

Modélisation par réseau de neurones

1 Le logiciel

Le logiciel `neuro.c` implémente un réseau à trois couches, optimisé par l'algorithme de rétropropagation. Il nécessite :

- un fichier de configuration, donnant les informations nécessaires pour réaliser un apprentissage,
- un fichier d'apprentissage et un fichier de test.

Ce logiciel doit être compilé, par exemple par la commande :

```
gcc -o neuro neuro.c -lm
```

1.1 Le fichier de configuration

Ce fichier doit contenir les informations nécessaires sur la structure du RN, les paramètres d'apprentissage et les fichiers de données.

Structure

La structure du RN est déterminée par la nature des fonctions d'activation et le nombre de neurones dans chaque couche.

- Les fonctions d'activation programmées sont la sigmoïde (S), la tangente hyperbolique (T), la gaussienne (G) et la fonction identité (I).
- Dans le logiciel utilisé, il n'y a qu'une seule sortie. Il suffit donc de préciser le nombre d'entrées (qui dépend du problème à traiter) et le nombre de neurones cachés, à estimer "au mieux".

Si on choisit un RN avec 5 entrées et 4 neurones cachés, avec la sigmoïde en couche cachée et la gaussienne en sortie, les deux premières lignes du fichier de configuration seront (dans cet ordre) :

```
S G  
5 4
```

Paramètres d'apprentissage

Il faut préciser, dans l'ordre :

- le nombre d'itérations,
- les paramètres ϵ et α , à déterminer empiriquement ($0.00001 \leq \epsilon \leq 0.2$, $0 \leq \alpha \leq 0.9$).

Les données

Suivent ensuite le nom des fichiers d'apprentissage et de test, ainsi qu'une indication (o/n) pour la normalisation des données :

- o l'utilisateur doit alors donner les bornes minimum et maximum pour chaque entrée
- n la normalisation est faite automatiquement

Dernières indications

- Un booléen, sous la forme o/n, permet de savoir si le RN est généré aléatoirement ou non. Si la réponse est n, il faut fournir un RN complet.
- Il faut préciser le pas pour l’affichage des résultats (nombre entier),
- le nombre d’entrées potentielles dans les fichiers de données (il peut y avoir, par exemple, 50 entrées potentielles sur lesquelles on ne choisit que 5 comme entrées effectives),
- enfin, on donne le numéro des colonnes qui correspondent aux variables d’entrée (attention : la première colonne porte le numéro 0).

Exemple de fichier de configuration

On donne le fichier :

```
S T
3 4
1000
0.01 0.0
f.app
f.tst
n
o
50
5
0 2 3
```

Comment interpréter ce fichier ?

Lancement d’un apprentissage

Soit `config` le nom du fichier de configuration. Pour lancer un apprentissage, il suffit de donner la commande :

```
neuro config
```

1.2 Les ensembles d’apprentissage et de test

Les fichiers d’apprentissage et de test présentent les données sous une forme matricielle.

- La colonne i comporte les différentes observations de la variable d’entrée i .
- La dernière colonne représente obligatoirement la sortie voulue.

2 Travail à effectuer

Il s’agit d’identifier un modèle neuronal de la série statistique “Gas Furnace Data” de Box et Jenkins.