

## ML Machine Learning

### Uma breve introdução

O termo Machine Learning vem de um ramo da inteligência artificial que faz uso de métodos e algoritmos para estabelecer, automaticamente, modelos a partir de uma coleção de dados que podem ou não estarem classificados. No primeiro contato, Machine Learning parece remeter aos robôs de filmes de ficção científica, mas no mundo real, o aprendizado de máquina significa muito mais delegar a tarefa de estabelecer as funções matemáticas que regem um determinado experimento para um ou mais algoritmos.

Imagine que estamos desenvolvendo um algoritmo para prever se você deveria levar uma blusa de frio ao sair de casa, a princípio, podemos apenas verificar se a temperatura está acima ou abaixo de 20°C, se sim, levamos a blusa, se não, simplesmente não levamos, neste caso, nosso algoritmo trabalha com uma série de regras preestabelecidas, e sempre trará o mesmo resultado todas as vezes para um determinado valor de entrada (temperatura).

Para o exemplo acima é bastante plausível utilizar um algoritmo com regras predefinidas (cheio de ifs) para decidir se você deveria levar ou não uma blusa ao sair de casa, mas se introduzirmos um pouco mais de complexidade, como por exemplo, se você precisar decidir levar ou não uma blusa de frio ao sair de casa para o trabalho, mas sabe que vai ficar fora de casa por mais de oito horas, fica um pouco mais difícil saber se deveria levar a blusa ou não levando em consideração apenas a temperatura naquele momento do dia.

Para este tipo de problema, seria mais interessante avaliar uma série de outras variáveis que possam influenciar a mudança de temperatura ao longo do dia. Observar todas essas variáveis pode ser um trabalho muito custoso para um pesquisador ou cientista, uma vez que ele terá que fazer uso de terabytes de dados históricos para tomar uma decisão, para um computador, no entanto, essa tarefa é muito mais factível, já que computadores foram feitos para realizar uma só tarefa: cálculos.

Com base neste exemplo podemos resumir utilizando uma lógica bastante simplista, que o aprendizado de máquina busca descobrir as equações que regem o comportamento de um determinado problema.

### Tipos de Algoritmos

O objetivo principal de um algoritmo de aprendizado de máquina é converter uma coleção de dados, mais comumente chamada de data set, em um modelo que pode ser aplicado em uma necessidade que é atendida pelo modelo. O data set em si pode conter dados que estão muito bem classificados, de forma que é possível fazer associações básicas aos dados de entrada e saída, no entanto, existem data sets não classificados, para estes, pode ser mais difícil identificar padrões referentes aos dados de entrada e saída.

Com base nas características do data set, podemos aplicar técnicas diferentes de ML para extrair informações importantes, como por exemplo, identificar padrões não observados anteriormente, ou reclassificar um grupo de dados, para facilitar, podemos classificar essas técnicas de aprendizado em dois tipos:

- **Aprendizado supervisionado:** O aprendizado é feito a partir da observação de um data set bem definido com padrões de entrada e saída classificados, para um valor  $x$  de entrada, um valor  $y$  de saída é garantido. Um exemplo da vida real é uma sala de aula, na qual os alunos recebem a informação categorizada diretamente da fonte, que é o professor.



- **Aprendizado não supervisionado:** Nestes casos, o aprendizado acontece a partir de padrões observados pelo próprio algoritmo, na tentativa de encontrar uma função que receba os inputs de  $x$  e processe gerando  $y$ , nestes casos, o data set não está classificado, ou seja, para uma determinada entrada, o valor da saída não está bem definido.



Dentro de cada um dos tipos de aprendizado descritos acima, podemos ainda aplicar técnicas diferentes a depender do tipo de modelo que estamos tentando estudar, alguns são mais utilizados em data sets classificados, outros são mais úteis para estudo utilizando aprendizado não supervisionado:

- **Classificação:** Algoritmos de aprendizado de máquina supervisionados que fazem escolhas probabilísticas entre duas ou mais classes distintas, como por exemplo, um algoritmo que identifica em imagens a diferença entre blusas e camisetas, ou que define se o animal em uma imagem é um gato ou cachorro por exemplo.
- **Regressão:** Um algoritmo de regressão é basicamente um algoritmo supervisionado que deduz um valor numérico a partir do modelo estudado. Na verdade os algoritmos de regressão e

classificação se sobrepõem bastante e é comum utilizar as duas formas para atacar um problema.

- Clustering (Agrupamento): O agrupamento de dados é uma forma de aprendizado de máquina não supervisionada, que busca identificar agrupamentos naturais no data set.

Agora que sabemos a teoria, vamos começar a entender como podemos aplicar um desses algoritmos na prática.