Université de mathématiques Aix-Marseille

et

DreamQuark

Etudes comparatives des méthodes algorithmiques d'hyperoptimisation

Master d'Ingénieurie Mathématiques et Modélisation

Année 2019/2020

Auteur :
Elie Lagache

Responsable de master :
Anne Nouri

Responsable du stage :
Sebastian Mueller

Responsable professionnel
Alexandre Cameron

"The human condition is plagued with a labyrinth of shortcomings, frailties and limitations that hinder man from reaching his fullest potential. Therefore, it only makes sense that we find ourselves at the next phase in human evolution where restricted man merges with the infinite possibilities of hyper-evolving technologies. This techno-human transmutation will prove to be 'the' quantum leap in human progression. The harmonization of technologically extending oneself, consciousness, artificial intelligence and machine learning will reverse the failures of genetic predisposition and limitation. " James Scott

Remerciements

Je tiens à remercier Monsieur Sebastian Mueller, Maitre de conférence à l'Université de d'Aix-Marseille, pour sa disponibilité et son encadrement pédagogique.
Je remercie, de plus, Alexandre Cameron, Data Scientist Chief de l'équipe project au sein de Dreamquark, avec qui j'ai réalisé la majeure partie du projet de mon mémoire et qui m'a accompagné pendant toute la durée de mon stage.
Je n'oublie pas de remercier ma responsable de master Mme Anne Nouri qui pu faire en sorte que je puisse finaliser et passer cet UE mémoire dans de bonne condition. Je remercie enfin, l'équipe project de Dreamquark avec qui j'ai participé à plusieurs études intéréssantes en lien avec le sujet du mémoire.

Table des matières

Chapitre 1

Introduction

Nous introduirons ce mémoire présentant tous les points essentiels du stage réalisé ainsi qu'un exemple introductif du sujet traité.

1.1 Machine learning

Le machine learning (apprentissage automatique ou statistique), représente une étude de l'intelligence artificielle. Il s'agit de laisser évoluer un processus algorithmique par l'intermediaire d'une "machine" afin de resoudre une problématique statistique (le plus souvent) difficile.
L'analyse réalisée est généralement associée à des arbres ou des courbes (par exemple la fonction de perte) et les problèmatiques à de la classification (binaire ou multi-classes). Il s'agit de l'étiquettage des données à une certaine classe. Un modèle de regression logistique en est un exemple classique.

1.2 Présentation du stage

1.2.1 Environnement professionel

L'entreprise d'acceuil est spécialisée dans l'intelligence artificielle concernant des problématiques de finance et d'assurance. En effet, Dreamquark developpe une machine virtuelle d'analyse prédictive utilisable par des personnes métiers (banques, assurances, marketing,..) sans qu'il soit nécéssaire de posséder forcément des connaissances appronfondies en data science. Les calculs sous-jacents découlent surtout des algorithmes de deep learning (apprentissage profond : réseaux de neurones). Le machine learning classique sera utilisé afin de comparer les performances réalisées par le produit developpé et c'est ainsi dans ce cadre que j'ai effectué mes travaux lors de ce stage.

1.2.2 Objectifs et sujets principaux

Le stage a comporté deux aspects principaux :

1. un travail de recherche théorique,
2. un travail opérationnel de test ou d'application de la théorie introduite.

Voici quelques exemples de sujets sur lesquels j'ai travaillé durant ce stage : les algorithmes de gradient boosting, **les algorithmes d'optimisation d'hyperparamètres**, la synthétisation de données pour la simulation de problèmatiques réelles, la detection de biais dans les modèles de machine learning. Ce mémoire portera essentiellement sur les algorithmes d'optimisation d'hyperparamètres.
La partie opérationnelle c'est déroulée essentiellement par l'intermédiaire de l'outil Python et de ces modules statistiques.