L'apprentissage automatique, (en anglais : machine learning, litt. « apprentissage machine, »), apprentissage artificiel ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes. On parle d'apprentissage statistique car l'apprentissage consiste à créer un modèle dont l'erreur statistique moyenne est la plus faible possible.  
L'apprentissage automatique comporte généralement deux phases. La première consiste à estimer un modèle à partir de données, appelées observations, qui sont disponibles et en nombre fini, lors de la phase de conception du système. L'estimation du modèle consiste à résoudre une tâche pratique, telle que traduire un discours, estimer une densité de probabilité, reconnaître la présence d'un chat dans une photographie ou participer à la conduite d'un véhicule autonome. Cette phase dite « d'apprentissage » ou « d'entraînement » est généralement préalable à l'utilisation pratique du modèle. La seconde phase est la mise en production : le modèle étant déterminé, de nouvelles données peuvent alors être soumises afin d'obtenir le résultat correspondant à la tâche souhaitée.   
Certains systèmes peuvent continuer à apprendre une fois en production, s'ils disposent d'un retour sur la qualité des résultats produits. C'est l'apprentissage en ligne, ou l'apprentissage continu.  
Selon le type de données utilisées pour l'apprentissage, on distingue :  
l'apprentissage supervisé : l'algorithme apprend à partir de données étiquetées (la réponse à la tâche, qui est la donnée de sortie, est donc connue pour chaque données d'entrée). L'objectif est de prédire les sorties pour de nouvelles données ;  
l'apprentissage non supervisé : l'algorithme apprend à partir de données non étiquetées. Il cherche à découvrir des structures sous-jacentes, cachées (qui peuvent par exemple être une densité de probabilité) ; des motifs dans les données permettent la classification ou le classement des données ;  
l'apprentissage semi-supervisé : il tire parti d'une grande quantité de données non étiquetées pour améliorer la performance du modèle, tout en utilisant une moindre quantité de données étiquetées pour guider son apprentissage. Il diminue les coûts d'étiquetage manuel des données ;  
l'apprentissage auto-supervisé : c'est une forme d'apprentissage non supervisé, où le modèle génère ses propres étiquettes à partir des données brutes. Le modèle peut ainsi créer des représentations internes utiles, sans nécessiter de données étiquetées manuellement.  
L'apprentissage automatique peut être appliqué à divers types de données, tels des graphes, des arbres, des courbes, ou plus simplement des vecteurs de caractéristiques, qui peuvent être des variables qualitatives ou quantitatives continues ou discrètes.

Si le modèle apprend de manière incrémentale, en fonction d'une récompense reçue par le programme pour chacune des actions entreprises, on parle d'apprentissage par renforcement  
Depuis l'antiquité, le sujet des machines pensantes préoccupe les esprits. Ce concept est la base de pensées pour ce qui deviendra ensuite l'intelligence artificielle, ainsi qu'une de ses sous-branches : l'apprentissage automatique.  
La concrétisation de cette idée est principalement due à Alan Turing (mathématicien et cryptologue britannique) et à son concept de la « machine universelle » en 1936, qui est à la base des ordinateurs d'aujourd'hui. Il continuera à poser les bases de l'apprentissage automatique, avec son article sur « L'ordinateur et l'intelligence » en 1950, dans lequel il développe, entre autres, le test de Turing.  
En 1943, le neurophysiologiste Warren McCulloch et le mathématicien Walter Pitts publient un article décrivant le fonctionnement de neurones en les représentant à l'aide de circuits électriques. Cette représentation sera la base théorique des réseaux neuronaux.  
Arthur Samuel, informaticien américain pionnier dans le secteur de l'intelligence artificielle, est le premier à faire usage de l'expression machine learning (en français, « apprentissage automatique ») en 1959 à la suite de la création de son programme pour IBM en 1952. Le programme jouait au Jeu de Dames et s'améliorait en jouant. À terme, il parvint à battre le 4e meilleur joueur des États-Unis,.  
Une avancée majeure dans le secteur de l'intelligence machine est le succès de l'ordinateur développé par IBM, Deep Blue, qui est le premier à vaincre le champion mondial d'échecs Garry Kasparov en 1997. Le projet Deep Blue en inspirera nombre d'autres dans le cadre de l'intelligence artificielle, particulièrement un autre grand défi : IBM Watson, l'ordinateur dont le but est de gagner au jeu Jeopardy!. Ce but est atteint en 2011, quand Watson gagne à Jeopardy! en répondant aux questions par traitement de langage naturel.  
Durant les années suivantes, les applications de l'apprentissage automatique médiatisées se succèdent bien plus rapidement qu'auparavant.  
En 2012, un réseau neuronal développé par Google parvient à reconnaître des visages humains ainsi que des chats dans des vidéos YouTube,.  
En 2014, 64 ans après la prédiction d'Alan Turing, le dialogueur Eugene Goostman est le premier à réussir le test de Turing en parvenant à convaincre 33 % des juges humains au bout de cinq minutes de conversation qu'il est non pas un ordinateur, mais un garçon ukrainien de 13 ans.  
En 2015, une nouvelle étape importante est atteinte lorsque l'ordinateur « AlphaGo » de Google gagne contre un des meilleurs joueurs au jeu de Go, jeu de plateau considéré comme le plus dur du monde.  
En 2016, un système d'intelligence artificielle à base d'apprentissage automatique nommé LipNet parvient à lire sur les lèvres avec un grand taux de succès,.  
Principes  
Les modèles d'apprentissage automatique contiennent des paramètres dont les valeurs sont ajustées tout au long de l'apprentissage. La méthode de la rétropropagation du gradient est capable de détecter, pour chaque paramètre, dans quelle mesure il a contribué à une bonne réponse ou à une erreur du modèle, et peut l'ajuster en conséquence.