# TP 0 Régression régularisée sur data set prêt_bancaire

En utilisant Python et les bibliothèques nécessaires, l'objet de ce TP est de réaliser des modèles de régression linéaire permettant de prévoir le montant d'un pret accordé à un client en fonction de son profil

1- **Analyse des données (EDA : Exploratory Data Analysis):**

A partir du fichier credit_immobilier_ISF.csv fourni :

-Effectuer les différentes étapes qui concernent la préparation des données en prévision des modélisations : Statistiques élémentaires, données manquantes, outliers, visualisations, ….

Commenter et justifier vos choix

**2- Modélisation : Montant du prêt fonction des features :**

Le label est la variable : **montant**

Découper le data set en train data set et test data set (30 % des données)

-**Régression linéaire classique :**

Donner la mse du test data set et du train data set ainsi que la rmse, les R2 et R2 ajustés

Commenter ces résultats

-**Régression régularisée Ridge :**

Faire varier l'hyperparamètre alpha (lambda) sur une plage de valeurs adéquates et représenter l'évolution des différents poids attribués aux variables explicatives sur un graphique : poids = f(alpha)

Représenter sur un graphique :

L'évolution de la mse du test data set en fonction de alpha

L'évolution du R2 du test data set en fonction de alpha

Déterminer automatiquement la valeur optimale de alpha avec la méthode Ridge Cross validation. Indiquer la valeur optimale obtenue

Indiquer la mse, la rmse, le R2 et les coefficients (poids) des features obtenus avec la regression Ridge utilisant cet alpha optimal

Commenter les résultats obtenus

-**Régression régularisée Lasso :**

Faire varier l'hyperparamètres alpha (lambda) sur une plage de valeurs adéquates et représenter l'évolution des différents poids attribués aux variables explicatives sur un graphique : poids = f(alpha)

Représenter sur un graphique :

L'évolution de la mse du test data set en fonction de alpha

L'évolution du R2 du test data set en fonction de alpha

Déterminer automatiquement la valeur optimale de alpha avec la méthode Lasso Cross validation. Indiquer la valeur optimale obtenue

Indiquer la mse, la rmse, le R2 et les coefficients (poids) des features obtenus avec la regression Lasso utilisant cet alpha optimal

Préciser en particulier les features éliminés (poids = 0) via le Lasso et commenter ces résultats

-**Régression régularisée ElasticNet :**

Effectuer la modélisation avec ElasticNet en gardant les valeurs de hyperparamètres par défaut et donner le R2 obtenu pour le train set et pour le test set

Choisir les hyperparamètres alpha et l1_ratio qui vous paraissent les meilleurs, modeliser et donner le R2 obtenu pour le train set et pour le test set

-**Choix des features à conserver par Récursivité :**

Effectuer une sélection automatique des features par récursivité, indiquer le nombre optimal de features à conserver, donner leurs noms et visualisera le score de la cross validation en fonction du nombre de features retenus

-**Hypothèses de validité pour la régression :**

Vérifier si les hypothèses pour la régression linéaire sont respectées :

Erreurs non corrélées, erreurs gaussiennes, homoscédasticité (même variance pour les différents résidus),…

Utiliser les histogrammes, QQPlots qui vous paraissent nécessaires et commenter

Rendre un compte rendu du TP format word ou pdf et le script développé (avec les commentaires nécessaires) tournant soit sous forme jupyter notebook, soit sous forme spyder

Noms des 2 fichiers à envoyer (XX étant le numéro du groupe de TP) :

GRXX_TP_0_Regression_regularisee_ISF_compte_rendu.doc ou .docx ou .pdf

GRXX_TP_0_Regression_regularisee_ISF_script.ipynb ou .py