprentissage, les affectations aux classes de la variable à expliquer.

(*) = observation mal classée.

	Setosa	Versicolor	Virginica
496 o1 - Observé : Setosa - Prévu : Setosa	1,00213	-0,00312	0,00443
497 o2 - Observé : Setosa - Prévu : Setosa	0,99615	0,00255	-0,00363
498 o4 - Observé : Setosa - Prévu : Setosa	0,99783	0,00301	-0,00427
499 o5 - Observé : Setosa - Prévu : Setosa	1,00278	-0,00404	0,00574
500 o6 - Observé : Setosa - Prévu : Setosa	0,99689	0,00149	-0,00212
501 o7 - Observé : Setosa - Prévu : Setosa	0,99678	0,00165	-0,00235
502 o8 - Observé : Setosa - Prévu : Setosa	1,00051	-0,00082	0,00116
503 o9 - Observé : Setosa - Prévu : Setosa	0,99688	0,00436	-0,00620
504 o11 - Observé : Setosa - Prévu : Setosa	1,00313	-0,00455	0,00646
505 o12 - Observé : Setosa - Prévu : Setosa	0,99564	0,00057	-0,00081
506 o13 - Observé : Setosa - Prévu : Setosa	1,00018	-0,00035	0,00049
507 o14 - Observé : Setosa - Prévu : Setosa	1,00202	-0,00296	0,00421
508 o15 - Observé : Setosa - Prévu : Setosa	1,00790	-0,01135	0,01612
509 o16 - Observé : Setosa - Prévu : Setosa	1,00445	-0,00643	0,00913

Rapport Général

Rapports et Graphiques

o144 - Observé : Virginica - Prévu : Virginica	-0,00857	-0,00286	0,99719
o145 - Observé : Virginica - Prévu : Virginica	-0,01188	0,00186	0,99049
o146 - Observé : Virginica - Prévu : Virginica	-0,00234	-0,01173	1,00979
o147 - Observé : Virginica - Prévu : Virginica	0,00693	-0,02441	1,02803
o148 - Observé : Virginica - Prévu : Virginica	0,00743	-0,02537	1,02928

Matrice de confusion pour le jeu d'apprentissage

par classe en lignes et les nombres d'observations prévues par classe en colonnes.

Pourcentage de mal classés : 0,000 %
 Pourcentage de bien classés : 100,000 %

Observé \ Prévu	Setosa	Versicolor	Virginica	Total
Setosa	43	0	0	43
Versicolor	0	38	0	38
Virginica	0	0	39	39
Total	43	38	39	120

Résultats pour les jeux de validation et de prévision

Ce rapport permet de visualiser les résultats de l'analyse :

- Résumé des paramètres de l'analyse
- Poids ajustés du réseau
- Poids généralisés pour chacune des espèces d'iris
- Résultats pour les données du jeu d'apprentissage
- Matrice de confusion pour le jeu d'apprentissage

et s'il y a des données pour les jeux de validation et de prévision :

- Résultats pour les jeux de validation et de prévision
- Matrice de confusion pour le jeu de validation

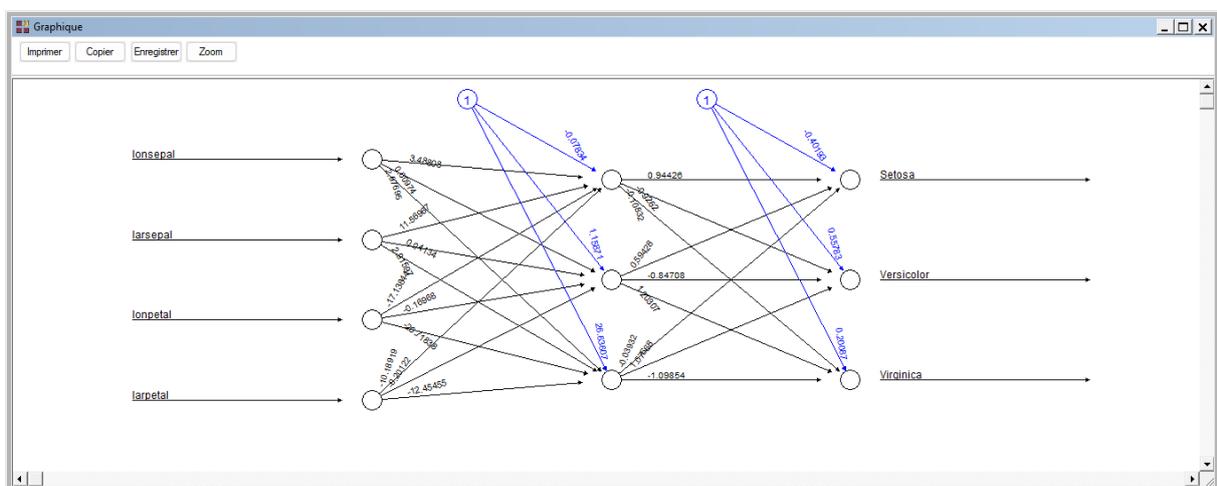
L'option Graphiques

Cette option permet d'obtenir deux graphiques pour l'analyse ANN.

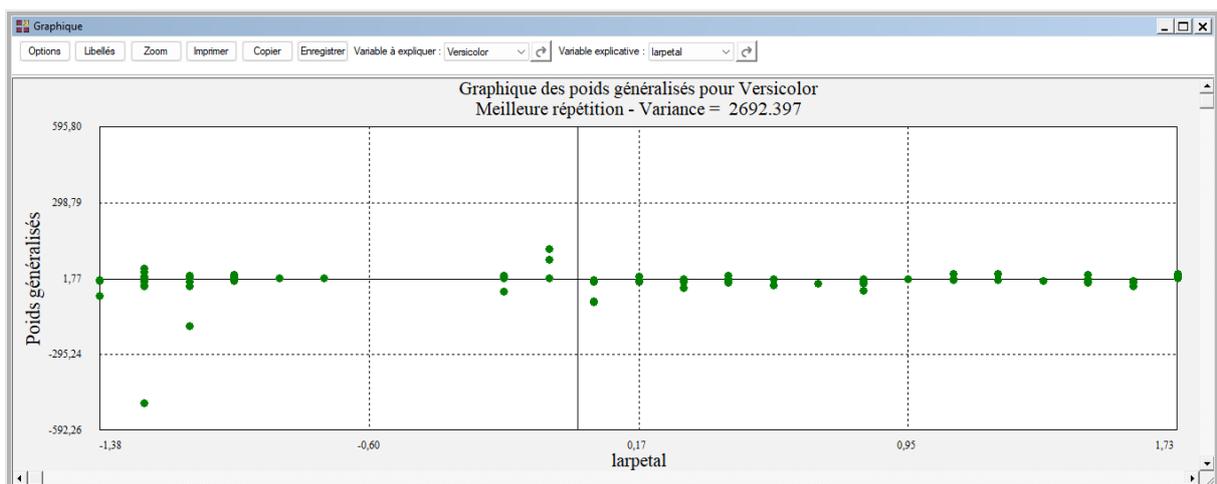
Graphique du réseau

Ce graphique affiche la structure du réseau, les poids synaptiques (lignes noires) et les biais (lignes bleues) :

- Les 4 nœuds en entrée associés aux 4 variables explicatives
- Les 3 nœuds pour l'unique couche cachée demandée dans cette analyse
- Les 3 nœuds de sortie associés à chacune des 3 espèces d'iris



Graphique des poids généralisés



Les poids généralisés (Intrator et Intrator, 2001) indiquent l'effet de la variable explicative sur la variable à expliquer : effet linéaire lorsque la variance des poids est petite, effet non linéaire lorsque cette variance est grande et aucun effet si les poids se rassemblent autour de 0.

Les variables à représenter dans ce graphique peuvent être sélectionnées à partir de la barre d'outils. Tous les graphiques des poids généralisés utilisent la même échelle pour l'axe vertical des poids, ce qui permet une analyse aisée.

Courbes ROC

Ces options ne sont pas disponibles car la variable à expliquer possède plus de 2 modalités.

Exemple 3 : Fichier Wines2 (Classement)

Les données sont le résultat de l'analyse chimique de 180 échantillons de vins issus de trois cultivars différents et provenant d'une même région en Italie.

(source : <https://archive.ics.uci.edu/ml/datasets/wine>).

Elles sont constituées des treize caractéristiques chimiques et spectroscopiques suivantes :

wine	N° du cultivar
alcohol	Alcool
malic acid	Acide malique
ash	Cendres
ash alkalinity	Alcalinité des cendres
magnesium	Magnésium
tot. phenols	Total des phénols
flavonoids	Flavonoïdes
non-flav. phenols	Phénols non-flavonoïdes
proanth	Proanthocyanidines
col. int.	Intensité de couleur
col. hue	Teinte de couleur
OD ratio	Rapport OD
proline	Proline

Cliquons sur l'icône ANN dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après.

Cliquons sur le bouton 'Sélection' pour définir les données du jeu d'apprentissage et utilisons la colonne 'Type' du fichier des données pour cela.

151 observations sont sélectionnées pour le jeu d'apprentissage.

Cliquons sur OK. Après quelques instants la fenêtre 'Rapports et Graphiques' s'affiche.

Réseaux de neurones artificiels

Type
Wine
 Alcohol
 Malic acid
 Ash
 Alcalinity of ash
 Magnesium
 Total phenols
 Flavanoids
 Nonflavanoid phenols
 Proanthocyanins
 Color intensity
 Hue
 OD ratio
 Proline
 LibVarY
 LibVarX
 LibObs
 LibCultivar

Neurones par couche cachée : 3

Seuil d'arrêt : 0,01

Nb. max. d'étapes : 100000

Nb. de répétitions : 1

Méthode
 Classement Régression

Prétraitement des données
 Aucun
 Centrage et réduction
 Min-Max

Variable à expliquer :
 Wine

Variables explicatives :
 Alcohol
 Malic acid
 Ash
 Alcalinity of ash
 Magnesium
 Total phenols
 Flavanoids
 Nonflavanoid phenols

(Libellé de la variable à expliquer :)

(Libellés des variables explicatives :)

(Libellés des observations :)

Algorithme : rprop+
 Fonction d'erreur : Erreur quadratique
 Fonction d'activation : Logistique
 Sortie linéaire
 Racine aléatoire : 321

Ok Annuler Sélection Supprimer Aide

Définition de la sélection

Et	Type	=	A
----	------	---	---

Liaison	Variable	Relation	Valeur ou variable
Et	Nonflavanoid phenols	=	Alcalinity of ash
Et non	OD ratio	<>	Alcohol
Ou	Proanthocyanins	<	Ash
Ou non	Proline	<=	Color intensity
	Total phenols	>	Flavanoids
	Type	>=	Hue
	Wine	débute	LibObs

Ok Annuler Ajouter Aide

L'option Rapports

Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML.

Voici un exemple de rapport au format Explorateur.

Rapports et Graphiques

Rapport ANN

- Résumé de l'analyse
- Poids ajoutés
- Poids généralisés M1
- Poids généralisés M2
- Poids généralisés M3
- Détails apprentissage
- Confusion apprentissage
- Détails validation et prévision
- Confusion validation

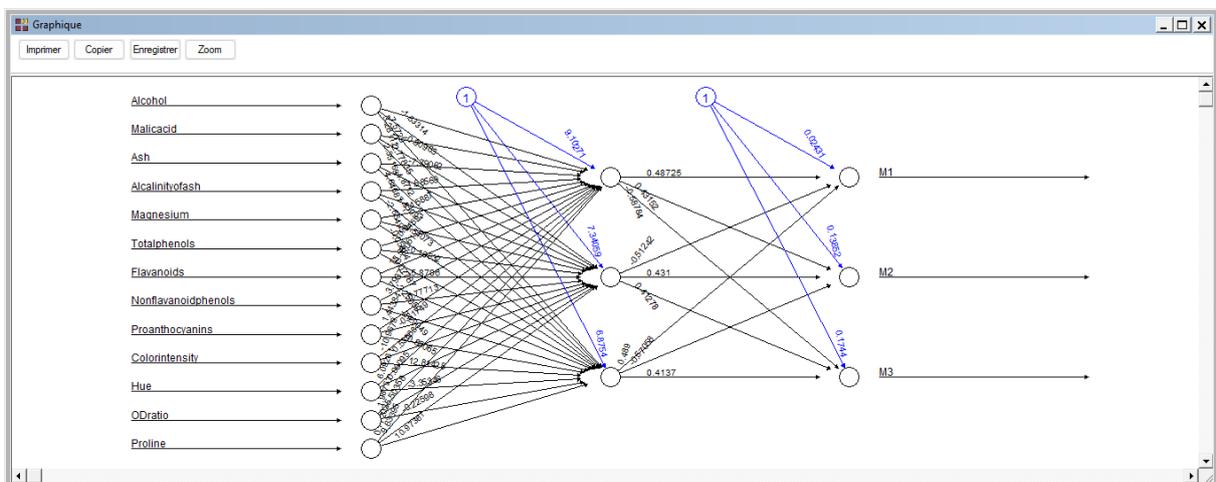
	1	2	3	4	5	6	7	8
1								
2	Résultats pour le jeu d'apprentissage							
3	Ce tableau affiche, pour la meilleure répétition (1) et pour le jeu d'apprentissage,							
4	les affectations aux classes de la variable à expliquer.							
5	(*) = observation mal classée.							
6								
7								
8			M1	M2	M3			
9	o1	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
10	o2	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
11	o3	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
12	o4	- Observé : M1 - Prévu : M1	1,00046	-0,00063	0,00038			
13	o5	- Observé : M1 - Prévu : M1	1,00019	-0,00013	0,00004			
14	o6	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
15	o7	- Observé : M1 - Prévu : M1	1,00055	-0,00055	0,00027			
16	o8	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
17	o9	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
18	o10	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
19	o11	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
20	o12	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
21	o13	- Observé : M1 - Prévu : M1	1,00056	-0,00054	0,00026			
22	o14	- Observé : M1 - Prévu : M1	1,00043	-0,00066	0,00042			

L'option Graphiques

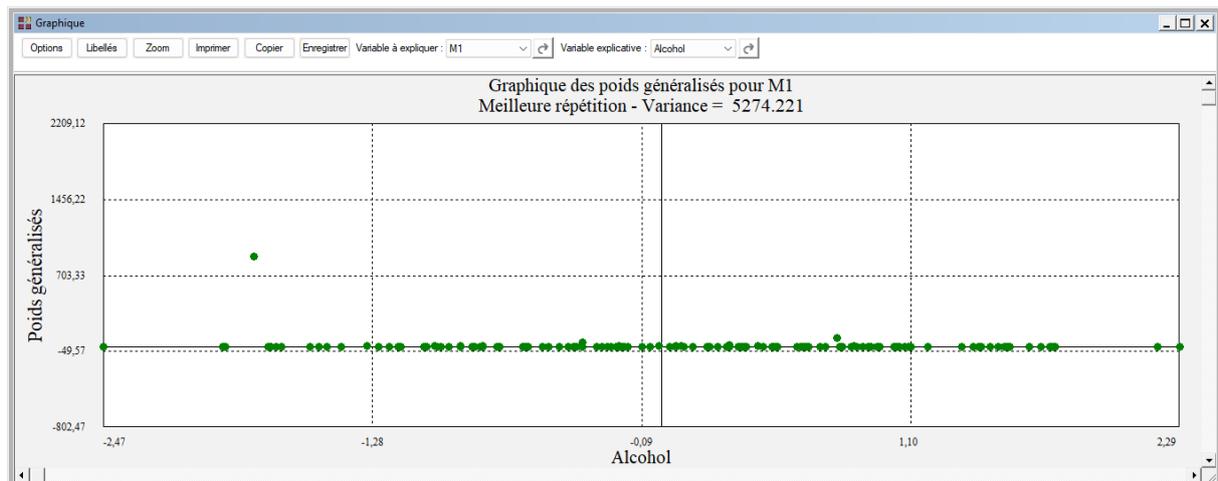
Graphique du réseau

Ce graphique affiche la structure du réseau, les poids synaptiques (lignes noires) et les biais (lignes bleues) :

- Les 13 nœuds en entrée associés aux 13 variables explicatives
- Les 3 nœuds pour l'unique couche cachée demandée dans cette analyse
- Les 3 nœuds de sortie associés à chacun des 3 cultivars



Graphique des poids généralisés



Exemple 4 : Fichier Boston (Régression)

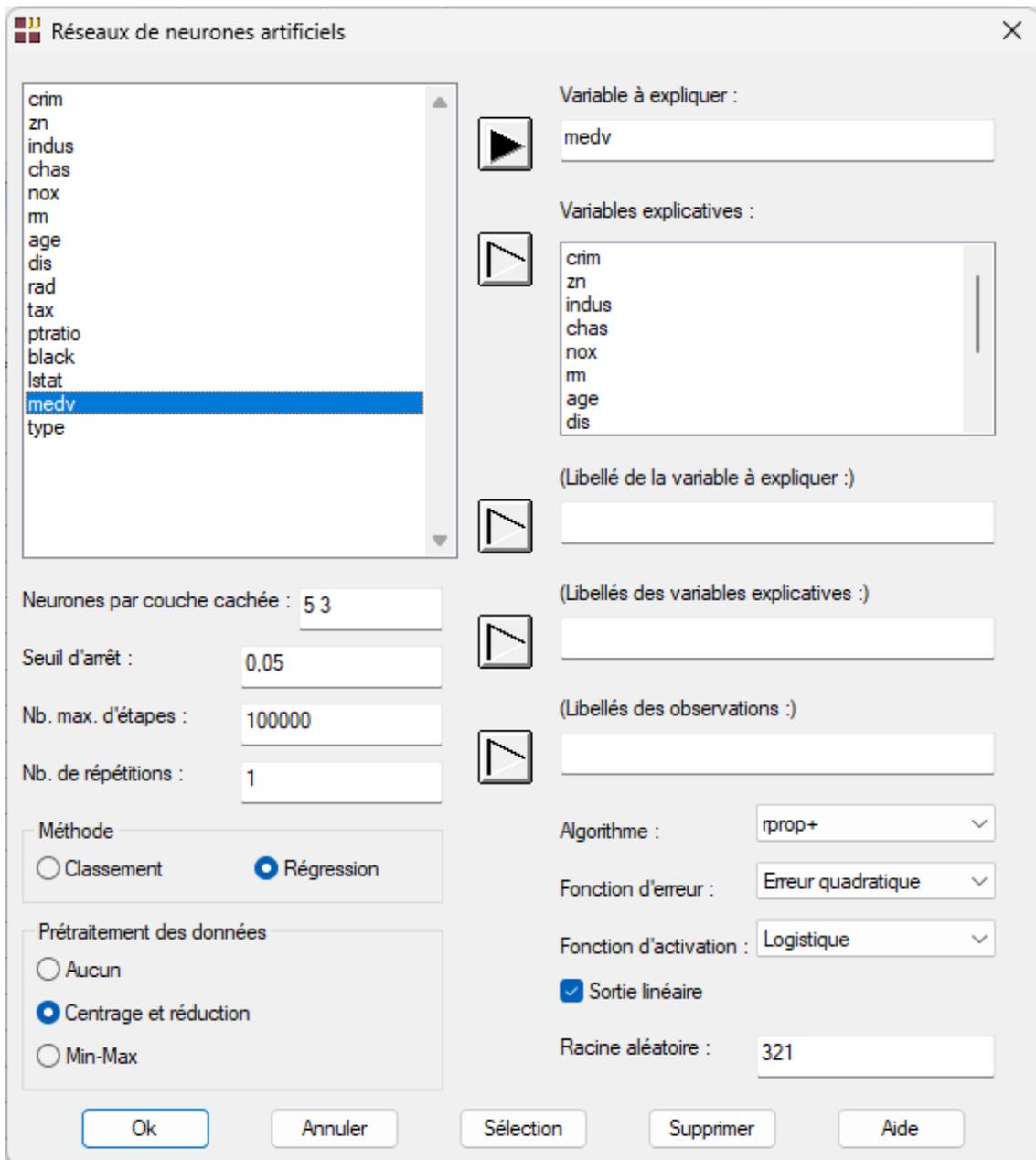
Ce fichier contient des informations collectées par le « U.S Census Service » concernant la valeur des habitations dans l'agglomération de Boston (source : <http://lib.stat.cmu.edu/datasets/boston>).

Les caractéristiques contenues dans le fichier sont les suivantes :

crim	taux de criminalité par habitant par ville
zn	proportion de terrains résidentiels zonés pour des lots de plus de 25 000 pieds carrés.
indus	proportion d'acres non commerciales par ville
chas	indicateur de proximité à la Charles River (1 = proche, 0 = sinon)
nox	concentration d'oxydes nitriques (partie pour 10 millions)
rm	nombre moyen de pièces par habitation
age	proportion des habitations construites avant 1940
dis	distance pondérée à cinq lieux d'emplois de Boston
rad	indice d'accessibilité aux autoroutes
tax	taux d'imposition foncière par \$ 10.000
ptratio	ratio élèves / enseignants par ville
black	$1000(Bk-0,63)^2$ où Bk = proportion de noirs par ville
lstat	% statut inférieur de la population
medv	valeur médiane de l'habitation (en milliers de \$)
type	indique si l'observation est utilisée pour l'apprentissage (A) ou la validation (V)

Le but est de prévoir la variable medv en utilisant ces données.

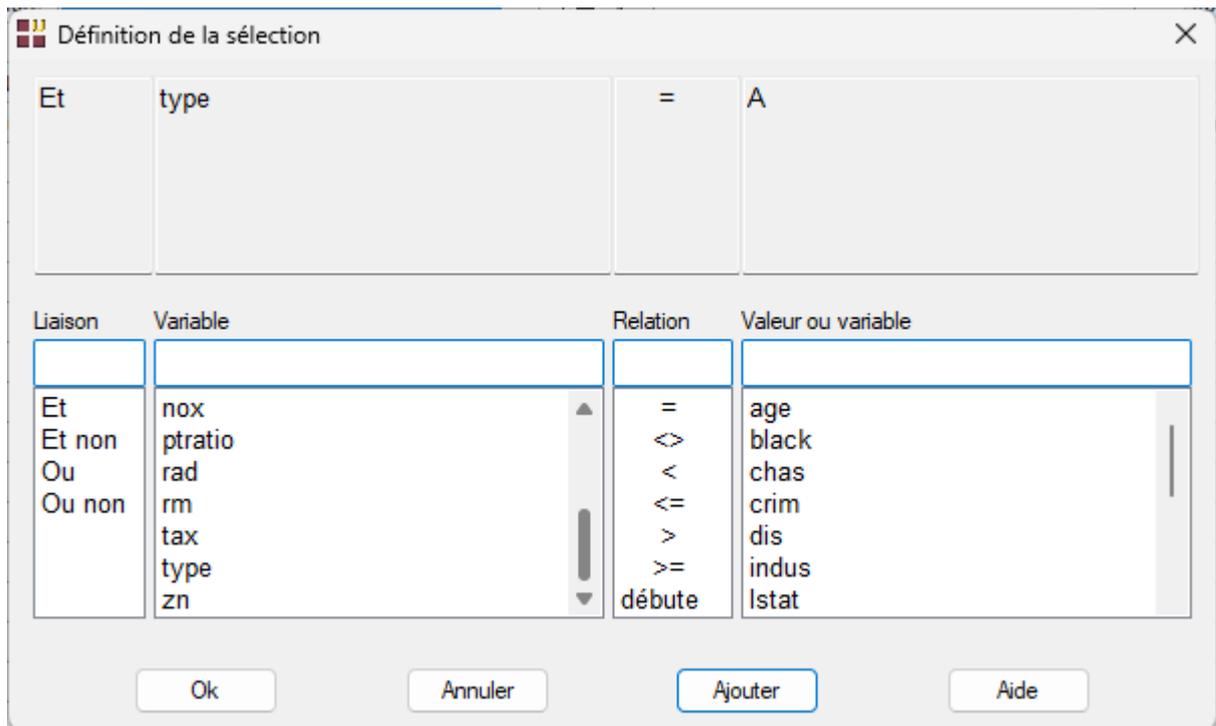
Cliquons sur l'icône ANN dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après.



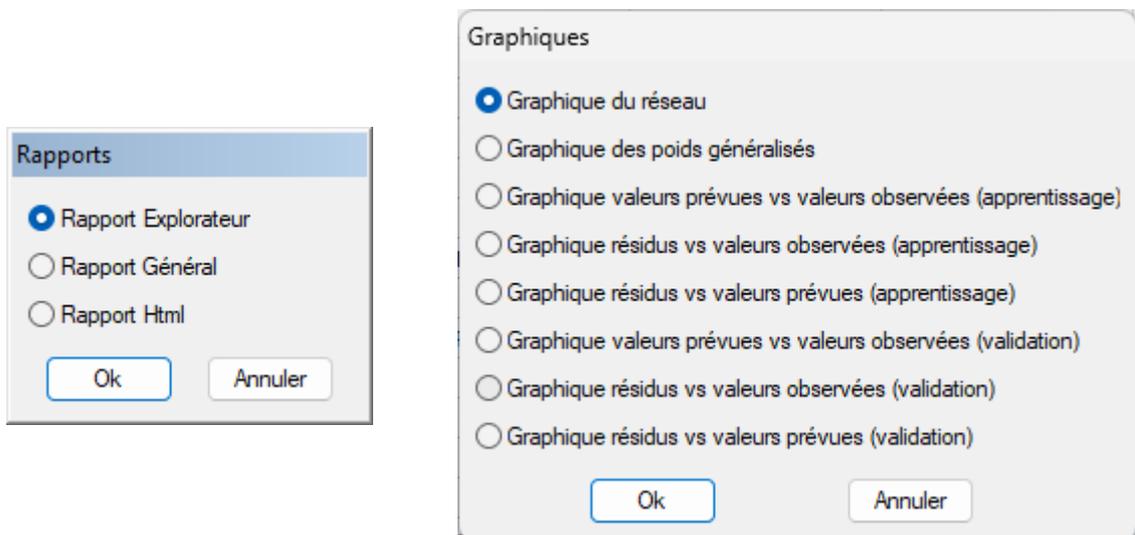
Cliquons sur le bouton 'Sélection' pour sélectionner les données du jeu d'apprentissage.

400 observations sont sélectionnées, les 106 observations restantes seront utilisées comme jeu de validation.

Cliquons sur Ok.



Après quelques instants la fenêtre 'Rapports et Graphiques' s'affiche.



L'option Rapports

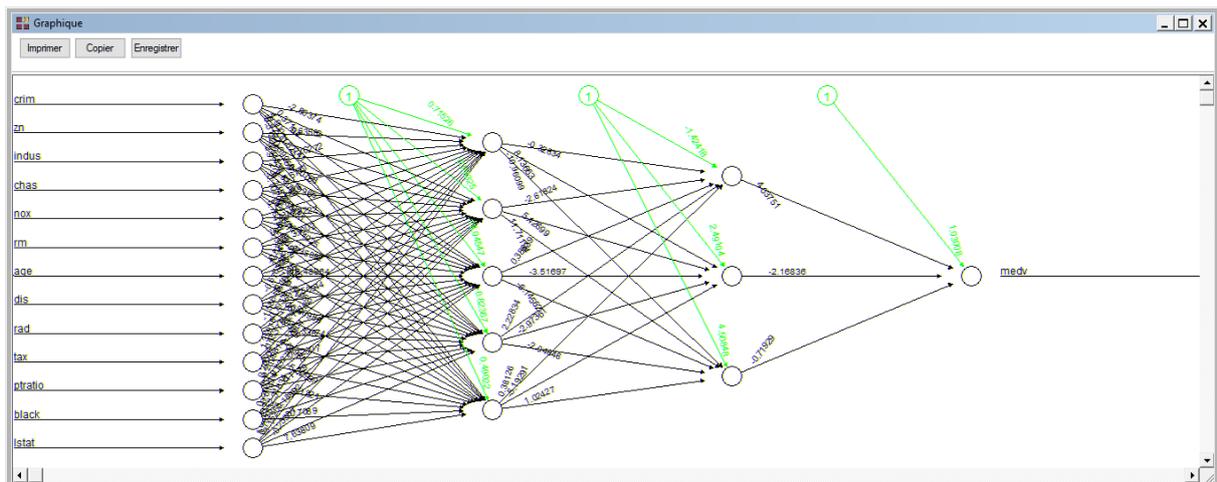
Cette option permet d'obtenir le rapport à l'écran sous la forme d'un explorateur, d'un tableur ou au format HTML.

Voici un exemple de rapport au format Explorateur.

	1	2	3	4	5	6	7	8
1								
2	Résultats pour les jeux de validation et de prévision							
3	Ce tableau affiche, pour la meilleure répétition (1) et pour les jeux de validation et de prévision,							
4	les valeurs observées (si validation), prévues et résidus (si régression) de la variable à expliquer.							
5	RMSE (racine carrée de l'erreur résiduelle) = 4.142							
6	MAE (erreur résiduelle absolue moyenne) = 2.610							
7								
8								
9		Observé	Prévu	Résidu				
10	o10	18,9	18,03107	0,86893				
11	o18	17,5	17,20920	0,29080				
12	o30	21,0	17,38466	3,60534				
13	o31	12,7	14,92329	-2,22329				
14	o33	13,2	12,62844	0,57356				
15	o43	25,3	22,13854	3,16146				
16	o48	16,6	21,26610	-4,66610				
17	o49	14,4	20,03524	-5,63524				
18	o57	24,7	22,15907	2,54093				
19	o66	23,5	26,68667	-3,18667				
20	o68	22,0	20,79350	1,20650				
21	o73	22,8	21,87600	0,92400				
22	o75	24,1	24,25279	-0,15279				

L'option Graphiques

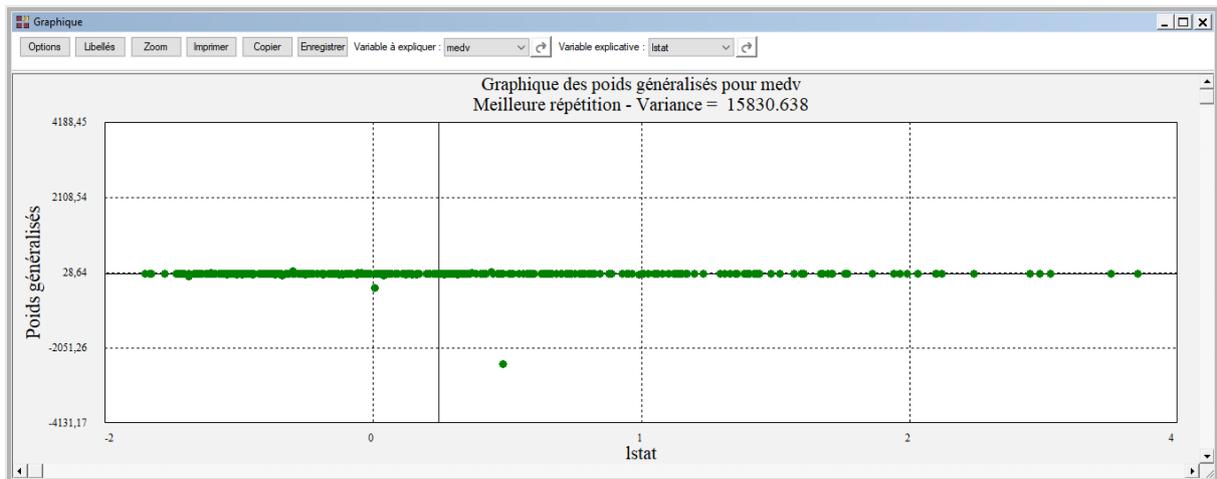
Graphique du réseau



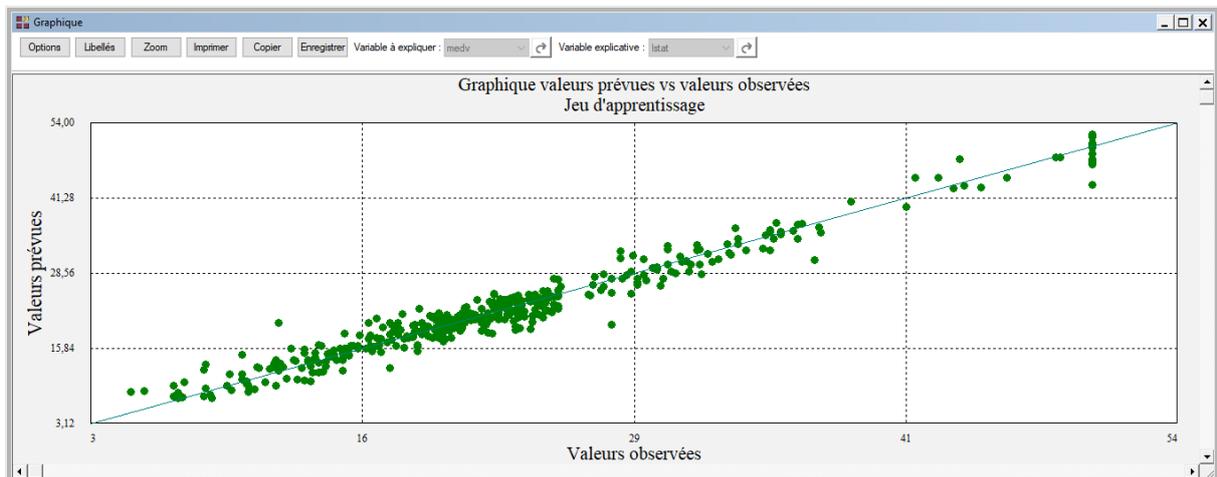
Ce graphique affiche la structure du réseau, les poids synaptiques (lignes noires) et les biais (lignes vertes) :

- Les 13 nœuds en entrée associés aux 13 variables explicatives
- Les 5 nœuds de la première couche cachée
- Les 3 nœuds de la seconde couche cachée
- Le nœud de sortie

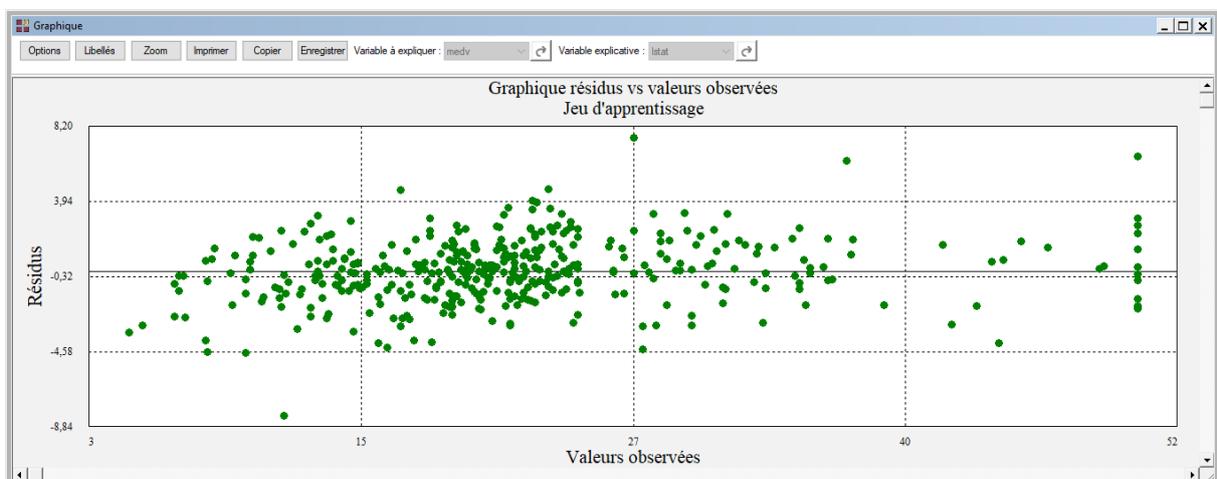
Graphique des poids généralisés



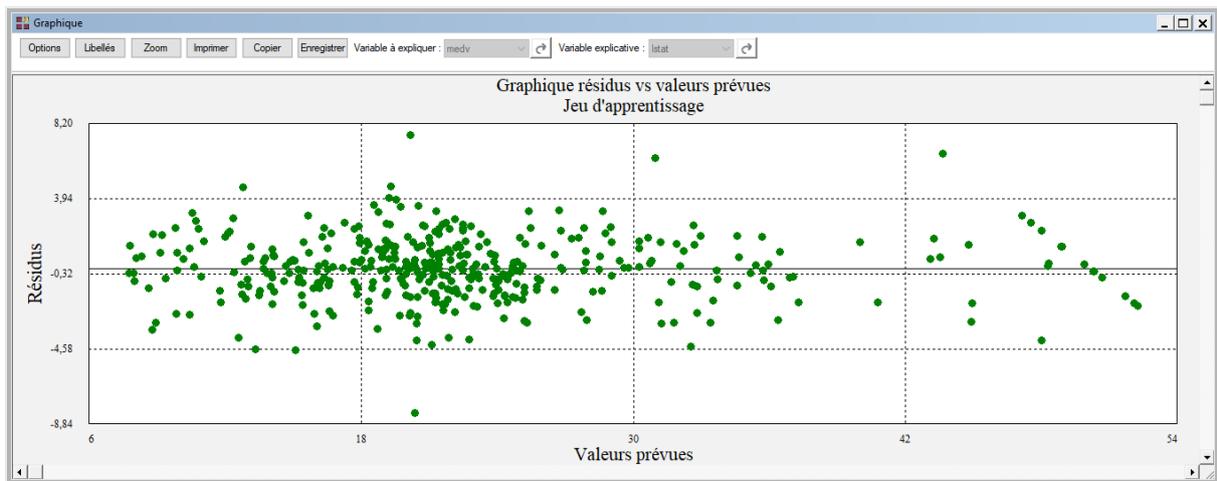
Graphique valeurs prévues vs valeurs observées (apprentissage)



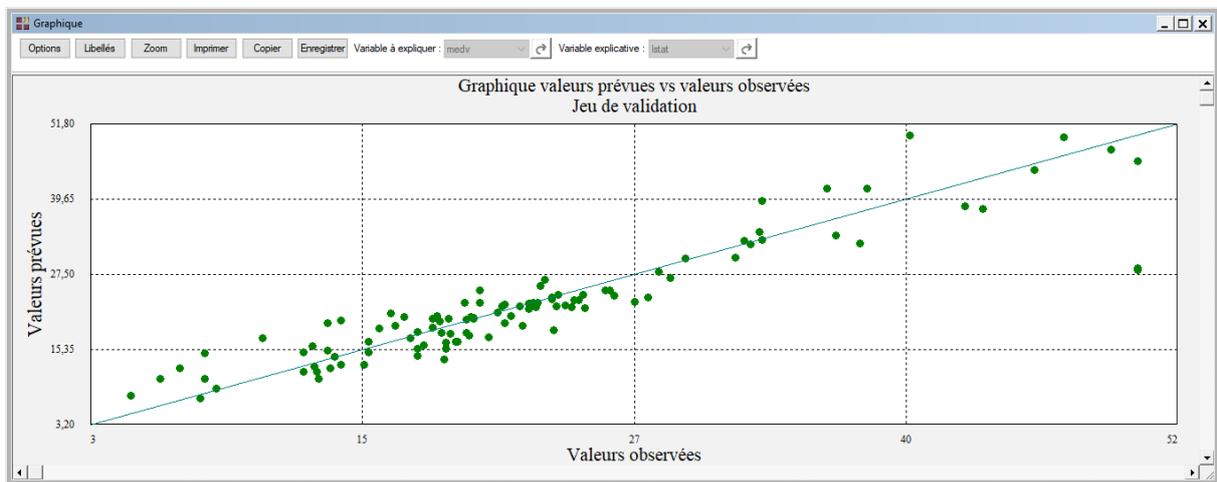
Graphique résidus vs valeurs observées (apprentissage)



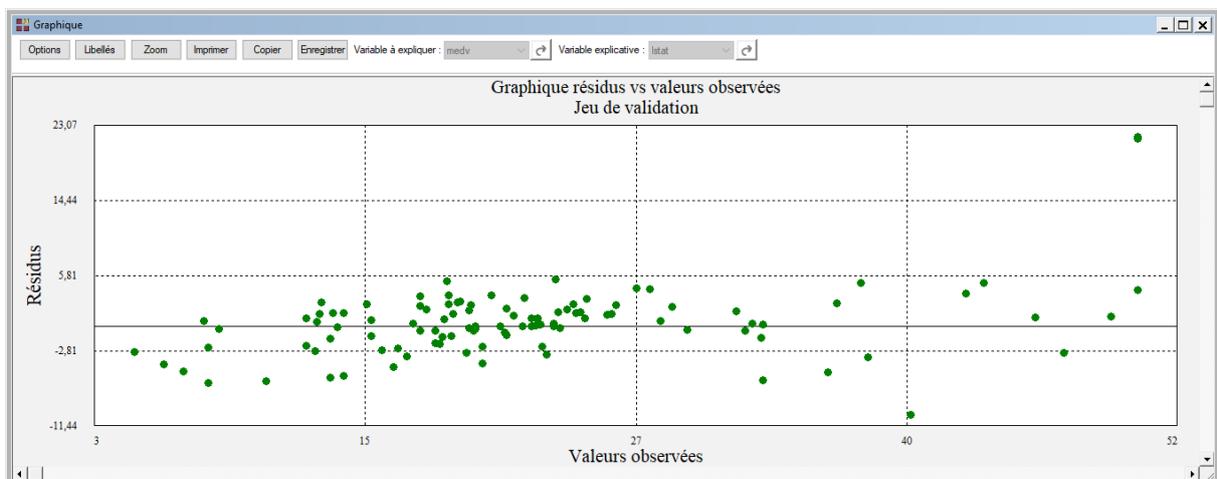
Graphique résidus vs valeurs prévues (apprentissage)



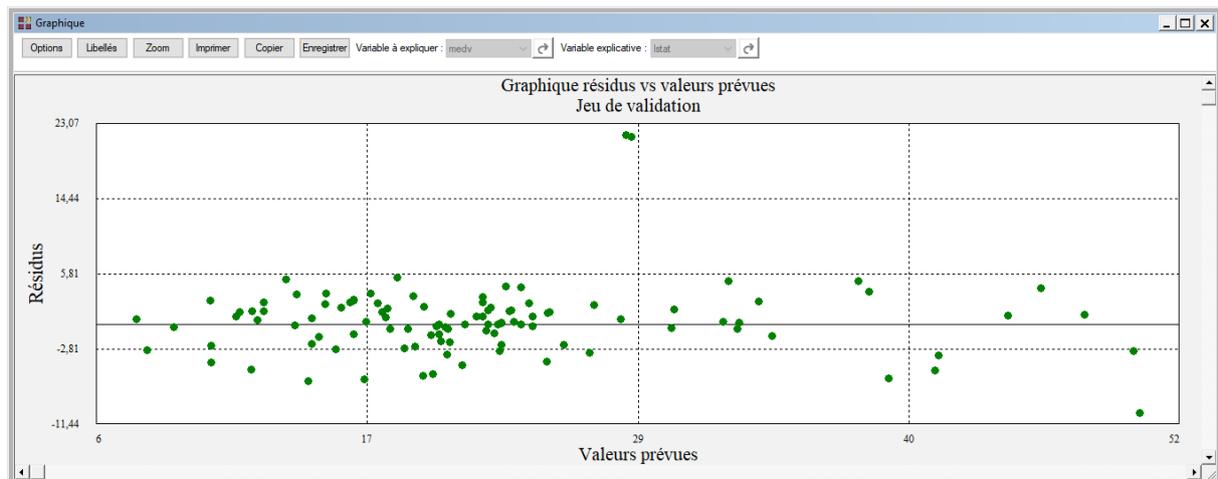
Graphique valeurs prévues vs valeurs observées (validation)



Graphique résidus vs valeurs observées (validation)



Graphique résidus vs valeurs prévues (validation)



Exemple 5 : Fichier Abalone (Régression)

Ce fichier contient des informations collectées par Warwick J Nash, Tracy L Sellers, Simon R Talbot, Andrew J Cawthorn et Wes B Ford (1994) « The Population Biology of Abalone (*Haliotis* species) in Tasmania. I. Blacklip Abalone (*H. rubra*) from the North Coast and Islands of Bass Strait ». Le but est de prévoir à partir des diverses caractéristiques mesurées l'âge des haliotis (ormeaux, oreilles de mer ou abalones) qui sont des mollusques marins à coquille unique vivant dans les eaux peu profondes du littoral.



Les caractéristiques contenues dans le fichier sont les suivantes :

Rings	Nombre d'anneaux
Age	Age
Sex	Sexe
Length	Longueur
Diameter	Diamètre
Height	Hauteur
Whole Weight	Poids total
Shucked Weight	Poids écaillé
Viscera Weight	Poids des viscères
Shell Weight	Poids de la coquille

Il s'agit de prévoir à partir de la variable Rings la variable Age en ajoutant 1,5 au nombre d'anneaux observé.

Cliquons sur l'icône ANN dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-après.

Cliquons sur le bouton Sélection pour définir le jeu d'apprentissage contenant 1000 observations.

Parmi les 502 autres observations, 500 sont affectées au jeu de validation et 2 au jeu de prévision.

Cliquons sur OK.

Réseaux de neurones artificiels

Type
Rings
Age
Sex
Length
Diameter
Height
Whole Weight
Shucked Weight
Viscera Weight
Shell Weight

Neurones par couche cachée : 3

Seuil d'arrêt : 0,05

Nb. max. d'étapes : 100000

Nb. de répétitions : 1

Méthode
 Classement Régression

Prétraitement des données
 Aucun
 Centrage et réduction
 Min-Max

Variable à expliquer : Rings

Variables explicatives :
Length
Diameter
Height
Whole Weight
Shucked Weight
Viscera Weight
Shell Weight

(Libellé de la variable à expliquer :)

(Libellés des variables explicatives :)

(Libellés des observations :)

Algorithme : rprop+

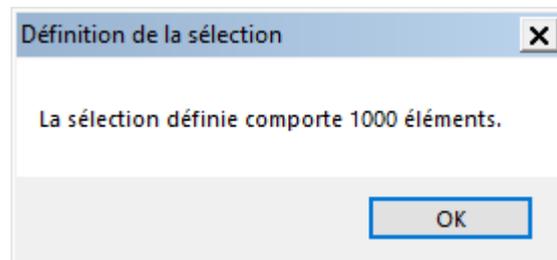
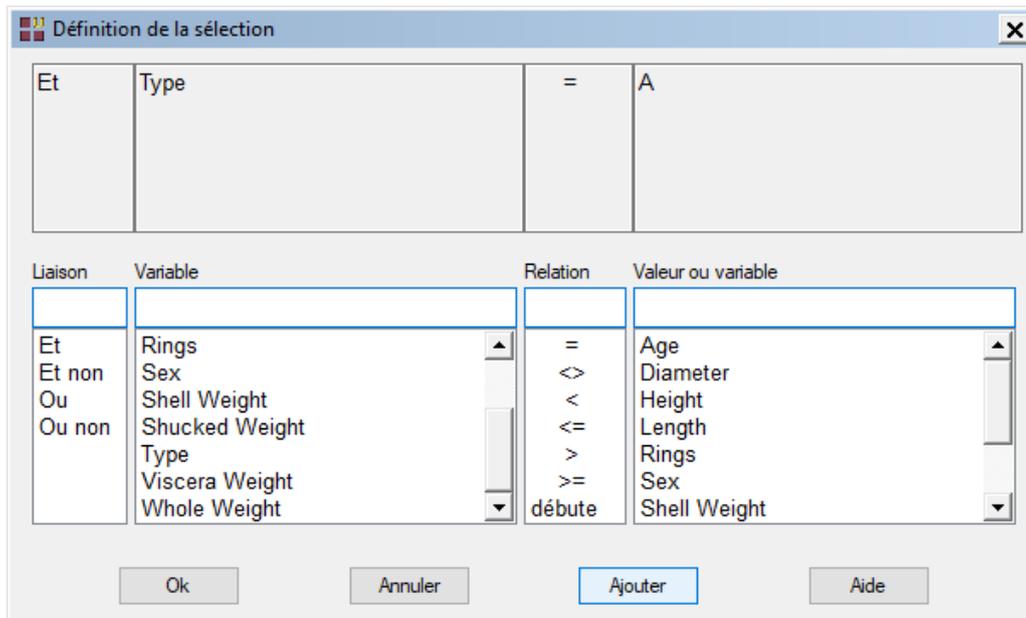
Fonction d'erreur : Erreur quadratique

Fonction d'activation : Logistique

Sortie linéaire

Racine aléatoire : 321

Ok Annuler Sélection Supprimer Aide



Après quelques instants la fenêtre 'Rapports et Graphiques' s'affiche.

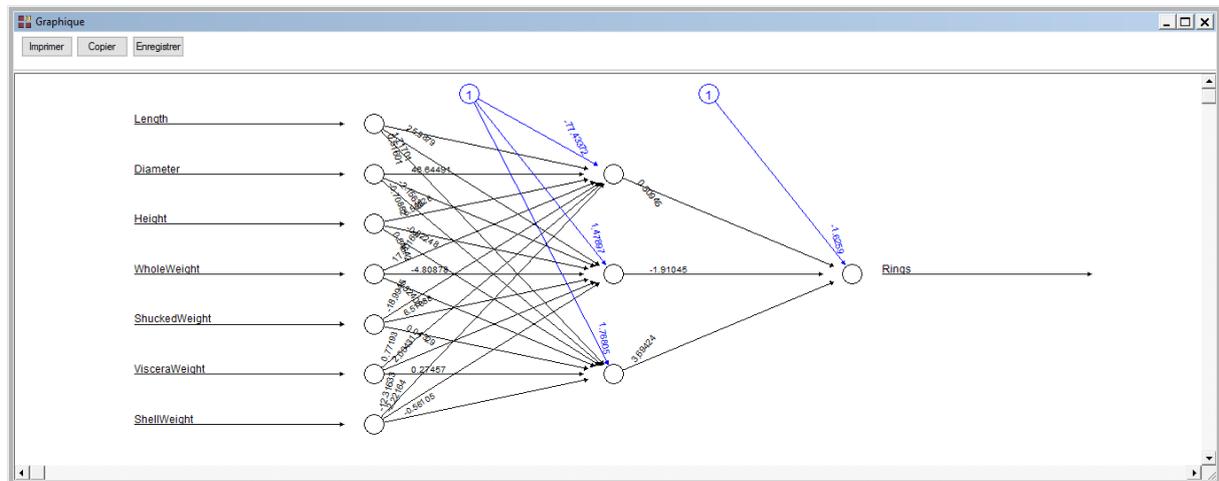
L'option Rapports

Voici un exemple de rapport au format Explorateur.

	1	2	3	4	5	6	7	8
1								
2	Résultats pour les jeux de validation et de prévision							
3	Ce tableau affiche, pour la meilleure répétition (1) et pour les jeux de validation et de prévision,							
4	les valeurs observées (si validation), prévues et résidus (si régression) de la variable à expliquer.							
5	RMSE (racine carrée de l'erreur résiduelle) = 2.112							
6	MAE (erreur résiduelle absolue moyenne) = 1.514							
7								
8								
9		Observé	Prévu	Résidu				
10	o1001		8,99307					
11	o1002		6,67383					
12	o1003	8	6,74293	1,25707				
13	o1004	11	7,40997	3,59003				
14	o1005	6	7,92844	-1,92844				
15	o1006	9	10,08547	-1,08547				
16	o1007	11	9,99318	1,00682				
17	o1008	11	8,35909	2,64091				
18	o1009	8	8,13069	-0,13069				
19	o1010	8	9,81522	-1,81522				
20	o1011	10	10,00668	-0,00668				
21	o1012	10	10,10764	-0,10764				
22	o1013	12	9,80425	2,19575				

L'option Graphiques

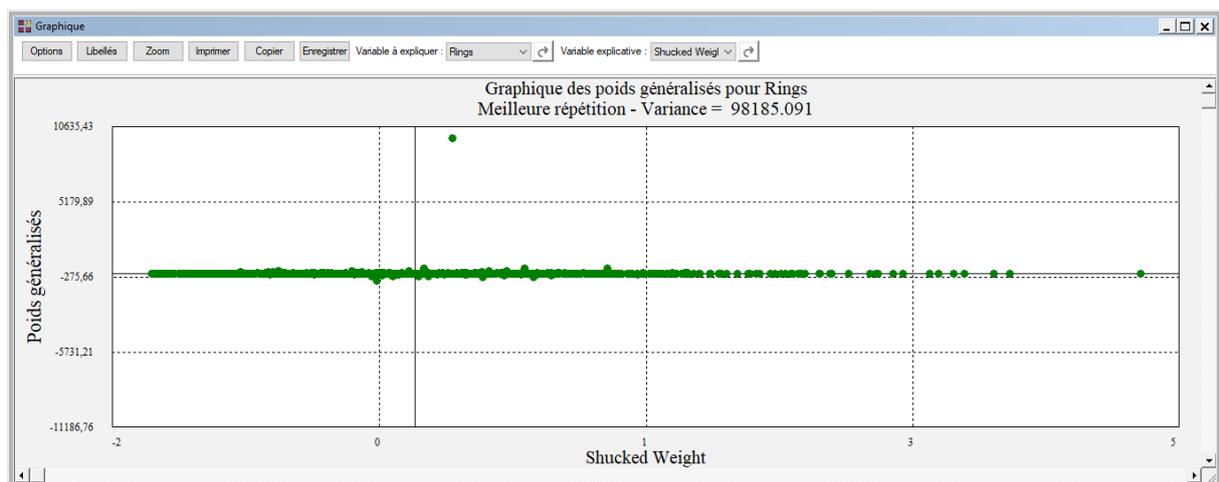
Graphique du réseau



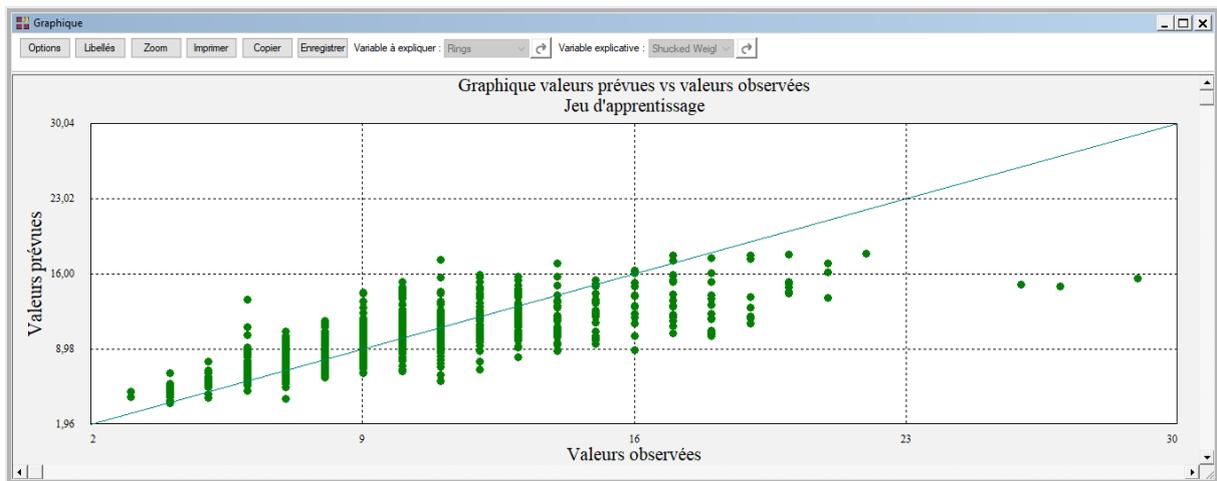
Ce graphique affiche la structure du réseau, les poids synaptiques (lignes noires) et les biais (lignes bleues) :

- Les 7 nœuds en entrée associés aux 7 variables explicatives
- Les 3 nœuds de la couche cachée
- Le nœud de sortie

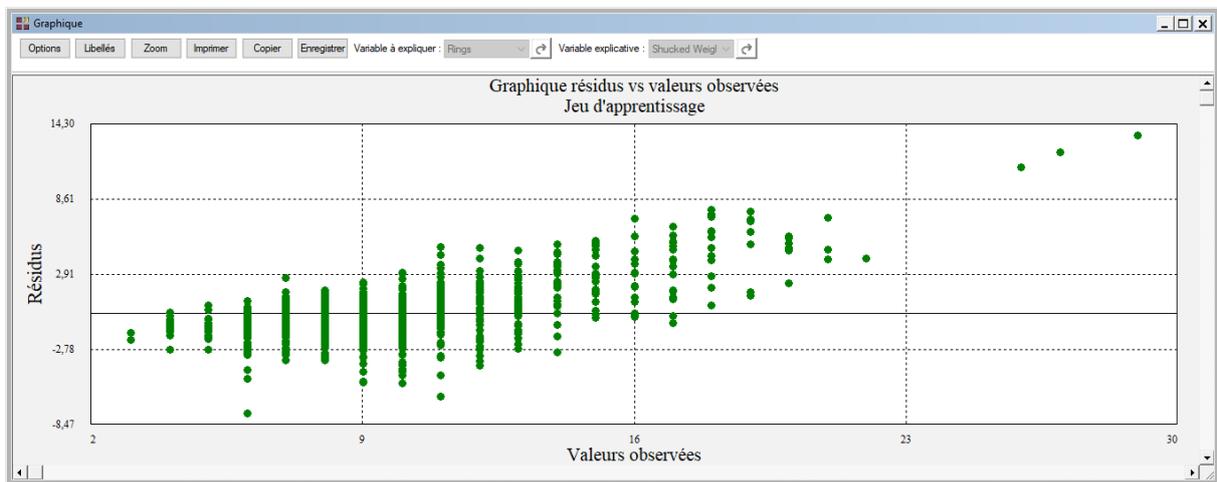
Graphique des poids généralisés



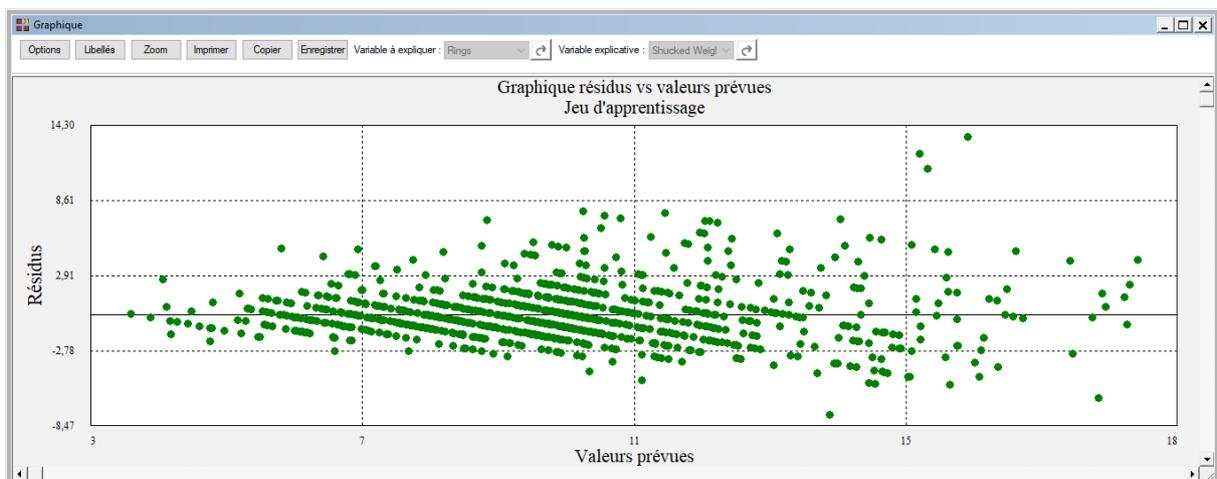
Graphique valeurs prévues vs valeurs observées (apprentissage)



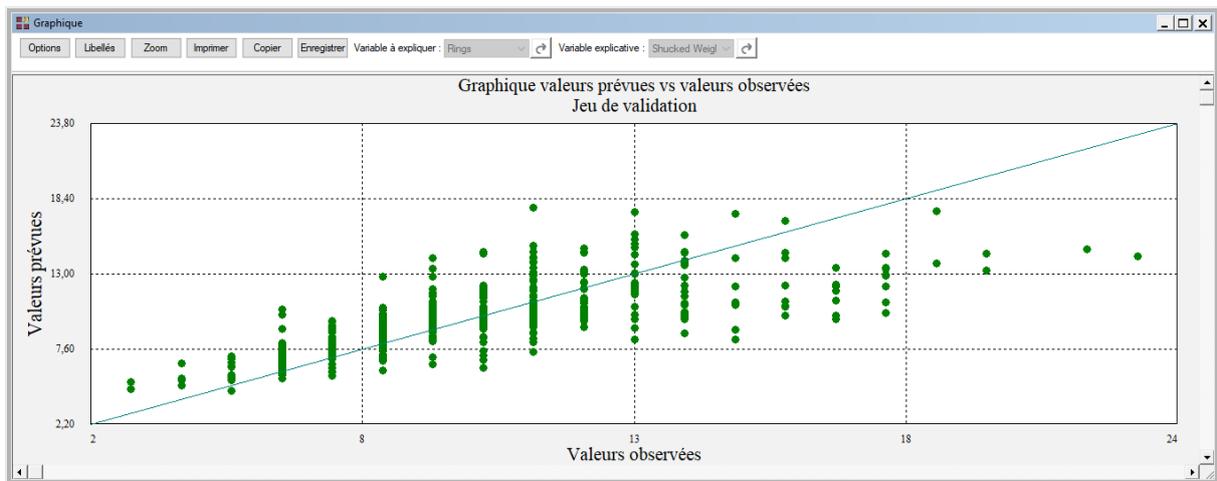
Graphique résidus vs valeurs observées (apprentissage)



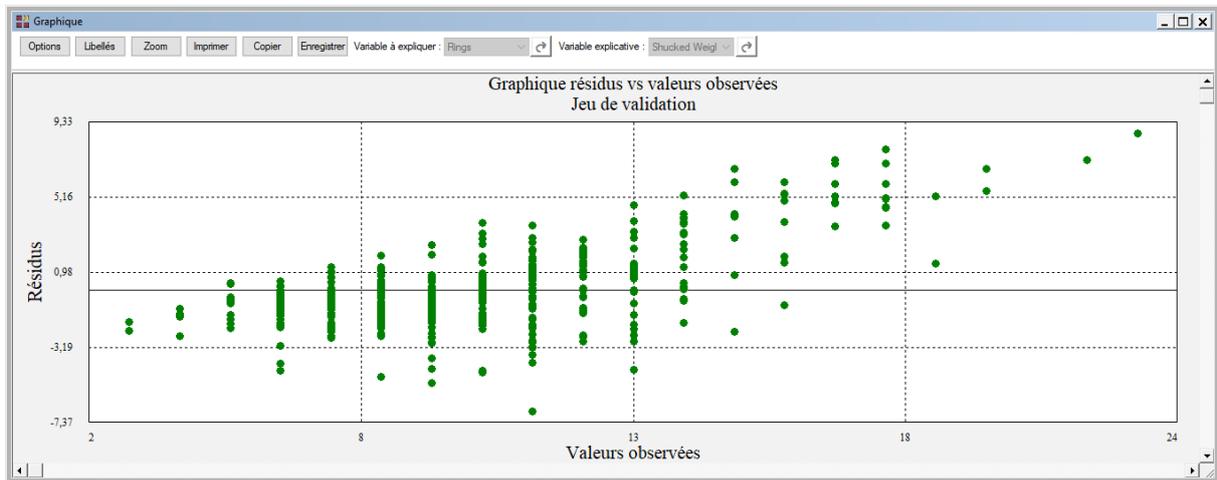
Graphique résidus vs valeurs prévues (apprentissage)



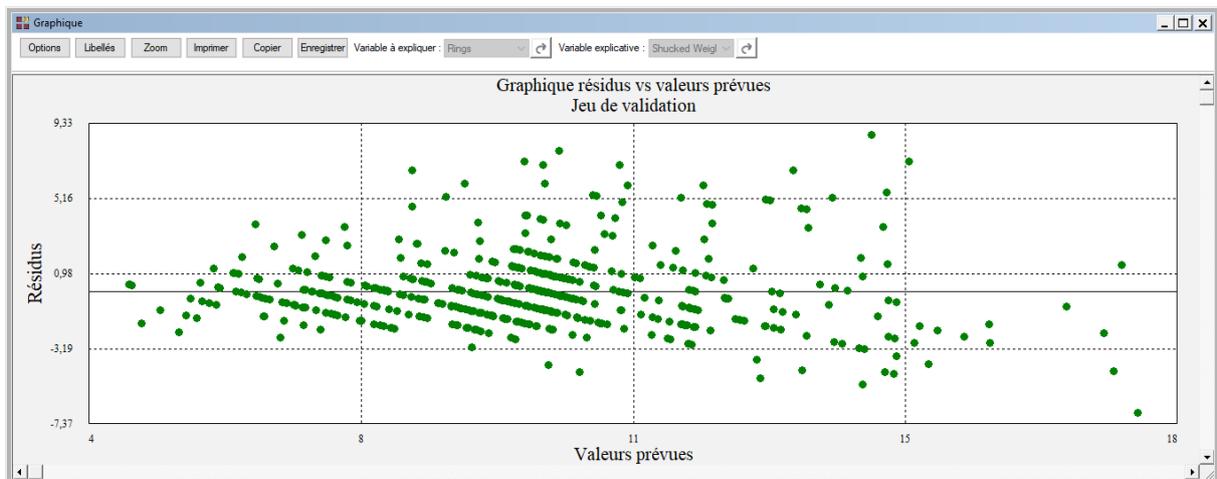
Graphique valeurs prévues vs valeurs observées (validation)



Graphique résidus vs valeurs observées (validation)



Graphique résidus vs valeurs prévues (validation)



Exemple 6 : Fichier Titanic (Classement)

Pour ce sixième exemple, nous utiliserons le fichier TITANIC. Ce fichier contient des informations concernant 714 passagers :

Statut	Décès ou Survie
Classe	Classe du passager (1 ^{ère} , 2 ^{ème} ou 3 ^{ème})
Sexe	Homme ou Femme
Age	Age du passager
Nbfse	Nombre de frères, sœurs ou époux, épouses à bord
Nbpe	Nombre de parents ou enfants à bord
Tarif	Tarif passager (en £)

Cliquons sur l'icône ANN dans le ruban Expliquer et renseignons la boîte de dialogue comme montré ci-dessous.

Réseaux de neurones artificiels

Statut

Age

Tarif

Nbfse

Nbpe

Classe

Sexe

Poids

LibVarQuant

LibVarQual

LibObs

Neurones par couche cachée : 2

Seuil d'arrêt : 0,01

Nb. max. d'étapes : 100000

Nb. de répétitions : 5

Méthode

Classement Régression

Prétraitement des données

Aucun

Centrage et réduction

Min-Max

Variable à expliquer : Statut

Variables explicatives : Age, Tarif, Nbfse, Nbpe

(Libellé de la variable à expliquer :)

(Libellés des variables explicatives :)

(Libellés des observations :)

Algorithme : rprop+

Fonction d'erreur : Erreur quadratique

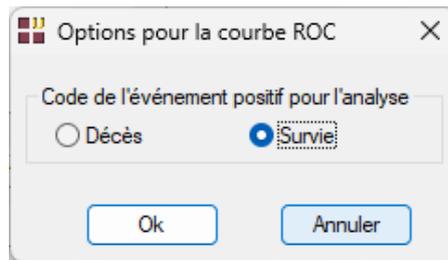
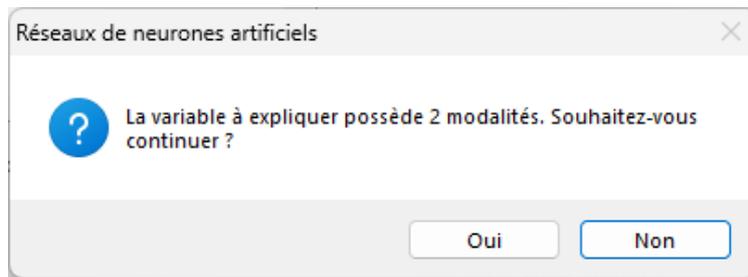
Fonction d'activation : Logistique

Sortie linéaire

Racine aléatoire : 12345

Ok Annuler Sélection Supprimer Aide

Confirmons le souhait de mettre en œuvre une méthode de classement et définissons le code de l'événement positif à Survie.



Après quelques instants la fenêtre du rapport s'affiche.

	1	2	3	4	5	6	7	8
1								
2	Résumé de l'analyse							
3	La meilleure répétition est la répétition numéro 4.							
4								
5								
6		Répétition 1	Répétition 2	Répétition 3	Répétition 4	Répétition 5		
7	[1.] Erreur	135,22165	141,82609	133,07017	133,08935	136,27551		
8	[2.] Seuil atteint	0,00933	0,00928	0,00971	0,00953	0,00928		
9	[3.] Nombre d'étapes	5714,00000	1000,00000	10038,00000	8674,00000	3830,00000		
10	[4.] "Intercept.to.1layhid1"	-1,05888	1,96177	-5,58803	5,62528	1,01897		
11	[5.] "Age.to.1layhid1"	2,14266	-1,66868	4,96536	-4,99515	-0,53039		
12	[6.] "Tarif.to.1layhid1"	-7,75837	9,63845	-15,01402	15,11029	5,04046		
13	[7.] "Nbfsse.to.1layhid1"	2,22492	-1,75123	5,74182	-5,77454	-0,65157		
14	[8.] "Nbpe.to.1layhid1"	0,47961	0,02459	1,35927	-1,36894	-0,53131		
15	[9.] "Intercept.to.1layhid2"	2,89930	1,59643	-19,31800	19,53604	-23,24553		
16	[10.] "Age.to.1layhid2"	-10,14291	-0,00524	-4,44705	4,47975	-16,43982		
17	[11.] "Tarif.to.1layhid2"	34,90324	0,00551	-19,46154	19,86157	-11,51028		
18	[12.] "Nbfsse.to.1layhid2"	-15,53953	-0,00308	-9,24615	9,23216	-11,48921		
19	[13.] "Nbpe.to.1layhid2"	-4,85107	-0,00060	-13,47325	13,69848	10,67177		
20	[14.] "Intercept.to.Deces"	-0,31550	0,08579	0,28182	1,28352	0,95861		
21	[15.] "1layhid1.to.Deces"	1,16738	-0,56589	0,52990	-0,52951	-0,67025		
22	[16.] "1layhid2.to.Deces"	0,55978	0,90726	0,47227	-0,47194	-0,44139		

La meilleure répétition est la répétition 4. Les calculs et graphiques seront basés sur cette répétition.

La matrice de confusion pour le jeu d'apprentissage indique que 74 % des observations sont bien classées.

L'aire sous la courbe ROC (AUC) est égale à 0,769.

Le modèle est donc modérément performant.

Rapports et Graphiques

Rapport ANN

- Résumé de l'analyse
- Poids ajustés
- Poids généralisés Deces
- Poids généralisés Survie
- Détails apprentissage
- Confusion apprentissage
- VP, FN, FP, VN, Sensibilité, Spécificité

	1	2	3	4	5	6	7	8
1								
2	Matrice de confusion pour le jeu d'apprentissage							
3	Ce tableau affiche pour la meilleure répétition (4) les nombres d'observations observées par classe en lignes et les nombres d'observations prévues par classe en colonnes.							
4								
5								
6	Pourcentage de mal classés : 25,770 %							
7	Pourcentage de bien classés : 74,230 %							
8								
9								
10	Observé \ Prévu	Deces	Survie	Total				
11	Deces		346	78	424			
12	Survie		106	184	290			
13	Total		452	262	714			
14								
15								
16								
17								
18								
19								
20								
21								
22								

Rapport Explorateur /

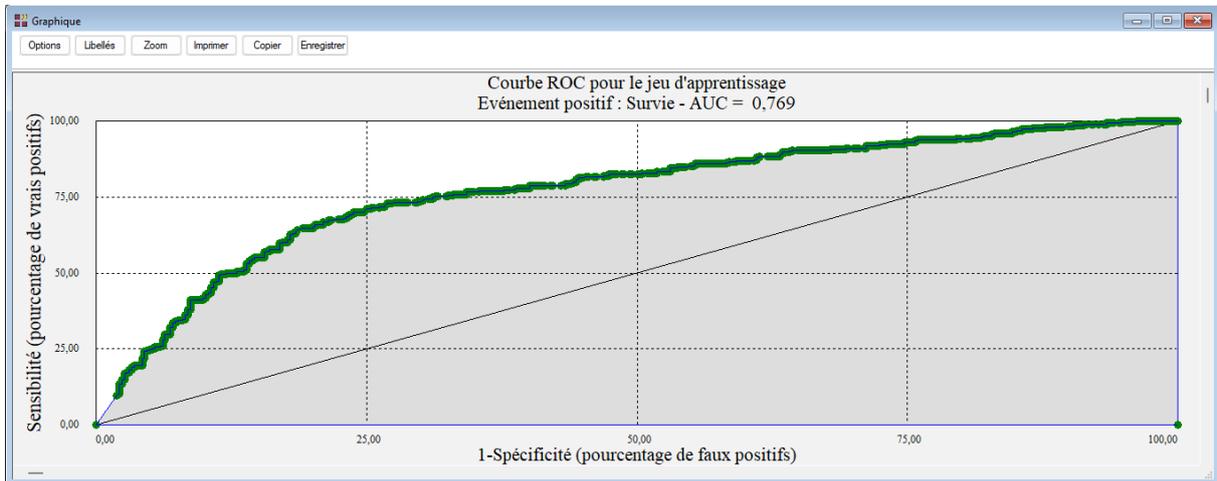
Rapports et Graphiques

Rapport ANN

- Résumé de l'analyse
- Poids ajustés
- Poids généralisés Deces
- Poids généralisés Survie
- Détails apprentissage
- Confusion apprentissage
- VP, FN, FP, VN, Sensibilité, Spécificité

	1	2	3	4	5	6	7	8
1								
2	VP, FN, FP, VN, SENSIBILITE, SPECIFICITE POUR LE JEU D'APPRENTISSAGE							
3								
4	Mesures = Mesures uniques							
5	VP = Nombres de vrais positifs							
6	FN = Nombres de faux négatifs							
7	FP = Nombres de faux positifs							
8	VN = Nombres de vrais négatifs							
9	Sensibilité en %							
10	Spécificité en %							
11								
12	Code de l'événement positif : Survie							
13	Aire sous la courbe (AUC) = 0,769							
14								
15								
16		Mesures	VP	FN	FP	VN	Sensibilité	Spécificité
17	1	Infmi	290	0	424	0	100,00000	0,0000
18	2	-0,23799	290	0	424	0	100,00000	0,0000
19	3	-0,23483	290	0	423	1	100,00000	0,2358
20	4	-0,23161	290	0	422	2	100,00000	0,4717
21	5	-0,22753	290	0	421	3	100,00000	0,7075
22	6	0,00109	290	0	420	4	100,00000	0,9434

Rapport Explorateur /



Note : Pour comparer les performances de plusieurs méthodes d'analyse, cet exemple est traité dans les six analyses AFD, ADB, KNN, BAYES, ANN et ARBRE.

Les variables internes créées par la procédure

Voici la liste des variables internes créées par la procédure. A noter que certaines des variables mentionnées ci-dessous peuvent ne pas apparaître, en fonction des options choisies.

<i>Variable</i>	<i>Contenu</i>
resapp	Résultats pour le jeu d'apprentissage
poids	Poids du réseau
libligresultmat	Libellés des lignes de la matrice des résultats
resultmat	Matrice des résultats
meilleurerep	Meilleure répétition
resresult	Résultats du réseau
poidsgene	Poids généralisés
libinda	Libellés des individus du jeu d'apprentissage
preva	Prévisions pour le jeu d'apprentissage
resida	Résidus pour le jeu d'apprentissage (régression)
libindvp	Libellés des individus des jeux de validation et de prévision
prevvp	Prévisions pour les jeux de validation et de prévision
residvp	Résidus pour les jeux de validation et de prévision (régression)
vpA	Vrais positifs (apprentissage)
fnA	Faux négatifs (apprentissage)
fpA	Faux positifs (apprentissage)
vnA	Vrais négatifs (apprentissage)
specificiteA	Spécificité (apprentissage)
sensibiliteA	Sensibilité (apprentissage)
aireA	Aire sous la courbe (apprentissage)
vpV	Vrais positifs (validation)
fnV	Faux négatifs (validation)
fpV	Faux positifs (validation)
vnV	Vrais négatifs (validation)
specificiteV	Spécificité (validation)
sensibiliteV	Sensibilité (validation)
aireV	Aire sous la courbe (validation)

Références

[Documentation du package R 'neuralnet' \(2019\)](#)

<https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>

Exemple 1

Ricco Rakotomalala – Université Lumière Lyon 2
« Deep Learning – Perceptrons simples et multicouches »

Ricco Rakotomalala – Université Lumière Lyon 2
« Tanagra_Optimal_Neurons_Perceptron.pdf »

Exemple 2

Fisher, R.A. (1936). « The use of multiple measurements in taxonomic problems. »
Ann. Eugenics 7, Pt. II, 179-188.

Exemple 3

<https://archive.ics.uci.edu/ml/datasets/wine>

Exemple 4

<http://lib.stat.cmu.edu/datasets/boston>

Exemple 5

Warwick J Nash, Tracy L Sellers, Simon R Talbot, Andrew J Cawthorn et Wes B Ford (1994) « The Population Biology of Abalone (Haliotis species) in Tasmania. I. Blacklip Abalone (H. rubra) from the North Coast and Islands of Bass Strait », Sea Fisheries Division, Technical Report No. 48 (ISSN 1034-3288)

Exemple 6

Kaggle.com
<https://www.kaggle.com/c/titanic>

GitHub repository
<https://github.com/alexperrier/packt-aml/blob/master/ch4>