

TP – Introduction à l’Apprentissage Statistique
 ISFA Lyon, Master 2 Actuariat, 2021/2022
 Pierrick Piette

TP 4 – XGBoost & Shap Values

Librairies nécessaires

xgboost

Données : [MedicalPremium.csv](#)

Même base de données qu’au TP 3 et 4

Premier XGBoost

1. Quels sont les nouveaux hyperparamètres à tuner par rapport au TP 3 ?
2. S’approprier l’impact d’un changement des hyperparamètres suivants sur le nombre d’arbres optimal : gamma, subsample, colsample_bynode, lambda, alpha.
3. Interprétez.

Tuning paramètre stochastique et gamma

1. On fixe lambda = 1, alpha = 0, max_depth = 3, eta = 0.01. Tuner les hyperparamètres subsample, colsample_bynode et gamma tel que subsample = colsample_bynode.
2. Quel est le temps moyen de l’évaluation d’un couple ?
3. Quel serait le temps d’évaluation si on testait l’ensemble des hyperparamètres ensemble en *grid search* avec 5 possibilités sur chacun ?

Valeurs de Shapley

1. Estimer l’XGBoost avec les hyperparamètres trouver précédemment sur l’ensemble de la base.
2. Avec xgb.plot.shap.summary analyser l’impact des variables.
3. Avec xgb.plot.shap analyser en détails les différentes variables.

Comparaison des interprétation

1. Comparer les valeurs de shapley avec les partial dependence plot pour l’XGBoost.
2. Comparer les partial dependence plot entre l’XGBoost et la forêt aléatoire.