

Mini Projet

Date de Remise du Mini-Projet: Jeudi 11 Juillet 2019 (14h30)

But du Projet:

Le but de ce projet est de vous permettre d'approfondir votre compréhension des Réseaux de Neurones (RN) en résolvant un problème tout en utilisant cette technique d'apprentissage automatique et en utilisant le modèle entraîné dans une application de votre choix.

Le travail DOIT être fait en binômes, ce qui veut dire exactement deux personnes pourront travailler ensemble sur ce devoir, ni plus ni moins.

Il vous est demandé ce qui suit :

1. Commencez par une lecture des descriptifs des différents ensembles de données (datasets) suggérés plus bas (sur le site UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=dateDown&view=table>) et faites des recherches rapides sur Internet sur chaque thème pour comprendre de quoi il s'agit.
2. Envoyez-moi un email au plus tard le Mercredi 19/06/2019 à 23h59 avec les deux datasets sélectionnés de la liste A à F donnée plus bas (N° 1 sera celui que vous aimeriez le plus et N° 2 le suivant). Votre choix devrait, par exemple, être **1. D ; 2. B**. Mentionnez aussi dans le même email le nom de votre partenaire (binôme) pour ce mini-projet.
3. Je vous confirmerai par email l'assignation d'un ensemble de données par binôme au plus tard le Vendredi 21/06/2019 in chaa Allaah. Je ferai de mon mieux pour vous assigner un dataset parmi ceux que vous aurez choisis, mais ceci n'est pas une garantie. (Au pire des cas, pour avoir une distribution uniforme de binômes/datasets, je ferai un tirage au sort juste et équitable.) Ceux qui n'exprimeront pas leurs choix, auront des problèmes assignés pris de ceux les moins choisis.
4. Faites la conception de votre solution et implémentez la en utilisant un script sur Matlab ou tout autre environnement qui vous permet de faire de l'apprentissage automatique avec réseaux de neurones. (Dans le cas où ce n'est pas Matlab, vous devez m'en informer au plus tard le 23/06/2019 et avoir mon accord dans les 48 heures qui suivront.)
 - a. Présentez le script qui vous permettra de varier la configuration du RNs (architecture, fonctions d'apprentissage, et tous les hyperparamètres que vous voudrez/pourrez.) et d'afficher les résultats obtenus dans chacun des cas, en vous assurant que vous sauvegarderez le réseau de neurones qui vous aura donné les meilleurs résultats.
 - b. Vous pouvez, dès à présent, commencer l'apprentissage sur nntool, le temps de comprendre le scripting. Vous pourrez alors finaliser votre script et l'utiliser sur des données que vous aurez déjà attaquées sur nntool. (Mais c'est juste un conseil ; vous pourrez attaquer le mini-projet comme vous voudrez. assurez-vous seulement que vous respecterez ce qui est exigé dans cet énoncé.)
5. Réfléchissez à une exploitation pratique de votre solution par réseaux de neurones et développez une application dans n'importe quel langage de programmation de votre choix qui permettra de réaliser votre idée. Le principe est que la classification se fera immédiatement pour n'importe quelle donnée qui sera présentée à l'application et le résultat sera exploité dans le sens de l'idée que vous aurez. A noter que votre application peut être basée sur le web, sur smartphone, ou en tant que programme « standalone ». La note sur cette partie prendra en considération l'innovation dans l'idée de votre application. Plus l'idée sera évidente, moins bonne sera la note sur cette partie.

A vous d'innover ! ;-)

Datasets proposés:

- A. **Drug Review Dataset (Drugs.com) Data Set** (2018) 63000000 instances; 6 attributs
- B. **Student Academics Performance Data Set** (2018) 300 instances; 22 attributs
- C. **Online Shoppers Purchasing Intention Dataset Data Set** (2018) 12330 instances; 18 attributs
- D. **WESAD (Wearable Stress and Affect Detection) Data Set** (2018) 215063 instances; 12 attributs
- E. **Detection_of_IoT_botnet_attacks_N_BaIoT Data Set** (2018) 7062606 instances; 115 attributs
- F. **Sports articles for objectivity analysis Data Set** (2018) 1000 instances; 59 attributs

Livrable :

1. Soumettre un rapport écrit qui :
 - a. rappelle le problème traité, les données choisies, le format de ces données, le traitement qui a été fait sur les données, la représentation des données (si elle est modifiée pour être plus appropriée à une utilisation avec les réseaux de neurones), etc. Vous devrez vous assurer que votre rapport définit clairement le problème que vous voulez résoudre ainsi que la raison de votre choix,
 - b. explique avec tous les détails nécessaires vos choix de conception du RN, et les résultats, y compris des tableaux récapitulatifs/comparatifs,
 - c. inclut les lignes directrices à suivre pour utiliser votre solution, et
 - d. dit qui a fait quoi de façon très précise (**des points seront déduits si la distribution des responsabilités n'est pas précisée dans le rapport.**) Chaque membre d'une équipe doit totalement maîtriser le projet aussi bien que les parties qu'il/elle a traitées. (Je n'aime pas les phrases comme : « nous avons tout fait ensemble » !)

Vous remettrez un rapport sous forme de document imprimé ainsi qu'un CD qui contiendra :

- votre rapport en format pdf,
- le jeu de données (clairement expliqué dans le rapport) ainsi que les données après traitement (le cas échéant),
- le script/programme pour créer différents réseaux de neurones et trouver la meilleure architecture et configuration pour le problème que vous avez choisi,
- les captures écrans nécessaires sauvegardées dans des répertoires bien structurés. Les noms des (fichiers des) captures et des répertoires doivent être aussi intuitifs et parlants que possible,
- les codes source et exécutable de votre application qui est construite pour exploiter le réseau de neurones, et
- une « Déclaration sur le plagiat et la malhonnêteté intellectuelle » remplie au stylo et signée par les deux étudiants. J'ai horreur du plagiat (copier-coller) : apprenez à compter sur vos propres efforts, développez vos compétences, et évitez de sérieux problèmes !!

Date de remise du rapport avec le CD : Jeudi 11 juillet, 2019 à 14h30.

Pour chaque 24h de retard à partir de cette échéance, 25% de la note seront déduits. Aucun rapport ne sera donc accepté après Lundi 15 juillet 2019 à 14h30.

Remarque importante:

Un bon rapport n'est pas nécessairement long ! A vous de juger comment avoir un rapport suffisamment complet qui reste aussi court que possible. (En d'autres termes, il ne s'agit pas de faire du remplissage mais bien de faire une présentation écrite, scientifique et aussi claire que possible). Pour vous aider, je limite la longueur du rapport à 10 pages avec une police Times New Roman, taille 11. (Les captures écrans, si vous en avez, doivent être incluses sur CD comme expliqué plus haut et des tableaux comparatifs inclus dans le rapport. Indiquez avec chaque entrée/résultat dans le tableau si la capture écran est incluse ou pas).

2. Il vous sera demandé de faire une démo de votre travail, i.e. le script de création et d'apprentissage du réseau de neurones ainsi que l'application qui l'exploite. Chaque équipe aura une 15aine de minutes pour présenter sa solution et répondre aux questions. Vous devez parfaitement maîtriser le problème et les différents détails de la conception de la solution. La date des présentations vous sera communiquée ultérieurement.

La note de votre équipe au projet sera relative à celle des autres binômes qui auront le même dataset!! De plus, La note de chaque membre d'un binôme ne sera pas nécessairement la même, chacun(e) selon sa contribution et sa compréhension du problème, de la solution, ses réponses aux questions lors de la présentation, etc. (Ne me dites pas « j'ai travaillé sur le mini-projet d'un autre module et il/elle a travaillé sur celui-ci !! »)

Notation du mini-projet:

Partie du mini-projet	Pourcentage de la note du mini-projet
Rapport	35%
Implémentation et résultats du script Réseau de Neurones	40%
Idée et implémentation de l'application	25%

N.B. :

Je n'accepterai aucune remise du projet ou parties du projet par email; vous devrez tout remettre dans ma boîte postale au niveau du département.

Bonus :

Toute innovation intéressante (non requise dans cet énoncé) sera récompensée. A vous de la mettre en relief et de l'expliquer.

Bon courage et.... surtout... Enjoy it ! 😊