

Qu'est-ce que l'apprentissage automatique ?

Deux définitions de l'apprentissage automatique sont proposées. **Arthur Samuel** le décrit comme : Il s'agit d'une définition plus ancienne et informelle.

"le domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés".

Tom Mitchell fournit une définition plus moderne :

"On dit qu'un programme informatique apprend de l'expérience E en ce qui concerne une certaine classe de tâches T et une mesure de performance P , si sa performance aux tâches dans T , telle que mesurée par P , s'améliore avec l'expérience E ."

Exemple : jouer aux dames.

E = l'expérience de nombreuses parties de jeu de dames

T = la tâche de jouer aux dames.

P = la probabilité que le programme gagne le prochain match.

En général, tout problème d'apprentissage automatique peut être classé dans l'une des deux grandes catégories suivantes :

Apprentissage supervisé et apprentissage non supervisé.

Apprentissage supervisé

Dans l'apprentissage supervisé, on nous donne un ensemble de données et nous savons déjà à quoi devrait ressembler notre sortie correcte, en ayant l'idée qu'il existe une relation entre *l'entrée et la sortie*.

Les problèmes d'apprentissage supervisé sont classés en problèmes de *"régression"* et de *"classification"*.

- ✓ Dans un problème de régression, nous essayons de prédire les résultats d'une sortie continue, ce qui signifie que nous essayons de faire correspondre les variables d'entrée à une fonction continue.
- ✓ Dans un problème de classification, nous essayons plutôt de prédire des résultats dans une sortie discrète. En d'autres termes, nous essayons de faire correspondre les variables d'entrée à des catégories discrètes.

Exemple 1 :

Étant donné des données sur la taille des maisons sur le marché immobilier, essayez de prédire leur prix. Le prix en fonction de la taille est une donnée continue, il s'agit donc d'un problème de régression.

Nous pourrions transformer cet exemple en un problème de classification en indiquant si la maison "se vend à un prix supérieur ou inférieur au prix demandé". Ici, nous classons les maisons en fonction du prix en deux catégories distinctes.

Exemple 2 :

(a) Régression - Étant donné une photo d'une personne, nous devons prédire son âge sur la base de cette photo.

(b) Classification - Dans le cas d'un patient présentant une tumeur, nous devons prédire si la tumeur est maligne ou bénigne.

Apprentissage non supervisé

L'apprentissage non supervisé nous permet d'aborder des problèmes avec peu ou pas d'idée de ce à quoi nos résultats devraient ressembler. Nous pouvons dériver une structure à partir de données dont nous ne connaissons pas nécessairement l'effet des variables.

Nous pouvons dériver cette structure en regroupant les données sur la base des relations entre les variables des données.

Avec l'apprentissage non supervisé, il n'y a pas de retour d'information basé sur les résultats de la prédiction.

Exemple :

Regroupement : Prenez une collection de 1 000 000 de gènes différents et trouvez un moyen de regrouper automatiquement ces gènes en groupes qui sont en quelque sorte similaires ou liés par différentes variables, telles que la durée de vie, l'emplacement, les rôles, etc.

Non-clustering : L'algorithme "Cocktail Party" vous permet de trouver une structure dans un environnement chaotique. (par exemple, identifier des voix individuelles et de la musique à partir d'un ensemble de sons lors d'une [soirée cocktail](#)).