

Introduction au Machine Learning

ARN

1 - Introduction

Abstract

Definition

Table des matières

1. Qu'est-ce que le Machine Learning ?	2
1.1. Comment faire du Machine Learning ?	2
1.2. Overfitting et Underfitting	2
1.2.1. Overfitting	2
1.2.2. Underfitting	2
2. Conclusion	3

1. Qu'est-ce que le Machine Learning ?

Le Machine Learning (ML) est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur la création d'algorithmes et de modèles permettant aux ordinateurs d'apprendre à partir de données. Au lieu de programmer explicitement un système pour effectuer une tâche, le ML utilise des données pour entraîner un modèle qui peut ensuite faire des prédictions ou prendre des décisions basées sur de nouvelles données.

Il existe 3 catégories principales de Machine Learning :

1. **Apprentissage supervisé** : Le modèle est entraîné sur un ensemble de données étiquetées, où chaque entrée est associée à une sortie connue. L'objectif est de prédire la sortie pour de nouvelles entrées.
 - Permet de résoudre des problèmes de classification et de régression.
 - Exemples : classification d'images, prévision de ventes, détection de fraudes.
2. **Apprentissage non supervisé** : Le modèle est entraîné sur un ensemble de données non étiquetées, et il doit découvrir des structures ou des motifs sous-jacents dans les données.
3. **Apprentissage par renforcement** : Le modèle apprend à prendre des décisions en interagissant avec un environnement, en recevant des récompenses ou des pénalités en fonction de ses actions.

1.1. Comment faire du Machine Learning ?

En premier il est important d'explorer les données pour comprendre leur structure et leur contenu. Cela peut inclure des étapes telles que la visualisation des données, le nettoyage des données (suppression des valeurs manquantes ou aberrantes). Ensuite, il faut passer par une étape de pré-processing des données pour les préparer à l'entraînement du modèle. Cela peut inclure la normalisation, la transformation des variables catégorielles en variables numériques, et la division des données en ensembles d'entraînement et de test.

Dans une approche supervisée nous aurons les étapes suivantes : data collection -> pre-processing -> feature extraction -> model selection -> performance evaluation

1.2. Overfitting et Underfitting

1.2.1. Overfitting

L'overfitting se produit lorsque le modèle apprend trop bien les détails et le bruit des données d'entraînement, au point qu'il ne parvient pas à généraliser sur de nouvelles données. Cela peut entraîner une performance médiocre sur l'ensemble de test. Pour éviter l'overfitting, il est important de :

- Utiliser des techniques de régularisation (comme L1 ou L2) pour pénaliser les poids du modèle.
- Utiliser des ensembles de validation pour évaluer la performance du modèle pendant l'entraînement.

1.2.2. Underfitting

L'underfitting se produit lorsque le modèle est trop simple pour capturer les relations sous-jacentes dans les données. Cela peut également entraîner une mauvaise performance sur l'ensemble de test.

2. Conclusion

Le Machine Learning représente un paradigme fondamentalement différent de la programmation classique dite algorithmique. Désormais, on exploite les données pour définir automatiquement les règles de classification, et non plus l'inverse.