## Projet de 2e année

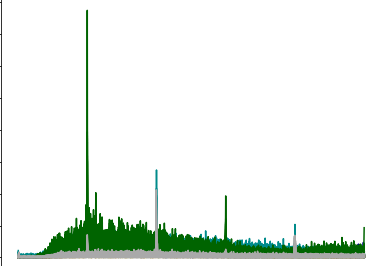
***Sujet – Détection de points d’impact dans des courbes pour la classification***

Nom du tuteur : Steven Golovkine

Entreprise ou organisme : Renault SA – Direction Qualité et Satisfaction Client – Pôle Data science

1. Contexte

Avec les récentes avancées technologiques, de plus en plus d’objets sont équipés de capteurs leur permettant, par exemple, de connaître la position d’autres objets dans leur environnement. Ces capterus fournissent un grand nombre de signaux pouvant être modélisés. On appelle ce genre de données, des données fonctionnelles. Ainsi, dans ce cas, une observation ne sera pas un vecteur mais une courbe. Renault, en tant que constructeur automobile, enregistre un grand nombre de données provenant des différents capteurs montés sur la voiture. Leur étude est donc indispensables. Celles-ci peuvent poser différents problèmes pour leur analyse : bruit, différents échantillonnages, points impactant, multidimensionnelles, … Dans ce projet, nous nous intéresserons plus particulièrement à la détection de points d’impact en vue de leur classification.



*Exemple de données fonctionnelles avec points d’impact*

1. Sujet proposé

Pour ce projet, on s’intéresse plus particulièrement à la détection de points d’impact dans les courbes pour la classification. Il peut arriver, lors de l’enregistrement de signaux, que des valeurs *a priori* aberrantes apparaissent. Cependant, ces valeurs peuvent indiquer un phénomène particulier. Par exemple, dans le cas de signaux indexés, non pas par le temps, mais par la fréquence, cela montre que quelque chose se produit à une certaine fréquence. Dans le cas, où le phénomène physique sous-jacent est connu, il est assez facile de connaître quels points peuvent être impactant sur la courbe. Mais, dans le pratique, ces points ne sont pas connus. De plus, ces points peuvent aussi ne pas apparaître au même instant pour chaque observation. L’objet de ce projet est de construire un modèle de classification automatique de signaux en se basant sur la détection des points d’impact. Celui-ci s’appuiera sur les données *open-*source du jeu de données *SemgHandMovementCh2* du *UEA & UCR Time Series Classification Repository*. Une description ainsi qu’un *pre-processing* des données sont données ici :

[*https://colab.research.google.com/drive/1nwrYtpPvQO-2AczbSzGV5-OEwvZ88o\_X*](https://colab.research.google.com/drive/1nwrYtpPvQO-2AczbSzGV5-OEwvZ88o_X)

1. Méthodologie envisagée

Méthodes d’apprentissage supervisée adaptées aux données fonctionnelles. Test de diverses méthodes pour la détection de point d’impact (détection de rupture de pente, détection d’*outliers,* …). Renault fournira un accompagnement méthodologique pour appréhender les spécificités liées à l’analyse des données fonctionnelles (*smoothing*, ACP fonctionnelle, régression fonctionnelle…).

1. Résultats attendus

Le rapport du projet devra comporter une comparaison des résultats de classification des différentes méthodes de détection de points d’impact. Le script R ou Python devra être fourni et reproductible.

1. Bibliographie

* Ramsay and Silverman, Functional Data Analysis, 2005, Springer
* Ramsay, Hooker and Graves, Functional Data Analysis with R and MATLAB, 2009, Springer
* Gijbels, Hall and Kneip, On the Estimation of Jump Points in Smooth Curves, Ann. Inst. Statis. Math., Vol. 51, No. 2, 231-251 (1999)
* Kneip, Poss and Sarda, Functional Linear Regression with Points of Impact, The Annals of Statistics, 20616, Vol. 44, No. 1, 1-30