

Fiche de synthèse:

Nom et Prénom: BENAMROUCHE Nadjib.

Objectif: Compte rendu

Au cours de ce cours de “fondement d’apprentissage automatique” on a vu les différentes techniques et modèles du Machine Learning.

On va commencer par ces trois types:

- L’apprentissage supervisé: consiste à superviser l’apprentissage de la machine en lui montrant des exemples (des données) de la tâche qu'elle doit réaliser.
- L’apprentissage non supervisé: l’algorithme doit opérer à partir d’exemples non annotés.
- L’apprentissage semi-supervisé.

Alors en suivant le processus suivant nous pouvons faire de la prédiction ou la classification à partir de nos données:

Data -> preprocessing -> extraction de caractéristique -> algorithme de classification/prédiction -> matching -> sortie.

Les différents modèles/méthodes de Machine Learning:

- Régression linéaire: utilisée principalement pour la prédiction [comment calculer la covariance et faire la droite de la ligne de la régression linéaire et le calcul des résidus].
- Régression logistique: utilisée pour la classification, elle peut être binomiale ou multiples.
- Arbre de décision: est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre (utilisé dans le cas de classification et de prédiction).
- Réseau de neurones: ce type de réseaux est généralement utilisé pour des problèmes complexes, ils ont une bonne précision par rapport aux autres modèles ML (si les problèmes à gérer sont des tâches complexes). Dans ce type d’algorithme et dans le cas supervisé on a des entrées et pour chaque entrée on a des poids et des biais qui serviront à orienter la solution du modèle par la suite avec par exemple un algorithme de descente de gradient. On calcule ensuite le taux d’erreurs et en cas de mauvais résultats (pas de bonne précision) on ajuste au fur et à mesure de la descente du gradient les poids des poids et des biais afin de converger vers la bonne solution.

Nous avons fait aussi l’ACP pour la réduction des dimensions (variables/individus) et cela en concernant le niveau de prédiction. L’ACP nous aide pour la compréhension des corrélations entre les variables.

Partie pratique: se familiariser avec le langage de programmation R avec RStudio et Jupyter notebook et ces différentes bibliothèques de machine learning tel que

- dplyr est une bibliothèque de manipulation de données.
- ggplot2 boxplot barplot etc sont des bibliothèques de visualisation de données.
- rpart qui propose une implémentation des méthodes de construction d'arbres de décision.