

TP4 – SAS et Entreprise Miner : chiffres manuscrits

{F. Denis, Yann.Esposito}@lif.univ-mrs.fr

19 janvier 2005

- ☞ Ce TP sera dédié à l'utilisation du module Entreprise Miner de SAS. Vous devez avoir fini le tutoriel en ligne (e-Learning) avant de commencer ce TP. Vous devez simplement vous familiariser avec Entreprise Miner de façon à pouvoir être autonome.

0 Configuration des chemins et récupération des échantillons

Dans un premier temps créez une nouvelle bibliothèque que vous nomerez PERSO dans votre répertoire personnel.

Lancez Entreprise Miner et dans les options de personnalisation, faites en sorte de sauvegarder les données sur votre compte. Option → User Preferences, ouvrir l'onglet Project, cliquez sur Browse puis choisissez K:\SASProjects\ par exemple.

File → delete current project

File → new project (chiffres manuscrits)

1 Récupération des données

Fermez ou minimisez Entreprise Miner pour revenir sur SAS.

Les fichiers contenant les données peuvent être trouvés au format csv dans

<http://www.lif.univ-mrs.fr/~esposito/enseignement.php>

Importez les données dans la bibliothèque PERSO, puis ouvrez Entreprise Miner.

Toujours sur ma page web, vous avez un lien vers un rapport créé avec SAS. Passez un moment à étudier le diagramme et essayez de trouver les différents paramètres passés aux algorithmes d'apprentissage. De la même manière vérifiez l'importance et l'apport d'un échantillon de validation pour tous les types de classifieur.

Vérifiez l'effet des différents types de topologie des réseaux de neurones. Vérifiez l'importance du nombre de couches, le fait de considérer que les données sont fortement bruitées ou pas du tout.

Étudiez le nombre d'étapes nécessaires à la convergence des réseaux de neurones.

Trouvez un moyen de sélectionner les attributs...

En bref, prenez le temps d'essayer la plupart des outils, et surtout passez du temps à comprendre chacun des éléments. En particulier, essayez de faire les choses dans l'ordre, regardez comment transformer vos données d'entrées. Essayez d'améliorer au mieux les réseaux de neurones. Vérifiez les effets entre les transformations faites aux données et aux résultats. Essayez d'ajouter de nouveaux attributs (pensez aux SVM) pour améliorer les résultats. N'oubliez pas que vos données sont des images...

