

Apprentissage automatique (Machine learning)

Syllabus

Visseho Adjiwanou, PhD.

12 June 2024

Introduction

Vous avez probablement entendu parler de “l’apprentissage automatique” sans être sûr de savoir exactement ce que c’est, en quoi il diffère des statistiques traditionnelles et ce que vous, en tant que scientifiques sociaux, pouvez en faire. Dans ce chapitre, nous allons démystifier l’apprentissage automatique, établir des liens avec ce que vous connaissez déjà en matière de statistiques et d’analyse de données, et approfondir certains des concepts et méthodes novateurs développés dans ce domaine. Bien que ce domaine trouve son origine dans l’informatique, il a été fortement influencé par les mathématiques et les statistiques au cours des 15 à 20 dernières années. Comme vous le verrez, bon nombre des concepts que vous allez apprendre ne sont pas entièrement nouveaux, mais sont simplement appelés autrement. Par exemple, vous êtes déjà familiers avec la régression logistique (une méthode de classification qui relève du cadre de l’apprentissage supervisé en apprentissage automatique) et l’analyse de regroupement (une forme d’apprentissage non supervisé). Vous apprendrez également des méthodes nouvelles plus exclusivement utilisées en apprentissage automatique, telles que les forêts aléatoires, les machines à vecteurs de support et les réseaux neuronaux. Nous limiterons les formalisations au minimum et nous concentrerons sur la transmission de l’intuition, ainsi que sur des conseils pratiques. Nous espérons que ce chapitre vous rendra à l’aise et familier avec le vocabulaire, les concepts et les processus de l’apprentissage automatique, et vous permettra d’explorer et d’utiliser ces méthodes et outils dans vos propres recherches et pratiques.

Qu’est-ce que l’apprentissage automatique?

Lorsque les humains améliorent leurs compétences avec l’expérience, on dit qu’ils apprennent. Est-il également possible de programmer des ordinateurs pour qu’ils fassent de même ? Arthur Samuel, qui a inventé le terme apprentissage automatique en 1959 (**samuel1959some?**), était un pionnier dans ce domaine, en programmant un ordinateur pour jouer aux dames. L’ordinateur jouait contre lui-même et contre des adversaires humains, améliorant ses performances à chaque partie. Finalement, après un entraînement (et

une expérience) suffisant, l'ordinateur est devenu un meilleur joueur que le programmeur humain. Aujourd'hui, l'apprentissage automatique a considérablement dépassé l'apprentissage du jeu de dames. Les systèmes d'apprentissage automatique ont appris à conduire (et à garer) des voitures autonomes, sont intégrés dans des robots, peuvent recommander des livres, des produits et des films qui nous intéressent (parfois), identifier des médicaments, des protéines et des gènes à étudier pour guérir des maladies, détecter le cancer et d'autres pathologies dans les radiographies et autres types d'imagerie médicale, nous aider à comprendre comment le cerveau humain apprend le langage, aider à identifier quels électeurs sont persuadables lors des élections, détecter quels étudiants sont susceptibles de nécessiter un soutien supplémentaire pour obtenir leur diplôme de lycée à temps, et aider à résoudre de nombreux autres problèmes. Au cours des 20 dernières années, l'apprentissage automatique est devenu un domaine interdisciplinaire englobant l'informatique, l'intelligence artificielle, les bases de données et les statistiques. Au cœur de l'apprentissage automatique, il s'agit de concevoir des systèmes informatiques qui s'améliorent avec le temps et l'expérience. Dans l'un des premiers livres sur l'apprentissage automatique, Tom Mitchell donne une définition plus opérationnelle, déclarant qu' : "Un programme informatique est dit apprendre de l'expérience E par rapport à une certaine classe de tâches T et une mesure de performance P , si sa performance à des tâches de la classe T , mesurée par P , s'améliore avec l'expérience E " (**mitchell1997machine?**). Nous aimons cette définition car elle est centrée sur la tâche et nous permet de considérer l'apprentissage automatique comme un outil utilisé dans un système plus large pour améliorer les résultats qui nous intéressent.

Encadré : Exemples d'apprentissage automatique commercial

Reconnaissance vocale : Les logiciels de reconnaissance vocale utilisent des algorithmes d'apprentissage automatique construits sur de grandes quantités de données d'entraînement initiales. L'apprentissage automatique permet à ces systèmes de s'adapter et de s'ajuster aux variations individuelles de la parole ainsi qu'aux différents domaines.

Voitures autonomes : Le développement continu des voitures autonomes applique des techniques issues de l'apprentissage automatique. Un ordinateur embarqué analyse en continu les flux vidéo et les capteurs entrants pour surveiller les environs. Les données entrantes sont comparées à des images annotées pour reconnaître des objets tels que les piétons, les feux de signalisation et les nids-de-poule. Afin d'évaluer les différents objets, des ensembles de données d'entraînement énormes sont nécessaires, où des objets similaires ont déjà été identifiés. Cela permet à la voiture autonome de décider des actions à entreprendre ensuite.

Détection de fraude : De nombreuses organisations publiques et privées sont confrontées au problème de la fraude et des abus. Les systèmes d'apprentissage automatique sont largement utilisés pour prendre des cas historiques de fraude et signaler les transactions frauduleuses au fur et à mesure qu'elles se produisent. Ces systèmes ont l'avantage d'être adaptatifs et de s'améliorer avec plus de données au fil du temps.

Publicités personnalisées : De nombreuses boutiques en ligne proposent des recommandations personnalisées promouvant des produits susceptibles d'intéresser les utilisateurs. En se basant sur l'historique d'achat individuel et sur ce que d'autres utilisateurs similaires ont

acheté dans le passé, le site prédit les produits qu'un utilisateur pourrait aimer et adapte ses recommandations. Netflix et Amazon sont deux exemples d'entreprises dont les logiciels de recommandation prédisent comment un client évaluerait un certain film ou produit et suggèrent ensuite les articles avec les évaluations prédites les plus élevées. Bien sûr, il y a quelques mises en garde ici, car ils ajustent ensuite les recommandations pour maximiser les profits.

Reconnaissance faciale : Les systèmes de surveillance, les plateformes de réseaux sociaux et les logiciels d'imagerie utilisent tous la détection et la reconnaissance faciale pour d'abord détecter les visages dans les images (ou vidéos) et ensuite les identifier pour diverses tâches. Ces systèmes sont entraînés en fournissant des exemples de visages à un système d'apprentissage automatique qui apprend ensuite à détecter de nouveaux visages et à identifier les individus connus. Le chapitre sur les biais et l'équité mettra en évidence certaines préoccupations concernant ces types de systèmes.

Box: Social Science machine learning examples

Potash et al (**Potash2015?**) worked with the Chicago Department of Public Health and used random forests (a machine learning classification method) to predict which children are at risk of lead poisoning. This early warning system was then used to prioritize lead hazard inspections to detect and remediate lead hazards before they had an adverse effect on the child.

Carton et al (**Carton2016?**) used a collection of machine learning methods to identify police officers at risk of adverse behavior, such as unjustified use of force or unjustified shootings or sustained complaints, to prioritize preventive interventions such as training and counseling.

Athey and Wager (**athey2019?**) use a modification of random forests to estimate heterogeneous treatment effects using a data set from The National Study of Learning Mindsets to evaluate the impact of interventions to improve student achievement.

Voigt et al (**Voigt2017?**) uses machine learning methods to analyze footage from body-worn cameras and understand the respectfulness of police officer language toward white and black community members during routine traffic stops.