**COMO ESCOLHER O ALGORITMO IDEAL?**

<https://www.datageeks.com.br/algoritmos-de-machine-learning/>

Com a evolução da inteligência artificial, pesquisadores da ciência da computação estão a todo momento sugerindo novas abordagens para tornar os computadores ainda mais autônomos e capazes de executar tarefas com precisão.

Por essa razão, tem aumentado a utilização de [algoritmos](https://pt.wikipedia.org/wiki/Algoritmo) de [Machine Learning](https://www.datageeks.com.br/machine-learning/) para resolver problemas complexos. Entre seus benefícios está a capacidade das máquinas aprenderem sem serem explicitamente programadas para determinada atividade.

Neste artigo apresentarei as 3 principais abordagens para aprendizagem de máquinas: supervisionada, não supervisionada e por reforço.

**O que é Machine Learning?**

Tradicionalmente, os programadores sempre desenvolveram sistemas especificando cada passo em código, de forma clara. Para isso, engenheiros de software precisavam conhecer bem os problemas que buscavam resolver e ter uma boa noção de todas as variáveis e possibilidades envolvidas.

Machine Learning (ML) é um subcampo da inteligência artificial que visa resolver problemas mais complexos, dos quais os programadores ainda não têm um conhecimento tão amplo.

Também se aplica quando é inviável mapear todas as variáveis e possibilidades. Para essa finalidade, os sistemas são estimulados a solucionar problemas de forma autônoma, aprendendo sobre eles e propondo uma saída.

Nesse contexto, os algoritmos analisam dados de entrada, processam e preveem saídas possíveis dentro de um intervalo definido. No processo, eles tentam diferentes abordagens e otimizam sua capacidade de chegar ao resultado.

**Os sistemas artificiais no ML passam por duas fases: treino e execução**. Na primeira, são alimentados por um conjunto de entrada para que sejam contextualizados com a situação específica que envolve o problema.

Então, na segunda fase, eles começam a deduzir os caminhos a partir do que aprenderam na etapa anterior. Essas duas fases variam a depender do tipo de algoritmo, como veremos mais adiante neste artigo.

Embora não seja novo, o Machine Learning é um campo do conhecimento em desenvolvimento, ou seja, ganha novas abordagens e novos modelos frequentemente. Essas evoluções buscam resolver problemas com ainda mais precisão e oferecer opções especiais para que cada ferramenta se adapte ao problema específico.

**Métodos de Machine Learning?**

A seguir, discutiremos os principais métodos do ML e suas características.

**Aprendizagem supervisionada**

A aprendizagem supervisionada é útil nos casos em que uma propriedade (rótulo) está disponível para um determinado conjunto de dados (conjunto de treinamento). Todas as entradas e saídas são conhecidas, mas precisam ser previstas para outras instâncias.

Então, esses dados são passados para o sistema de aprendizagem, que tem como função descobrir caminhos e ajustar seu próprio modelo para chegar aos resultados esperados.

Um exemplo prático seria ensinar uma máquina a reconhecer e categorizar e-mails, separando os que são relevantes dos que são SPAM.

Na fase de treinamento, são transmitidas para o algoritmo amostras de ambos os casos, visto que os dois blocos de informação são plenamente conhecidos. Já na fase de execução, o sistema terá a capacidade de determinar se um e-mail novo é ou não uma mensagem indesejada.

**Ou seja, o software já sabe quais entradas estão associadas com quais saídas, mas precisa aprender um meio de entender essa associação.**

Nesse processo, ele tenta identificar padrões e estabelecer previsões que ajudam a otimizar a abordagem. Esse modelo tem esse nome porque é como se o operador humano estivesse sempre dando assistência ao sistema, ensinando-o de uma forma direta.

Os resultados desse tipo de algoritmo são geralmente marcados como classificação e regressão. O primeiro diz respeito a uma forma de mapear elementos iguais em categorias específicas, como no exemplo acima, “spam” e “não spam”. Já o segundo consiste em identificar uma tendência para os dados que permite, inclusive, predizer o futuro com base em dados históricos.

Uma aplicação comum é no setor de segurança, com o uso de ferramentas para identificar comportamentos suspeitos em uma rede e garantir proteção contra ataques.

**Aprendizagem não-supervisionada**

Já na aprendizagem não-supervisionada o desafio é descobrir relações implícitas em um conjunto de dados não rotulados. Dessa forma, o algoritmo fica encarregado de identificar padrões para rotular os dados.

Um exemplo muito comum é a empresa que fornece como base para um sistema de ML um conjunto de dados sobre os clientes e espera que ele identifique possíveis atributos comuns e padrões de comportamento para criar ofertas específicas e segmentadas. No começo, a companhia não conhece bem as informações a ponto de saber as possíveis saídas.

É como um aprendizado sem assistência humana, com o sistema ditando os caminhos a seguir. Nesse modelo de algoritmo, as estratégias comuns são *clustering* e redução de dimensão.

Abordagens como essas podem ser utilizadas em sistemas de recomendação, em que, com base em informações colaborativas ou específicas, um software consegue filtrar conteúdos ideais em meio a uma miríade deles.

É muito usado em lojas de e-commerce, para recomendar produtos que possam ser interessantes para o cliente que visita o site. Similarmente, há serviços de streaming que sugerem músicas ou vídeos que devem ser vistos por um usuário. Tanto **Netflix**como **Amazon** utilizam essa abordagem para criar seus algoritmos de recomendação.

**Aprendizagem por reforço**

Nesse método de aprendizagem, que segue o estilo utilizado no início dos estudos sobre Inteligência Artificial, o computador é estimulado a aprender com base em tentativas e erros, otimizando o processo na prática direta. Com essa abordagem, é possível, por exemplo, ensinar um sistema a priorizar hábitos em detrimento de outros, com recompensas proporcionais ao acerto.

A aprendizagem por reforço foi inspirada por psicólogos comportamentalistas, que acreditavam na eficácia de recompensas e punições na educação dos seres humanos. Também lembra o procedimento de adestração de animais domésticos.

É um método, portanto, baseado na construção de experiência. A partir disso, o algoritmo sabe quais caminhos são melhores que outros e quais processos são mais ágeis.

Exemplos de aplicação são os veículos autônomos e máquinas que jogam xadrez. O sistema aprende com múltiplas tentativas, que envolvem erros como uma jogada ruim ou um choque contra um obstáculo.

**Quais os principais algoritmos de machine learning?**

Existe uma infinidade de algoritmos utilizados em machine learning, cada um com uma finalidade específica.

Há também características que podem inviabilizar a escolha do modelo mais preciso para determinado problema, como a utilização alto poder computacional.

**Abaixo, uma lista com os algoritmos mais populares:**

* [Regressão Linear](https://www.datageeks.com.br/regressao-linear/)
* Regressão logística
* Árvores de Decisão
* [Classificação Naive Bayes](https://www.datageeks.com.br/naive-bayes/)
* [Extreme Gradient Boosting – XGBoost](https://www.datageeks.com.br/xgboost/)
* SVM (Support Vector Machine)
* Ensemble Methods
* Algoritmos de Agrupamento (Clustering)
* Decomposição em valores singulares
* Análise de Componentes Principais
* Análise de componentes independentes
* KNN (K-vizinhos mais próximos)
* Floresta aleatória ou Random Forest
* Algoritmos de redução dimensional

**Como escolher o algoritmo ideal?**

Como são várias as formas de se chegar a um resultado com ML, e por existir diversas abordagens, é preciso saber bem como escolher o algoritmo mais adequado ao problema proposto. Veremos algumas dicas a seguir.

**Finalidade do problema**

É importante conhecer bem a finalidade e o contexto do caso que deve ser processado pela máquina, com detalhes específicos, pois eles ajudam a selecionar o melhor jeito de tratar os dados.

Caso exista um objetivo bem definido, com saídas conhecidas, o ideal pode ser um método de aprendizagem supervisionada. Dentro dessa abordagem, caso o resultado seja um valor numérico, é possível usar regressão. Caso sejam categorias, classificação.

**Quantidade de dados**

Outro aspecto relevante é a quantidade de dados que será utilizada para alimentar o software. Redes neurais, por exemplo, geralmente precisam de grandes conjuntos de entrada.

É importante conhecer bem e se ater ao tamanho das bases para estabelecer qual método seguir, pois cada um lida com a quantidade de uma forma diferente. Além disso, caso a base de dados seja muito grande e o algoritmo demasiado complexo, pode ser preciso fazer uma avaliação a respeito da capacidade computacional necessária para a execução do algoritmo.

**Complexidade do problema e acurácia**

A complexidade do problema também é uma questão importante, pois isso determina o grau de acurácia desejado. Com isso, é possível selecionar qual algoritmo se encaixa melhor nesse grau de precisão analisando fatores como a periculosidade de um erro e as consequências que eles podem trazer.

**Tempo**

E, por fim, o tempo deve ser levado em conta. Existem problemas que vão requerer decisões em tempo real. Outros, no entanto, com um intervalo aceitável. Isso deve ser pensado na hora de escolher o tipo de ML que resolva o caso da melhor forma.

Neste artigo, vimos como as soluções de Machine Learning diferem e são classificadas em três grandes categorias principais e em subcategorias menores. Cada uma tem sua própria característica e natureza.

Portanto, se adapta a problemas específicos de contextos especiais. Para selecionar o algoritmo que melhor resolve o seu caso é preciso analisar alguns fatores, como: complexidade, tempo e quantidade de dados.