

Machine learning

Visseho Adjiwanou, PhD.

10 June 2024

Objectifs du cours :

- Comprendre les concepts fondamentaux de l'apprentissage automatique.
- Apprendre à implémenter et évaluer différents algorithmes de machine learning en utilisant R.
- Explorer des applications pratiques et avancées du machine learning.
- Développer des compétences en utilisation des packages R dédiés au machine learning.

Prérequis :

- Connaissances de base en R et en statistiques.
- Connaissances en algèbre linéaire et en calcul différentiel sont recommandées.

Session 1 : Introduction et concepts de base

- Introduction à l'apprentissage automatique
- Types d'apprentissage : supervisé, non supervisé, semi-supervisé, par renforcement
- Cycle de vie d'un projet de machine learning
- Préparation des données
- Travail pratique : Introduction aux packages tidyverse et data.table pour la manipulation des données.

Apprentissage automatique supervisé (I)

Session 2 : Régression linéaire et extensions

- Concept de la régression linéaire
- Méthode des moindres carrés
- Évaluation des performances : RMSE, MAE, R^2
- Travail pratique : Implémentation de la régression linéaire avec `lm()` et visualisation avec `ggplot2`.

Session 3 : Régression logistique

- Concept de la régression logistique
- Fonction sigmoïde et probabilités
- Coût et optimisation (descente de gradient)
- Travail pratique : Classification binaire avec la régression logistique en utilisant `glm()`.

Session 4 : Algorithmes de classification

- k-Nearest Neighbors (k-NN)
- Support Vector Machines (SVM)
- Arbres de décision et forêts aléatoires
- Travail pratique : Utilisation des packages `class` pour k-NN, `e1071` pour SVM, et `randomForest` pour les forêts aléatoires.

Session 5 : Évaluation et validation

- Techniques de validation croisée
- Métriques de performance : précision, rappel, F1-score, courbe ROC-AUC
- Problèmes de surapprentissage et sous-apprentissage
- Travail pratique : Utilisation du package `caret` pour la validation croisée et l'évaluation des modèles.

Étude de cas 1

Apprentissage automatique non supervisé (II)

Session 6 : Clustering et méthodes non supervisées

- Introduction au clustering
- Algorithmes : k-means, DBSCAN, agglomératif
- Réduction de dimensionnalité : PCA, t-SNE
- Travail pratique : Application du clustering avec `stats` et `dbscan`, et de la réduction de dimensionnalité avec `prcomp` et `Rtsne`.

Session 7 : Traitement du langage naturel (NLP)

- Introduction au NLP
- Représentation des textes : sac de mots, TF-IDF, word embeddings
- Modèles de langage : RNN, LSTM
- Travail pratique : Analyse de sentiments sur des textes avec les packages `tm`, `text2vec`, et `keras` pour les RNN.

Session 8 : Apprentissage par renforcement (2h)

- Concepts de base de l'apprentissage par renforcement
- Environnement, agent, état, action, récompense
- Algorithmes : Q-learning, SARSA
- Travail pratique : Implémentation d'un agent simple en utilisant le package reinforcement-learning.

Session 9 : Réseaux de neurones et deep learning (partie 1) (2h)

- Introduction aux réseaux de neurones
- Perceptron multicouche (MLP)
- Fonction d'activation et propagation avant
- Travail pratique : Implémentation d'un MLP simple avec le package keras.

Session 10 : Réseaux de neurones et deep learning (partie 2) (2h)

- Entraînement des réseaux de neurones (backpropagation)
- Optimisation et régularisation
- Réseaux convolutionnels (CNN) pour le traitement des images
- Travail pratique : Création et entraînement d'un CNN pour la classification d'images avec keras.

Session 11 : Conclusion et perspectives

- Révisions et discussion sur les avancées récentes du domaine
- Perspectives de carrière dans le machine learning
- Présentation des projets de groupe
- Travail pratique : Présentation et évaluation des projets finaux.

Étude de cas 2

Ressources et matériel recommandé :

Livres : “An Introduction to Statistical Learning” de Gareth James, Daniela Witten, Trevor Hastie, et Robert Tibshirani “Deep Learning with R” de François Chollet et J. J. Allaire
Sites web : R-bloggers Coursera Kaggle