## Projet de 2e année

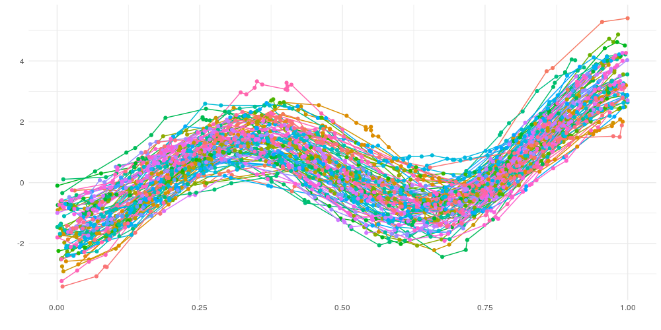
***Sujet – Classification de courbes après recalage***

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 de leur environnement. Ces capteurs 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 par 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. Celles-ci peuvent poser différents problèmes pour leur analyse : bruit, différents échantillonnages, *time lag*, multidimensionnelles, … Dans ce projet, nous nous intéresserons plus particulièrement au *time lag* et au recalage de courbes pour la classification.



*Example de données fonctionnelles – Chaque courbe représente une observation*

1. Sujet proposé

L’objet de ce projet est de construire un modèle de classification automatique de signaux en recalant les courbes pour améliorer les résultats de classification. Celui-ci s’appuiera sur les données *open-source* du jeu de données *CharacterTrajectories* du *UEA & UCR Time Series Classification Repository.* Une description ainsi qu’un pré-processing des données sont donnés ici :

https://colab.research.google.com/drive/1UL99rv\_d9ZYxMLphInirvObiHElt7DWg

1. Méthodologie envisagée

Méthodes d’apprentissage supervisée adaptées aux données fonctionnelles. Test de diverses méthodes de recalage, transformation linéaire des courbes ou usage de la distance DTW, par exemple. Renault fournira un accompagnement méthodologique pour appréhender les spécificités liées à l’analyse des données fonctionnelles (*smoothing*, ACP fonctionnelle, …).

1. Résultats attendus

Le rapport du projet devra comporter une comparaison des résultats de classification des différentes méthodes de recalage testées. 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
* Ramsay and Li, Curve Registration, J. R. Statist. Soc. B (1998) 60, Part 2, pp. 351-363
* Kneip, Li, MacGibbon and Ramsay, Curve Registration by Local Regression, The Canadian Journal of Statistics, Vol. 28, No. 1 (March., 2000), pp. 19-29
* Petitjean. Description des alignements formés par DTW. 2011. Hal-00647522