

Enseignant(s)

VIDAL Nicolas

Email(s)

nvidal@myges.fr

4A IBD - Deep Learning - Kaggle Dataset Analysis with Keras

1 Matières, formations et groupes

Matière liée au projet :

Formations : -

Nombre d'étudiant
par groupe :

3 à 4

Règles de constitution des groupes: **Imposé**

Charge de travail
estimée par étudiant :

30,00 h

2 Sujet(s) du projet

Type de sujet : **Imposé**

3 Détails du projet

Objectif du projet (à la fin du projet les étudiants sauront réaliser un...)

Utiliser le Framework Tensorflow 2.10 pour composer des modèles issus du Deep Learning et obtenir des résultats proches de l'état de l'art sur différents Datasets. Un soin tout particulier sera à apporter à l'étude des hyper paramètres des modèles étudiés.

Descriptif détaillé

Après avoir installé et configuré Tensorflow 2.10 sur leur machine, ou en utilisant Google Colaboratory les étudiants devront procéder à l'étude de différents modèles sur des Datasets en faisant varier les différents hyper paramètres de ces derniers :

- * Modèles précédents
 - Modèle Linéaire
 - Perceptron Multicouches
- * Nouveaux modèles
 - Conv Net(s)
 - ResNets / HighwayNets
 - RNN(s)
 - Transformers

Il sera fortement conseillé aux étudiants de s'appuyer sur Tensorboard pour visualiser comparer et retranscrire les performances de ces modèles.

Il sera important de faire apparaître pour chaque dataset, pour chaque modèle :
L'influence de tous les hyperparamètres des modèles (structure, fonctions d'activation, ...), ainsi que des paramètres des algorithmes d'apprentissages (learning rate, momentum, optimizer, ...)

Etude d'un ou plus datasets Kaggle parmi :

Pressentis pour cette année (pas encore certain, mais vous pouvez commencer à vous y entraîner) :

<https://www.kaggle.com/competitions/nfl-big-data-bowl-2023/overview/timeline>

<https://www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/overview/timeline>

<https://www.kaggle.com/competitions/goodreads-books-reviews-290312/overview/timeline>

ATTENTION les deadlines de ces compétitions sont début Janvier !

(juste pour s'entraîner, ne peut pas être choisi seule :)

<https://www.kaggle.com/competitions/novozyms-enzyme-stability-prediction/overview/timeline>

Pour information, l'année précédente le choix se portait sur : 2021/2022

(Entraînement)

<https://www.kaggle.com/c/tabular-playground-series-jan-2022/overview>

(Difficile)

<https://www.kaggle.com/c/jigsaw-toxic-severity-rating>

(Très Difficile)

<https://www.kaggle.com/c/g-research-crypto-forecasting/overview>

(Très Difficile)

<https://www.kaggle.com/c/tensorflow-great-barrier-reef/data>

Ouvrages de référence (livres, articles, revues, sites web...)

<http://www.deeplearningbook.org/>

<https://www.tensorflow.org/>

<https://keras.io/>

Outils informatiques à installer

Keras / Tensorflow 2.10 / Jupyter / PyCharm / Colab

1	Etape intermédiaire	Première(s) soumission(s) à la compétition Rendu : notebook Jupyter + pdf / Présentation	lundi 28/11/2022 23h59
2	Rendu final	Soutenance + Rapport final Rendu : notebook Jupyter + pdf / Présentation	dimanche 26/02/2023 23h59

5 Soutenance

Durée de présentation
par groupe :

15 min

Audience : **A huis clos**

Type de présentation :

Présentation / PowerPoint - Démonstration

Précisions :