

# 1 Introdução

Texto da introdução.

## 1.1 O que é machine learning?

Aprendizagem de máquina é a área da ciência da computação que tem por objetivo o desenvolvimento de aplicações capazes de aprender estratégias para execução de uma determinada tarefa a partir da exposição a informações relevantes, diferenciando-se de sistemas tradicionais onde o comportamento deve ser pré-programado [1]. Aplicações baseadas em aprendizagem de máquina são capazes de identificar padrões de interação complexos entre variáveis de um conjunto de dados com alta dimensionalidade para realizar tarefas de classificação, regressão, agrupamento e outras [2].

Considere, por exemplo, o cenário onde tem-se um conjunto de dados provenientes de um estudo clínico; os dados incluem, para cada paciente, informações demográficas (idade, gênero, renda, educação), informações clínicas coletadas antes do tratamento (respostas observadas em entrevistas ou testes psicológicos) e informações clínicas coletadas após o tratamento, incluindo a avaliação final sobre a eficácia do tratamento. Uma aplicação de aprendizagem de máquina seria capaz de ingerir tal conjunto de dados e aprender a realizar determinadas tarefas. Uma tarefa possível é prever o desfecho da intervenção testada mediante a apresentação das informações demográficas e clínicas de um novo paciente. Outra possibilidade é utilizar padrões de similaridade no conjunto de dados para criar agrupamentos, facilitando a identificação de subgrupos dentro da população alvo e permitindo analisar os efeitos da intervenção de maneira mais assertiva para os diferentes subgrupos.

Por trás dessas aplicações, encontram-se modelos computacionais com múltiplos parâmetros e o processo de aprendizagem consiste em encontrar um conjunto de parâmetros que produz os melhores resultados para a tarefa em questão; nesse sentido, a aprendizagem de máquina aproxima-se de métodos estatísticos tradicionais. A diferenciação entre métodos estatísticos e de aprendizagem de máquina é feita, muitas vezes, com base em critérios históricos, mas é possível identificar características próprias das abordagens de aprendizagem de máquina: ao contrário da abordagem estatística inferencial, por exemplo, a aprendizagem de máquina não assume nenhuma relação entre variáveis; em certa medida, pode-se dizer que é uma abordagem agnóstica de teoria [1, 2].

## 1.2 Quais os tipos de técnicas de machine learning?

As diferentes técnicas de aprendizagem de máquina podem ser classificadas de acordo com a estratégia adotada durante o processo de aprendizagem. As principais categorias são: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço [1].

### **1.2.1 Supervisionada**

Na aprendizagem supervisionada, a aplicação é exposta a um conjunto de dados que contém informações sobre o desfecho para cada uma das observações. A informação sobre o desfecho pode ser um conjunto de variáveis ou uma única variável; em ambos os casos as variáveis podem representar quantidades ou categorias. O fato de a aplicação ter acesso às informações de desfecho, geralmente providas por um agente externo (um pesquisador que registrou os desfechos manualmente), confere o caráter de supervisão à este processo. Técnicas de aprendizagem de máquina para regressão e classificação (support vector machines, árvores de decisão, redes neurais) pertencem a esta categoria.

### **1.2.2 Não supervisionada**

Na aprendizagem não supervisionada, o conjunto de dados analisado não contém informações sobre o desfecho para as observações; perde-se assim a característica de supervisão. Espera-se que a aplicação identifique, de maneira autônoma, os padrões de relacionamento existentes entre as variáveis do conjunto de dados. Técnicas de aprendizagem de máquina populares para tarefas de agrupamento e redução de dimensionalidade (k-means clustering, PCA, TSNE) pertencem a esta categoria.

### **1.2.3 Por reforço**

Na aprendizagem por reforço, as aplicações adquirem conhecimento a respeito da tarefa ao longo do tempo por meio da obtenção de feedback sobre seu desempenho.

## **1.3 A construção de uma aplicação de machine learning**

### **1.3.1 Análise descritiva**

### **1.3.2 Pré-processamento**

Na etapa de pré-processamento, busca-se preparar o conjunto de dados de treinamento, colocando-no em um estado adequado à técnica de aprendizagem de máquina que se pretende utilizar. Tratamentos comumente realizados na etapa de pré-processamento são: seleção de características, transformações, imputações e balanceamento de classes.

A seleção de características consiste em eliminar do conjunto de dados as variáveis que tenham pouca contribuição para a aprendizagem da tarefa. Em um conjunto de dados onde todas as observações são de pessoas brasileiras, a variável de nacionalidade não contribui para a explicação do desfecho que se busca prever, portanto pode ser removida.

Transformações são aplicadas de acordo com os requisitos da técnica de aprendizagem de máquina em uso. Por exemplo, algumas técnicas de aprendizagem de máquina são suscetíveis à influência de variáveis com escala muito

superior às demais; nesses casos é comum transformação das variáveis para uma escala padronizada (medida em desvios padrão a partir da média).

### **1.3.3 Treinamento do modelo**

### **1.3.4 Validação do modelo**

## **2    Desenvolvimento**

Texto do desenvolvimento.

### **2.1   Parte 1**

Texto da parte 1.

### **2.2   Parte 2**

Texto da parte 2.

### **3 Conclusão**

Texto da conclusão.

## References

- [1] Qifang Bi et al. “What is Machine Learning? A Primer for the Epidemiologist”. In: *American Journal of Epidemiology* (Oct. 2019). ISSN: 1476-6256. DOI: 10.1093/aje/kwz189. URL: <http://dx.doi.org/10.1093/aje/kwz189>.
- [2] Jaime Delgadillo. “Machine learning: A primer for psychotherapy researchers”. In: *Psychotherapy Research* 31.1 (Dec. 2020), pp. 1–4. ISSN: 1468-4381. DOI: 10.1080/10503307.2020.1859638. URL: <http://dx.doi.org/10.1080/10503307.2020.1859638>.