

Python et Machine Learning
TRAVAUX DIRIGÉS
Septième partie

Salim Lardjane

Université de Bretagne Sud

1. Importer sous Python les données ozone.csv du répertoire Programmes sur le Forum. Il s'agit de différentes variables météorologiques relevées dans une ville américaine au cours d'une année. On souhaite prévoir la quantité d'ozone dans l'atmosphère de cette ville (upo3) à partir des autres variables et de leurs interactions.
2. On n'utilisera pas la variable "day". La supprimer du tableau de données.
3. Normaliser l'ensemble des variables explicatives et créer les variables correspondant à leurs interactions deux à deux.
4. Diviser les données en un ensemble d'apprentissage et un ensemble de test.
5. Evaluer la performance des modèles suivants sur l'échantillon test : régression par 5-plus proches voisins, régression linéaire par mco, régression ridge avec la valeur par défaut du paramètre de contrôle, regression lasso avec la valeur par défaut du paramètre de contrôle, régression elastic-net avec les valeurs par défaut des paramètres de contrôle.
6. Quel modèle retenir à ce stade ? Comment améliorer les choses ?