

**TP5 – SAS et Entreprise Miner**  
{Francois.Denis, Yann.Esposito}@lif.univ-mrs.fr  
26 janvier 2005

☞ Ce TP sera dédié à l'utilisation de SAS Enterprise Miner. Il s'agira d'aller plus loin qu'avec le tutoriel.

## 0 Configuration des chemins et récupération des échantillons

**0.1.** Dans un premier temps créez une nouvelle bibliothèque que vous nomerez PERSO. Cette bibliothèque devra pointer dans un répertoire situé sur votre compte personnel (dans K:\).

**0.2.** Lancez Enterprise 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

**1.1.** Fermez ou minimisez Enterprise Miner pour revenir sur SAS. Les fichiers contenant les données peuvent être trouvés au format csv dans (choisissez la version avec identifiant)

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

Importez les données dans la bibliothèque PERSO, puis ouvrez Enterprise Miner. Lors de l'importation, les noms des attributs passent de X1, ..., X257 à VAR1, ..., VAR257. L'attribut VAR257 correspond à la classe. Notez que le type de cette classe est dite numérique alors qu'elle devrait être nominale. Modifiez cette attribution.

Je vous demande de ne pas utiliser le fichier de test pour l'apprentissage, ni pour vérifier quel est le meilleur classifieur.

**1.2.** Si vous ne l'avez pas fait la semaine dernière, je vous demande de bien vérifier le gain de qualité apporté lorsque les données sont séparées en échantillon d'apprentissage et échantillon de validation.

Veillez à vérifier l'amélioration pour plusieurs classifieurs.

## 2 Score

**2.1.** Dans un premier temps, attribuez un profit de 1 si le bon chiffre est trouvé et une dépense de 5 si une erreur est commise. Ajoutez un nœud Data Set Attributes pour cela.

**2.2.** Avec le nœud Assessment vérifiez quel est le modèle qui semble le meilleur.

2.3. Choisissez le meilleur réseau de neurones et exportez le code de score associé.

(a) Juste **avant** la ligne : DATA &\_PREDICT ; SET &\_SCORE ; écrivez :

```
%let _PREDICT=x ;  
%let _SCORE=Perso.Zipmontest;
```

Le jeu de données X est créé pendant le scoring. Par défaut il est enregistré dans la bibliothèque WORK.

(b) Juste avant la fin (dernier RUN; QUIT;) écrivez

```
proc sort data=&_PREDICT;  
  by descending EP_VAR257_;  
run;
```

```
proc print data=&_PREDICT;  
  var VAR257 U_VAR257 EP_VAR257_ P_VAR2570 P_VAR2571 P_VAR2572 P_VAR2573  
      P_VAR2574 P_VAR2575 P_VAR2576 P_VAR2577 P_VAR2578 P_VAR2579 ;  
run;
```

Essayez de comprendre la signification de ces lignes notamment en lisant les commentaires ajoutés par défaut au code.

Exécutez le programme, que se passe-t-il ?

