

TP3: Régression Linéaire

Yannig Goude: yannig.goude@edf.fr

M2 "Mathématiques de l'aléatoire"

Exercice 1: régression linéaire, lien consommation/température

Importer les données `data_conso_hebdo0.txt` déjà étudiées dans le TP1.

1. représenter le nuage de point de la consommation en fonction de la température
2. quelle(s) transformation(s) de la variable température paraît adéquate pour ajuster un modèle de régression linéaire entre ces variables?
3. proposer au moins 3 modèles. Estimer et comparer les différents modèles de régression correspondant (utiliser la fonction `lm`, `summary` et leurs sorties: R^2 , R^2 ajusté...)
4. ajouter au nuage de point les effets estimés par les modèles.
5. tirer aléatoirement $N = 100$ observations et en faire un jeu de données test. Estimer les modèles précédant sur les données restantes (données d'apprentissage) et prévoir la consommation des données de test. Mesurer les performances en prévision de ces modèles en utilisant le critère du RMSE (Root Mean Square Error):

$$RMSE(Y, \hat{Y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

Ce résultat confirme-t-il votre choix de modèle?

Exercice 2: régression linéaire et modélisation de phénomènes périodiques

1. en reprenant les résultats du TP1, construire une variable `Load.detrend` représentant la consommation corrigée de sa tendance (linéaire ou quadratique, cf TP1) et la représenter.
2. pour modéliser des phénomènes cycliques il est naturel d'utiliser une décomposition en série de Fourier. Considérer ici le modèle de régression linéaire suivant:

$$y_t = \sum_{q=1}^K a_q \cos(\omega_q t) + b_q \sin(\omega_q t) + \varepsilon_t \quad (1)$$

avec $\omega = 2\pi/T$ ou T est la période du phénomène cyclique, K , a_q et b_q étant des paramètres à estimer à l'aide des données.

3. écrire un script `r` permettant d'estimer pour $K = 1, \dots, 10$ le modèle 1. Représenter en fonction de K le R^2 ajusté. Selon ce critère, quelle valeur de K choisissez vous?
4. représenter sur un graphique `Load.detrend` et l'estimation de l'effet cyclique obtenu. Zoomer sur la première année.
5. en utilisant la fonction `train` du package `caret` calculer le critère RMSE pour $K = 1, \dots, 10$ par validation croisée V-fold (prendre $V = 10$, utiliser la fonction `lapply`). Représenter ce critère en fonction de K . Cela confirme-t-il votre choix de K ?

Exercice 3: régression linéaire et prévision de consommation

1. en utilisant les même approches que dans l'exercice 1 et 2, proposer un modèle de prévision de la consommation électrique par régression linéaire.
2. analyser en détail le modèle que vous obtenez (analyse des résidus, diagnostics...)
3. importer le jeu de données `data_conso_hebdo1.txt`, effectuer une prévision de la consommation et l'évaluer par le critère du RMSE.

Exercice 4: régression linéaire et prévision de consommation

Reprendre les jeux de données étudiée au TP2 et proposer un modèle de prévision par régression linéaire.