

Programme de formation Machine Learning orienté Python - Les basiques

• Objectifs

Cette formation vous permettra d'acquérir les bases du Machine Learning. À la fin de ces trois jours vous serez en mesure de créer et de déployer vos propres modèles prédictifs, à l'aide de l'écosystème scientifique du langage Python. Vous aurez alors à votre disposition tout un ensemble de bibliothèques et d'outils open-source, tels que Numpy, Scikit-Learn ou TensorFlow. Les concepts théoriques étudiés seront accompagnés d'exemples pratiques, utilisant des données issues de problèmes réels et de domaines variés (biologie, analyse d'images, climatologie, etc).

- Comprendre en quoi consiste l'apprentissage automatique
- Identifier les problèmes pour lesquels ce type de méthode saura apporter une réponse pertinente
- Connaître les différents types de méthodes et savoir utiliser les principaux algorithmes

• Pré requis

Des connaissances de base en statistiques et sur Python

• Durée

3 jours

• Public

Administrateurs, Analystes

• Plan de formation

1 - Les bases de l'apprentissage automatique

Cette première partie définit l'apprentissage automatique.

Elle présente les différents types de méthodes et elle donne les connaissances nécessaires pour identifier quels problèmes se prêtent facilement à l'apprentissage automatique.

Qu'est-ce que l'apprentissage automatique ?

L'apprentissage automatique appartient au domaine de l'intelligence artificielle, nous verrons quelles sont ses spécificités.

Catégorisation des algorithmes

Nous présenterons les différents types de méthodes d'apprentissage automatique. Nous verrons que plusieurs classifications sont possibles, qui correspondent à différents

aspects de l'apprentissage automatique.

- Apprentissage supervisé / Apprentissage non-supervisé
- Classification / Régression
- Problème binaire / Problème multi-classes

Présentation de quelques applications remarquables

Nous reviendrons sur quelques problèmes résolus par des méthodes d'apprentissage automatique.

2 - Manipulation des données

Cette partie traite de l'acquisition et de la représentation des données. La préparation des données est une étape clé dans les méthodes d'apprentissage automatique. Les techniques que nous étudierons nous permettront de découvrir une méthode d'apprentissage automatique supervisé (la

décomposition en composantes principales) et une méthode d'apprentissage automatique non-supervisé (l'algorithme des k-moyennes).

Quelles données utiliser ?

Nous présenterons les types de données fréquemment utilisés dans des problèmes d'apprentissages automatiques. Nous verrons comment normaliser ces données ainsi que les avantages de ce pré-traitement.

Constitution d'un ensemble de données d'apprentissage

Lors de l'utilisation d'une méthode par apprentissage supervisé, la constitution et l'utilisation d'un ensemble de données d'apprentissage pertinent sont fondamentales. Nous étudierons quelques bonnes pratiques :

- vérification de l'adéquation entre les données d'apprentissages et les données réelles
- identification des classes rares et stratégies pour les prendre en compte

Réduction de la dimension des données

Nous présenterons la méthode d'analyse en composantes principales et l'utiliserons pour identifier les variables corrélées et réduire la dimension des données.

Regroupement des données

Nous présenterons l'algorithme des k-moyennes et l'utiliserons pour regrouper les données d'un problème en un nombre restreint d'ensembles.

3 - Régression

Les méthodes de régression permettent de quantifier et de comprendre la relation entre plusieurs variables. Nous verrons comment ces méthodes permettent de réaliser des prédictions ou d'optimiser un environnement de production.

Méthode des moindres carrés ordinaire

L'algorithme des moindres carrés ordinaire est l'une des méthodes de régression les plus simple.

Nous verrons son principe, comment l'utiliser et quels sont ses limites.

Autres algorithmes de régression

Nous présenterons et utiliserons quelques algorithmes plus avancés, proposé par la bibliothèque Scikit-Learn.

4 - Classification par apprentissage supervisé

L'apprentissage supervisé se déroule en deux temps. Tout d'abord la méthode est entraînée à reconnaître chacune des classes à partir d'un ensemble de données d'apprentissage. Cet apprentissage aboutit à la production d'un modèle, capable de déterminer pour de nouvelles données à quelles classes elles appartiennent.

Problèmes binaires

Les problèmes de classifications les plus simples concernent l'attribution d'une donnée à une classe parmi deux. Nous présenterons et utiliserons la méthode Séparateur à Vaste Marge (SVM).

Paramétrage des méthodes par validation croisée

Les méthodes d'apprentissage comprennent généralement de nombreux paramètres. La validation croisée permet de déterminer automatiquement les meilleures valeurs pour ces paramètres.

Classification multi-classes

Nous verrons comment les problèmes de classifications peuvent être étendus de 2 à N classes. Nous montrerons comme les Séparateurs à Vaste Marge sont modifiés pour gérer de tels problèmes. Nous présenterons et utiliserons la méthode Random Forest.

5 - Réseaux de neurones

Les réseaux de neurones et l'apprentissage profond rencontrent un succès croissant dans la résolution de problèmes complexes dans des domaines tels que la finance, la médecine ou la production industrielle.

Principe

Nous décrirons rapidement le fonctionnement d'un réseau de neurones.

La bibliothèque TensorFlow

Nous présenterons la bibliothèque TensorFlow et nous l'utiliserons pour entraîner un réseau de neurones.

6 - Vers des modèles plus complexes

Apprentissage semi-supervisé

L'apprentissage semi-supervisé se situe entre l'apprentissage supervisé et l'apprentissage non-supervisé. Les données d'apprentissages sont bien présentes mais souvent incomplètes.

Apprentissage par renforcement

L'apprentissage par renforcement fait intervenir des interactions régulières avec un validateur.

L'algorithme propose un résultat et le validateur lui indique quelles parties sont fausses et qu'elles sont celles qui sont correctes.

Ces indications permettent d'améliorer les connaissances de la méthode et aboutissent à la proposition d'un nouveau résultat, plus pertinent.

Apprentissage multi-tâches

Lors de l'étude d'une nouvelle discipline, il est fréquent que nous faisons appel aux connaissances et compétences acquises dans d'autres domaines. Par exemple le fait d'étudier l'anglais peut faciliter l'apprentissage d'une deuxième langue étrangère.

L'apprentissage multi-tâche comprend des méthodes capables de résoudre simultanément plusieurs problèmes, de manière à ce que la progression pour un problème influence favorablement la progression sur les autres problèmes