

Apprentissage Statistique Automatique I



Présentation

Description

Programme :

- I. Modèles de régression
 1. Régression linéaire - Régression logistique
 2. Méthodes de pénalisation (LASSO - RIDGE - Elastic Net)
 3. Réduction de la dimension (régression des moindres carrés partiels)
 4. Arbre de régression
- II. Apprentissage supervisé (classification)
 1. Régression logistique
 2. Arbre de classification
 3. Méthode des k plus proches voisins
 4. Méthodes ensemblistes (random forest, bagging, boosting)
 5. Support Vector Machine
- III. Apprentissage non-supervisé (clustering)
 1. k-means
 2. Réduction de la dimension (ACP, ...)
 3. Mélange de gaussiennes

Logiciels :

Les travaux pratiques de ce module seront réalisés avec Python ou R

Objectifs

1. Maîtriser les méthodes classiques d'apprentissage statistique automatique
2. Mettre en œuvre les méthodes d'apprentissage à l'aide du langage Python ou R
3. Identifier les méthodes les plus pertinentes selon le contexte applicatif (apprentissage supervisé et non supervisé)
4. Identifier des situations de sur ou sous apprentissage

5. Calibrer un modèle d'apprentissage et évaluer l'ajustement par validation croisée

Pré-requis obligatoires

Modules SM1 et SM2.

Bibliographie

1. The Elements of Statistical Learning. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Second edition, Springer, 2009.
2. Pattern Recognition and Machine Learning. Christopher Bishop. Springer, 2006.
3. Le modèle linéaire par l'exemple. Jean-Marc Azais, Jean-Marc Bardet. Deuxième édition, Dunod, 2012.
4. <https://scikit-learn.org/stable/Documentation> de scikit-learn