

Introspection

Benjamin GUIGON

January 2021

1 Lien

Article écrit par Benjamin GUIGON

<https://github.com/benjaminiguigon/PSBX/blob/main/Article%20ACP%20.pdf>

2 ACP

J'ai choisi de rédiger un article sur l'*ACP* car c'est une technique de machine learning indispensable qu'il faut connaître. On peut penser a priori (je le pensais), que l'aspect mathématique de l'*ACP* était facile. Cependant en lisant l'article que j'ai résumé, je me suis rendu compte que ce n'était pas du tout trivial et que comprendre le principe général était assez important.

Après avoir résumé 10 travaux de mes camarades, je peux maintenant faire un point sur mon travail au regard de ceux des autres. Certains de mes camarades ont fait des travaux remarquables. Ils étaient clairs, précis, fournis et surtout il y avait des sujets très bien choisis car ils rentrent parfaitement dans la ligne d'apprentissage de certain d'entre nous.

Mon article sur l'*ACP* est mathématiquement plus compliqué que beaucoup de travaux de mes camarades. C'est à la fois un point positif et négatif. En effet le côté mathématique est très présent avec des lignes de calculs pas évidentes ainsi que des principes non triviaux, cependant notre classe n'a pas un niveau homogène en mathématique. Donc mon travail est peut être trop axé sur les mathématiques et pas assez sur la compréhension générale. Cependant il permet aux plus curieux de comprendre le principe de l'*ACP* et les sous jacents mathématique qui la régissent.

J'aurais peut être pu aussi appliquer un cas concret d'une *ACP* sur R ou python en parallèle de mon explication mathématique.

3 XGBoost

Il est depuis quelques années très populaires en Data science d'utiliser l'algorithme de la descente de gradient pour résoudre des problèmes complexes. Cependant, depuis quelques années, un nouvel algorithme appelé XGBoost est sous le feu des projecteurs grâce à ses caractéristiques qui le rendent un des plus puissants pour beaucoup de problèmes.

J'ai choisi ce sujet car je voulais absolument maîtriser cette technique de machine learning qui est d'un côté intéressante et d'un autre important de maîtriser pour arriver avec de solides bases dans le monde du travail.

Mon travail sur XGBoost sur R est très complet car j'ai fait plusieurs exemples, sur plusieurs datasets et avec plusieurs lignes directrices. C'est donc un travail qui résume en long et en large l'utilisation de XGBoost sur des exemples concrets. Cependant, mon script est par moment axé pour des personnes qui connaissent certains principes de machine learning ou de pré-processing. J'ai fait le choix de faire plusieurs scripts et de laisser aux lecteurs l'autonomie de se renseigner sur les aspects plus théorique (*dummies, sparse matrix etc...*

Le package XGBoost peut être très long à comprendre par nous même (comme beaucoup de packages) mais ce type de travail permet de passer beaucoup de temps sur l'article que nous écrivons mais d'avoir en contre partie un gain de temps sur le travail des autres élèves. Je voulais apprendre le package sur les arbres de décision Rpart et grâce à mes camarades j'ai un article résumé avec un exemple concret pour me familiariser avec le sujet avant d'apprendre plus en détail de mon côté.