**Master 2 Économétrie et Statistiques, parcours Économétrie Appliquée**

**SVM et réseau de neurones**

*Gabin Lagre et Astrid Valicon*



Année 2021/2022

Sommaire

[I. Introduction 4](#_Toc85033558)

[II. Présentation et préparation des données 4](#_Toc85033559)

[III. Méthodologie utilisée 6](#_Toc85033560)

[IV. Application et modélisation 6](#_Toc85033561)

[V. Conclusion et discussion 6](#_Toc85033562)

# **Introduction**

Une onde gravitationnelle s’apparente à une ondulation dans l’espace-temps. Cette ondulation est causée par des par des processus très violents et energétique de l’Univers. La fusion de deux trou noirs en fait partie!

On peut reprendre ce qui est écrit dans “overview” et le traduire en français en vrai c’est une bonne intro je trouve.

Lien Kaggle : <https://www.kaggle.com/c/g2net-gravitational-wave-detection>

# **Présentation et préparation des données**

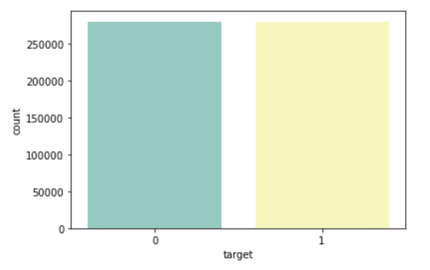
Nous avons à notre disposition quatre fichiers. Les deux premiers sont des dossiers comprenant des fichiers .npy qui regroupent eux même des séries chronologiques. Ces dernières contiennent des mesures simulées d’ondes gravitationnelles à partir d’un réseau de trois interféromètres à ondes gravitationnelle. Il s’agit de LIGO Hanford, LIGO Livingston et Virgo. Chacune des séries temporelles (une pour chaque détecteur) contiennent soit le bruit du détecteur, soit le bruit du détecteur plus un signal d’onde gravitationnelle simulé. Nous précisons également que chaque cycle dure deux secondes et est échantillonné à 2048 Hz. Notre tâche consiste à déterminer quand le signal est présent dans les données (lorsque « target » = 1). Cette problématique consiste en l’application d’une classification binaire, si le signal est détecté ou non. Le fichier « train » est le fichier sur lequel nous allons travailler. Le fichier « test » nous permettra de comparer nos prévisions.

Il existe plusieurs variables pouvant impacter la forme des ondes gravitationnelles. Il s’agit de la masse, l’emplacement du ciel, la distance, les spins de trou noir, l’angle d’orientation binaire, la polarisation gravitationnelle d’onde, l’heure d’arrivée et la phase à coalescence (fusion). Ces paramètres ne sont pas intégrés dans notre jeu de données mais ont été randomisés et utilisés pour générer les signaux présents dans la base.

Les deux autres fichiers sont des fichiers .csv. L’un correspond aux valeurs cibles indiquant si le signal associé contient une onde gravitationnelle (target = 1) ou non (target = 0). Ce sont les données dites d’apprentissage. L’autre correspond aux données test, qui nous permettra de vérifier à la fin nos résultats et les comparer.

Nous commençons par importer nos données d’apprentissage en format .csv (que nous appelons « train\_label »), puis en format .npy (que nous appelons « train\_path »).

***Train\_label*** : Cette base comporte 560000 observations et 2 colonnes. La première colonne correspond à l’identifiant (id) et la seconde correspond à la valeur cible, qui prend la valeur 0 ou 1. Il n’y a pas de valeurs nulles. Grâce à la commande value\_counts(), nous pouvons voir que 280070 observations prennent la valeur 0 et 279930 prennent la valeur 1 *(cf graphique 1)*. Il y a donc au total, 279930 signaux d’ondes gravitationnelles dans la base de données, ce qui représente 49,98% des données.



*Graphique 1 : Répartition des cibles = 0 et cibles = 1 dans nos données*

***Train\_path*** : Ces données ont été importées grâce à la librairie glob. Il y a 560 000 fichiers .npy dans ce dossier. Nous en choisissons un pour l’observer. L’échantillon comporte 3 lignes car ce sont les séries chronologiques de 3 interféromètres à ondes gravitationnelles évoqués plus haut. Il y a 4096 colonnes. Ces chiffres sont les mêmes pour l’ensemble des fichiers .npy. Chaque fichier .npy (comprenant 3 lignes et 4096 colonnes) et assigné à une valeur cible (soit 1 soit 0).

# **Méthodologie utilisée**

Carte trop bien pour savoir quelle méthode utilisée 🡪 <https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html>

SVM :

* <https://scikit-learn.org/stable/modules/svm.html#classification>

Réseau de neurones

# **Application et modélisation**

# **Conclusion et discussion**