

# Resumo do artigo "A Few Useful Things to Know About Machine Learning"

André Rato[45517], Diogo Faustino[40968], José Alexandre[45223]

9 de março 2021

## Resumo

Neste trabalho faz-se um resumo do artigo intitulado "A Few Useful Things to Know About Machine Learning"[1].

## 1 Descrição do artigo analisado

Neste artigo, é feita uma apresentação de alguns conhecimentos importantes adquiridos por investigadores através das suas experiências. Este artigo está dividido em 12 secções.

Na primeira secção, são apresentados os 3 componentes que todos os algoritmos de *machine learning* possuem:

- a **representação** numa *machine learner* é o conjunto de classificadores/funções que podem ser aprendidos (espaço de hipóteses) - se uma função não está no espaço de hipóteses, não pode ser aprendida;
- a função de **avaliação** avalia o modelo da *machine learning*;
- a **otimização** é o método utilizado para procurar o melhor modelo de aprendizagem.

Nas segunda e terceira secções, são focados os temas da generalização dos dados utilizados e da quantidade dos mesmos ser suficiente ou não para a representação e avaliação dos modelos.

Overfitting é o tema principal da quarta secção, onde é referido que uma das maneiras de interpretar overfitting é partir o erro de generalização em duas componentes: **bias** (tendência de a *machine learner* aprender a mesma coisa errada) e **variance** (tendência para aprender coisas aleatórias).

Na quinta secção, o autor refere que generalizar corretamente torna-se mais difícil consoante a dimensão dos dados. Os algoritmos de *machine learning* dependem de um raciocínio baseado em similaridade.

O autor aborda ainda, na secção 6, que todas as garantias teóricas que possam ser tomadas não são o que aparentam e podem gerar erros no final. Este problema pode ser corrigido recorrendo à informação presente na sétima secção,

onde o autor explica que a engenharia de recursos é a chave para o sucesso na área.

Como regra geral, um algoritmo pouco inteligente com muitos dados supera um algoritmo muito inteligente com uma quantidade modesta de dados. Porém, quanto mais dados, maior a quantidade de problemas (secção 8).

Na nona secção, o autor defende que aprender muitos modelos é melhor do que aprender apenas um.

Simplicidade não implica precisão é abordado na secção 10 e a diferença entre representável e aprendível na secção 11.

Na última secção do artigo é referido que a relação causa-relação não se aplica à *machine learning*.

## Referências

- [1] Domingos, P. - A Few Useful Things to Know About Machine Learning. Tapping into the “folk knowledge” needed to advance machine learning applications. United States: Andrew A. Chien. 0001-0782. Volume 55 2012, 78–87. Last accessed 9 March 2021