MASTER : Traitement de l'Information et Exploitation des Données
Fouad Badran, Cécile Mallet, Awa Niang, Charles Sorror, Sylvie Thiria

Master TRIED,

TPB03 : Rapport

Sujet :

Utilisation d'un MLP pour la classification de chiffres manuscrits

Réalisé par :

Exemple de Rapport

Année universitaire : ----/----

Résumé de l'énoncé :

Objectifs :

1) Comparaison de différentes architectures (1ère partie)

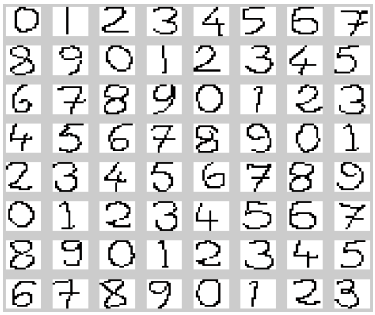
2) Etude de l'importance du codage (2ème partie)

3) Utilisation des masques et poids partagés (3ème partie)

Données brutes :

- Base de 480 chiffres chacun représenté par un vecteur de 256 pixels codés ±1, qui correspond à une image 16x16.

  Extrait :



- Réponses désirées : 480 vecteurs de longueur 10 codés 1 à l'indice de la bonne réponse et -1 ailleurs.

Répartition des données :

- Ensemble d'Apprentissage : chiffres manuscrits de 1 à 300
- Ensemble de Validation : chiffres manuscrits 301 à 400
- Ensemble de Test : chiffres manuscrits 401 à 480

1ère Partie : Optimisation de l'architecture du classifieur MLP

Comme demandé dans l'énoncé, on a étudié les architectures suivantes :

- Réseau sans couche cachée à sortie linéaire
- Réseau sans couche cachée à sortie tangente hyperbolique
- Réseau à une couche cachée dont les neurones utilisent la fonction tangente hyperbolique; les neurones de sortie du réseau ayant une fonction d'activation linéaire. Pour ce réseau, on a fait varier le nombre de neurones cachés.

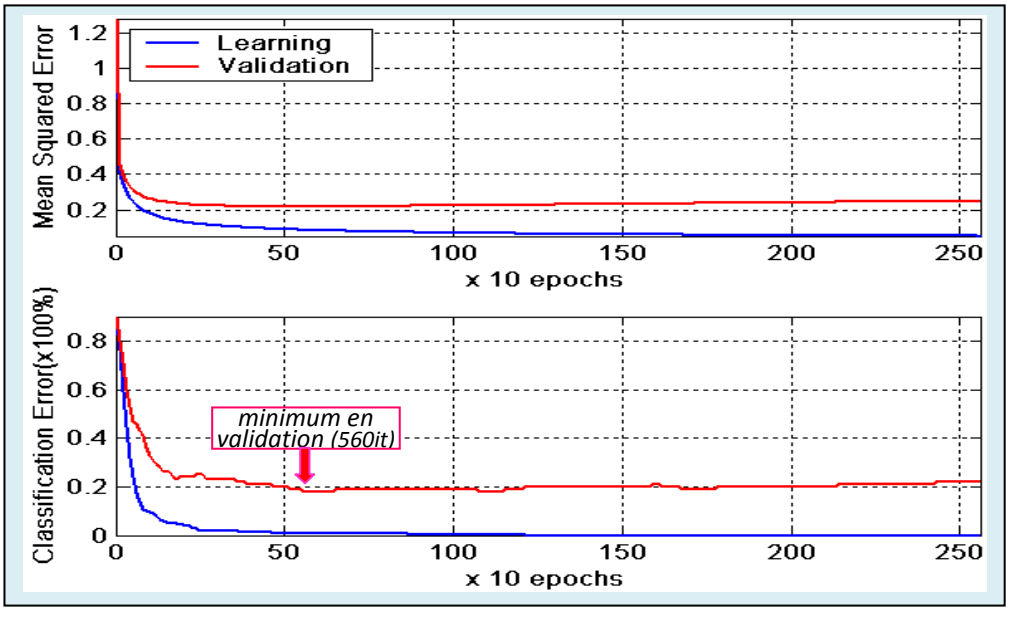
Les résultats en terme d'erreur et de performance **sur la base de test** sont présentés ci-après sous forme de tableau. Ils sont accompagnés des courbes d'apprentissage pour la 1ère partie du TP. Pour chaque situation, on a effectué plusieurs expériences. Les résultats présentés ici sont les meilleurs obtenus. Pour chaque présentation de résultat on indiquera, en plus de l'erreur et de la performance, les informations complémentaires suivantes :

- m = le nombre de neurones sur la couche cachée (s'il y a lieu).
- itmpv = l'indice de l'itération à laquelle le minimum de la performance en validation a été détecté (sachant que par ailleurs, lorsque le graphe est présenté, le nombre d'itérations effectuées apparait en abscisses).

• Réseaux sans couche cachée

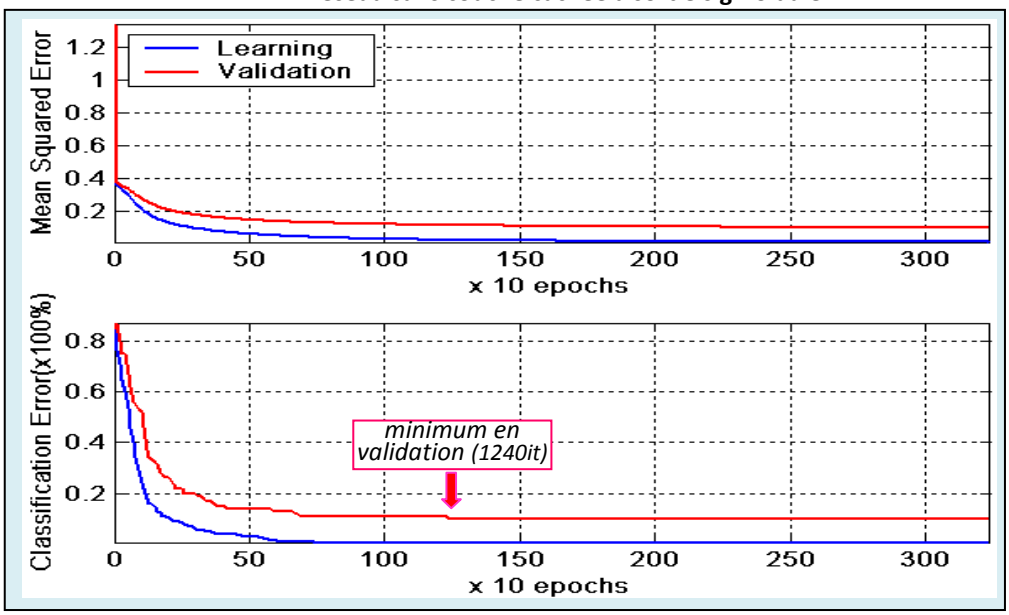
On présente ici les courbes d'apprentissage des architectures sans couche cachée en commençant par le réseau à sorties linéaire puis ensuite celui à sorties tangentes hyperboliques. Les courbes bleues se rapportent à l'ensemble d'apprentissage alors que les rouges sont celles calculées sur l'ensemble de validation.
L'erreur quadratique moyenne (MSE) est représentée (pour ces 2 ensembles) par les courbes du haut. Celles du bas correspondent à la performance, c'est-à-dire au pourcentage de chiffres mal classés (pour l'ensemble considéré).

Réseau sans couche cachée à sortie linéaire



Pour le 1er réseau, on constate une décroissance rapide des courbes qui commencent à se stabiliser à partir de la 200 ou 300ième itération. Le minimum de la performance sur l'ensemble de validation se situe à la 560ième itération.

Réseau sans couche cachée à sortie sigmoïdale



Avec le réseau à sorties sigmoïdales, où l'on constate aussi une décroissance rapide des courbes et une stabilisation à partir de 200 itérations environ, la performance minimale sur l'ensemble de validation apparaît à la 1240ième itération.

Les résultats sur l'ensemble de test de ces 2 réseaux sans couche, sont présentés dans le tableau ci-après. On observe un net avantage du 2ème modèle (à sortie sigmoïdale) sur le 1er modèle (à sortie linéaire) montrant l'intérêt d'utiliser un modèle non linéaire.

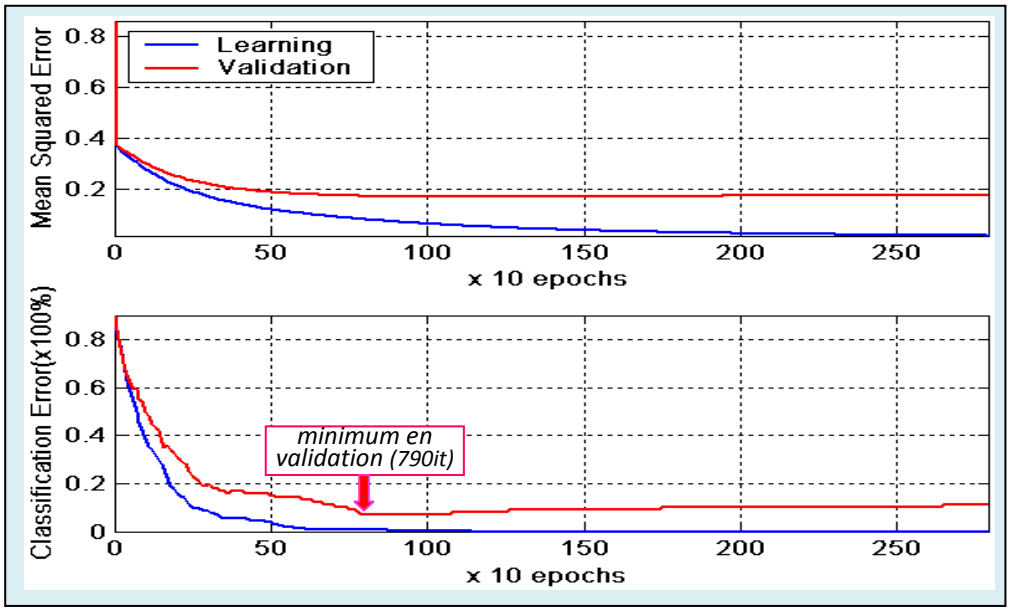
Tableau1 : Résultats des modèles neuronaux sans couches cachées

Type de réseau	**Réseau linéaire**	**Réseau sigmoïdal**
E_{test} **(Erreur sur la base de test)**	**0.196**	**0.108**
Performance : % de chiffres mal classés.	**18.8**	**11.3**
Informations complémentaires	itmpv=560	itmpv=1240

- **Réseau à une couche cachée.**

La 3ème architecture utilisée est constituée de sorties linéaires, et de neurones cachés à fonction d'activation tangente hyperbolique. Pour ce réseau, on fait varier le nombre m de neurones de la couche cachée. Les résultats optimaux semblent se situer pour des valeurs de m allant de 10 à 20 environs. On présente ici les courbes d'apprentissage pour l'un des meilleurs résultats obtenus avec m=15.

Réseau avec à sortie sigmoïdale avec 15 neurones cachés



Des résultats comparables on été obtenus pour des valeurs de m=10 ou m=20 par exemples, comme le montre le tableau 2 ci-dessous. Pour m=15 la performance minimale en validation se produit à la 790ième itération.

Tableau2 : Résultats des réseaux à 1 couche cachée

Informations complémentaires	**m=10** itmpv=2520	**m=15** itmpv=790	**m=20** itmpv=1540
E_{test} **(Erreur sur la base de test)**	**0,161**	**0,171**	**0,167**
Performance : % de chiffres mal classés.	**15**	**13,8**	**15**

• Informations complémentaires

On précise ici que la valeur du pas de gradient utilisé pour les résultats présentés ci-dessus était de **0,1**. Pour ce travail on a aussi testé d'autres valeurs comme **0.05**, **0.2** ou **0.5**. Cela n'a pas modifié significativement les résultats ou les a même parfois plutôt dégradés.
On a également, par ailleurs, réduit la taille de l'ensemble d'apprentissage en n'utilisant que les 200 premières formes de la base de données. Le gain du temps d'apprentissage est constatable, mais, en contrepartie, les résultats sont moins probants.

• Commentaires

Pour cette série d'expériences on remarque que c'est le réseau sans couche cachée à fonction sigmoïde qui présente le meilleur compromis (temps d'apprentissage/performance). Ceci peut ce comprendre en partie si l'on effectue le décompte de nombres de paramètres. Il y en a 2570 pour un réseau sans couche cachée et il faut au moins 10 cellules cachées pour en obtenir un nombre à peu près équivalent soit 2680. On sait qu'à partir d'une certaine complexité, il n'est plus utile d'augmenter le nombre de connexions qui, s'il devient trop important, conduit même à une dégradation des résultats. Le réseau sans couche semble donc comporter un nombre approprié de connexions.
Si un réseau sans couche permet d'effectuer une classification acceptable c'est probablement que l'hypothèse linéaire est plus ou moins envisageable pour ce cas particulier d'étude. Cette hypothèse linéaire est d'autant plus envisageable si l'on considère le rapport entre la dimension (de l'espace) du problème qui est de 256 et le nombre de points utilisés dans cet espace qui n'est que de 300. Les résultats sont cependant très dépendants des conditions initiales des poids.