

# Dossier de Réseaux de Neurones

MASTER ESA  
Abdoul Aziz NDOYE

Date de remise le 30 avril 2020

## Consignes

Le projet peut être réalisé **seul** ou en **binôme**. Chaque groupe doit rendre un dossier contenant les sources et code des programmes réalisés. Le dossier doit contenir un compte rendu et une analyse de la mise en œuvre des méthodes développées. Le projet peut être réalisé sous **R**, **Python** ou **SAS** (Vous pouvez utiliser SAS Studio et l'interface SAS Viya).

## Objectif du projet

L'objectif de ce projet est la mise en oeuvre d'un réseau de neurones pour modéliser la décision d'octroi de crédit. Cette modélisation sera appliquée sur les données de crédit ("credit\_scoring") que vous trouverez en pièce jointe avec le descriptif des variables.

Ces données sont constituées de 1000 exemples de demandeurs de crédits dont 700 solvables. Elles disposent de 21 variables décrivant les caractéristiques et les informations personnelles de chaque demandeur notamment

- Le détail du compte courant,
- La durée du crédit, L'historique du credit, L'objet du crédit
- Le montant du crédit, Le compte épargne, le statut professionnel, l'âge, le type logement
- ...

Vous partitionnez votre base de données en deux sous-ensemble. Un ensemble d'apprentissage comprenant 70% des données (700 individus) et un

ensemble de validation comprenant le reste (300 individus), et cela de manière stratifié, c'est-à-dire que dans les deux sous-ensemble, la part des individus ayant un défaut sur le remboursement de leur prêt est conservé par rapport à la population d'origine.

## **Mise en œuvre**

Vous construisez un PMC (Perceptron Multi-Couches) en choisissant l'architecture qui s'adapte mieux à vos données en déterminant le nombre de couche caché optimal, mais également le nombre de neurones de chacune de ces couches et le taux d'apprentissage. Il s'agira donc de choisir plusieurs PMC avec un nombre différent de couches (et de neurones) cachées, de les comparer afin d'avoir la meilleure architecture possible. Vous discuterez de vos choix suivant le critère de sélection utilisé.

## **Comparaison de modèles**

Vous évaluez ensuite la performance de votre réseau de neurones construit que vous comparez aux performances des méthodes suivantes :

- la regression logistique classique et,
- les arbres de décisions,
- les forêts aléatoires,
- le gradient boosting,
- SVM