

Aprendizado de Máquina

1. Introdução e definições

A área de pesquisa conhecida como aprendizado de máquina (ML, do inglês *machine learning*) inicialmente despontou como uma sub-especialidade da inteligência artificial.

Vejamos algumas **definições** para ML:

- “*Machine learning is the science (and art) of programming computers so they can **learn from data***”. (Aurélien Géron, 2017)
- “*Field of study that gives computers the ability to **learn without being explicitly programmed***”. (Arthur Samuel, 1959)

Em suma, como o próprio nome sugere, o ponto central em *machine learning* está associado à busca por modelos computacionais que se mostrem capazes de aprimorar sua habilidade em realizar uma determinada tarefa à medida que são expostos aos dados disponíveis.

1.1. Por que ML se tornou tão importante?

Nos últimos anos, o volume de dados disponíveis – das mais variadas naturezas, como texto, imagem e áudio – atingiu escalas sem precedentes (de *tera* a *petabytes*), praticamente impossibilitando que seres humanos lidem diretamente com as informações coletadas.

Da perspectiva de ML, porém, a disponibilidade de uma quantidade expressiva de dados – desde que sejam representativos com relação à tarefa que se pretende realizar – é um fator que pode contribuir para que o aprendizado seja bem-sucedido.

Outros fatores que certamente contribuíram para esta expansão de ML e, em especial, de *deep learning*, são:

- Maior disponibilidade de recursos computacionais para armazenamento e processamento – GPUs (*graphics processing units*), TPUs (*tensor processing units*), *clusters*, *cloud computing*.
- Boas estratégias e técnicas para realizar o treinamento, *i.e.*, o ajuste dos parâmetros do modelo escolhido.
- Existência de *frameworks* para o desenvolvimento eficiente de soluções de ML.



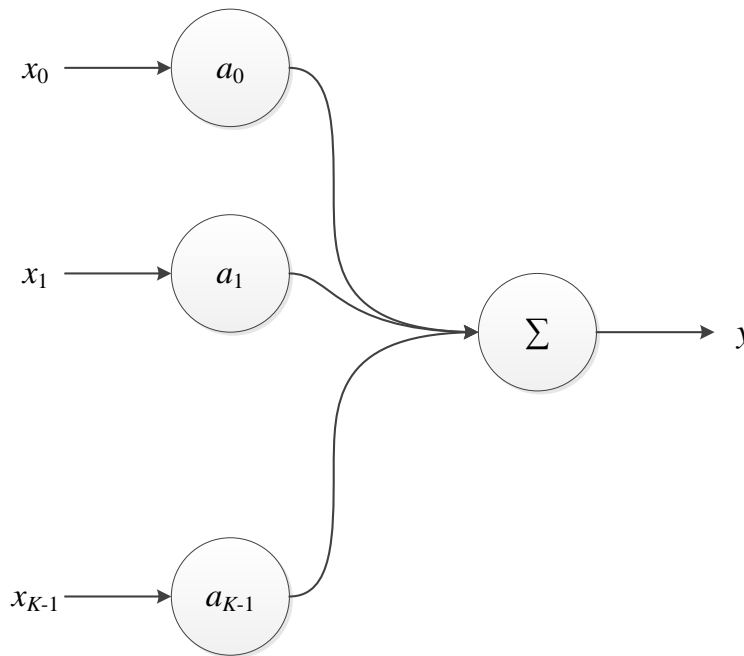
1.2. Em que consiste o aprendizado?

Para se compreender melhor esta questão, é pertinente pensar em três escolhas que são absolutamente vitais em *machine learning*: (1) estrutura; (2) critério; e (3) algoritmo de otimização.

A primeira escolha está relacionada à estrutura do modelo que será utilizado para realizar a tarefa desejada. Em diversas aplicações, como regressão, predição e controle, o modelo a ser construído é responsável por gerar uma resposta (saída). Neste caso, a estrutura está relacionada a como pode ser caracterizado o mapeamento entrada-saída gerado pelo modelo.

Neste contexto, é preciso considerar aspectos relacionados à linearidade, causalidade, presença ou não de realimentação (memória), etc.

- **Exemplo:**



Combinador linear

$$y = \sum_{i=0}^{K-1} a_i x_i$$

- **Tópicos relacionados:**

- ✓ Regressão linear.
- ✓ Regressão logística.
- ✓ Redes neurais artificiais.
- ✓ Máquinas de vetores-suporte.
- ✓ Árvores de decisão.
- ✓ *Deep learning*.

A estrutura escolhida para o sistema, a qual é caracterizada pela existência de diversos parâmetros (e.g., os coeficientes a_i da combinação linear), precisa ser ajustada ou adaptada tendo em vista um objetivo específico.

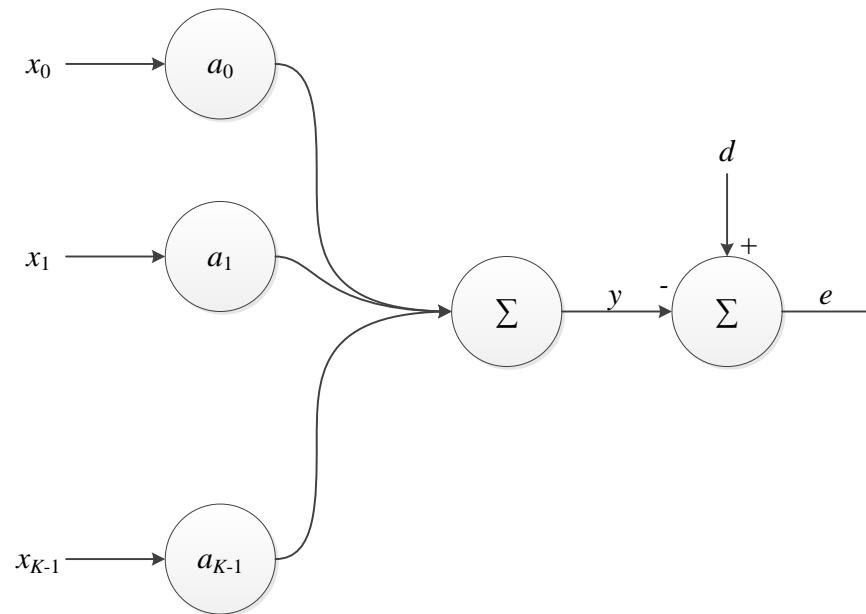
A segunda escolha tem a ver com o critério de adaptação, o qual exprime matematicamente o que se deseja atingir durante o processo de aprendizado, revelando o grau de sucesso alcançado pelo modelo na realização da tarefa.

Tal escolha gravita em torno de um aspecto importante: a quantidade de informação a priori que se encontra disponível com respeito ao comportamento desejado.

- **Tópicos relacionados:**

- ✓ Aprendizado supervisionado.
- ✓ Aprendizado não-supervisionado.
- ✓ Função de erro (custo).
- ✓ Regularização.
- ✓ Validação cruzada.
- ✓ Capacidade de generalização.
- ✓ Aprendizado baseado na teoria da informação.

- Exemplo:



- Deseja-se encontrar os valores dos coeficientes a_i que façam com que a saída y seja a mais parecida possível com a referência d .
- Dada a disponibilidade de uma resposta desejada para o padrão de entrada fornecido ao sistema, é possível observar o sinal de erro (e) e, portanto, estabelecer como critério a minimização de uma função diretamente ligada a

ele, como, por exemplo, o erro quadrático médio (MSE, do inglês *mean-squared error*):

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^N |e_k|^2.$$

Via de regra, o objetivo deve ser atingido no maior nível possível. Em outras palavras, busca-se o conjunto de parâmetros do sistema que **otimize** uma determinada função matemática que representa o objetivo a ser alcançado. Logo, a definição do critério dá origem a um **problema de otimização**. No exemplo do combinador linear, teríamos:

$$\min_{a_i} \text{MSE} = \min_{a_i} \frac{1}{N} \sum_{k=1}^N |e_k|^2$$

Por fim, uma vez estabelecido o objetivo, é preciso escolher o algoritmo responsável por encontrar os parâmetros ótimo do modelo.

Mais uma vez, no caso do combinador linear, existe uma solução em forma fechada baseada na pseudoinversa de Moore-Penrose que minimiza o MSE.

- **Tópicos relacionados:**

- ✓ Otimização não-linear.
- ✓ Retropropagação do erro (*error backpropagation*).
- ✓ Solução fechada ou método iterativo.
- ✓ Esquema em batelada ou padrão-a-padrão.
- ✓ *Expectation-maximization* (EM).

1.3. Alicerces de ML

Os conceitos, modelos e algoritmos existentes em ML são desenvolvidos e compreendidos tendo como base diferentes temas da matemática e engenharia, tais como:

- Otimização não-linear.

- Probabilidade e estatística.
- Álgebra linear.
- Teoria da informação.
- Teoria de grafos.
- Computação.

2. Paradigmas de aprendizado

2.1. Supervisionado

Para cada amostra (ou padrão) de treinamento disponível, existe uma resposta desejada que é conhecida. Neste caso, dizemos que os dados são **rotulados**.

Sendo assim, é possível definir uma função custo que expresse uma medida de erro entre a resposta que o modelo gerou e a resposta esperada para cada padrão do conjunto de treinamento.

2.2. Não-supervisionado

Não há uma saída desejada associada a cada padrão, de modo que os dados são **não-rotulados**. Neste cenário, desejamos que o modelo seja capaz de capturar, representar ou evidenciar propriedades ou regularidades existentes no conjunto de dados.

2.3. Semi-supervisionado

Trata-se de um meio-termo entre os dois tipos anteriores de aprendizado. Assim, para a construção do modelo, são utilizados tanto dados rotulados quanto não-rotulados.

2.4. Reforço

Embora não seja possível indicar a saída correta para cada amostra, como ocorre no caso supervisionado, temos acesso a uma informação sobre a qualidade da saída gerada pelo modelo na forma de um sinal de recompensa ou punição.

3. A “febre”

The New York Times

Data Science: The Numbers of Our Lives

By Claire Cain Miller

April 11, 2013



HARVARD BUSINESS REVIEW calls data science “the sexiest job in the 21st century,” and by most accounts this hot new field promises to revolutionize industries from business to government, health care to academia.

The field has been spawned by the enormous amounts of data that modern technologies create — be it the online behavior of Facebook users, tissue samples of cancer patients, purchasing habits of grocery shoppers or crime statistics of cities. Data scientists are the magicians of the Big Data era. They crunch the data, use mathematical models to analyze it and create narratives or visualizations to explain it, then suggest how to use the information to make decisions.

In the last few years, [dozens of programs](#) under a variety of names have sprung up in response to the excitement about [Big Data](#), not to mention the six-figure salaries for some recent graduates.

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules



Data: 06 de Maio de 2017



Computer Wins on 'Jeopardy!': Trivial, It's Not

By JOHN MARKOFF FEB. 16, 2011



330

[See how this article appeared when it was originally published on NYTimes.com](#)



DeepMind

Google's Go-playing AI still undefeated with victory over world number one

AlphaGo has won its second game against China's Ke Jie, sealing the three-game match in its favour



Alex Hern

@alexhern

Thursday 25 May 2017
09.50 BST



Chinese Go player Ke Jie reacts during his second match against DeepMind's game-playing AI, AlphaGo.
Photograph: China Stringer Network/Reuters

Google's Go-playing AI has won its second game against the world's best player of



Rumours have swirled about Uber's autonomous car ambitions for a while, and the company has finally confirmed them

Project Kinect for Azure combines AI and Microsoft's next-gen depth camera

Kinect's legacy lives on.

DAN THORP-LANCASTER 7 May 2018



Pentagon & Congress

Pentagon outlines its first artificial intelligence strategy

By: Matt O'Brien, The Associated Press

February 12



4. Alguns marcos históricos

Anos 1940: John von Neumann – autômatos celulares / auto-replicação.

1943: Warren McCulloch e Walter Pitts – modelo lógico (“booleano”) do neurônio.

1946 – 1953: *Macy conferences – Cybernetics.*

1948: Norbert Wiener – livro *Cybernetics*.

1948/1949: William Grey Walter – “tortoises”.

1950: Alan Turing – artigo intitulado *Computing Machinery and Intelligence* – Teste de Turing.

1950: Claude Shannon – *Programming a Computer for Playing Chess*.

1955: Allen Newell / Herbert Simon – *Logic Theorist*: Prova de teoremas do “*Principia Mathematica*” de Russell e Whitehead. Simon: “solução do problema mente/corpo”.

1956: Dartmouth Conference: organizada por Marvin Minsky, John McCarthy, Claude Shannon, Nathan Rochester. Na proposta, foi utilizada a expressão *artificial*

intelligence (AI). Pretende-se discutir “computers, natural language processing, neural networks, theory of computation, abstraction and creativity”. Marca o início da **era de ouro** da AI.

1958: Frank Rosenblatt – *perceptron*.

- *“Perceptron may eventually be able to learn, make decisions, and translate languages”.*

1959: Herbert Simon, J. C. Shaw, Allen Newell – *General Problem Solver*: busca em grafos como base.

1964 – 1966: ELIZA – Joseph Weizenbaum – primeiro *chatterbot*.

Final dos anos 1960: Marvin Minsky e Seymour Papert: *micro-worlds, blocks world*.

Exemplo – Terry Winograd e SHRDLU (1968 – 1970): instruções para operações com “blocos”.

1960 – 1970: Amplo otimismo sobre o desenvolvimento da AI. Previsões de que o problema estará resolvido em um período de tempo “curto”. Expressivo financiamento governamental / militar (e.g., DARPA).

Anos 1960 - anos 1970: algoritmos evolutivos surgem em diversos trabalhos (Rechenberg, Schwefel, Fogel, Holland et al.). O clássico livro de Holland é de 1975.

1969: Minsky e Papert – livro *Perceptrons*: críticas que desacreditam o campo das redes neurais, que entra numa “era das trevas”.

1974 – 1980: “Inverno” – decepção com as dificuldades encontradas. Cortes de financiamento. Críticas à AI: John Lucas, John Searle et al.

1976: Weizenbaum – livro *Computer Power and Human Reason* – questões éticas do uso da AI (decisão / escolha).

1980 – 1987: *Boom* da AI – Sistemas especialistas / redes neurais.

1980: Marco para os *sistemas especialistas* (Edward Feigenbaum), baseados em regras (de *experts*). XCON economiza milhões de dólares da Digital Equipment Corporation.

Anos 1980: Financiamento expressivo (EUA, Reino Unido, Japão).

1982: Rede de Hopfield, Mapa de Kohonen – renascimento do paradigma conexionista.

1986: Rumelhart, Hinton, Williams, McClelland / Werbos (1974) – consolidação do processo conhecido como retropropagação de erro (*error backpropagation*).

1986: Trabalho de Farmer, Packard e Perelson sobre redes imunológicas. Nos anos 1990, a área de sistemas imunológicos artificiais se consolida (Forrest, Kephart, Dasgupta et al.). Posteriormente, importantes desenvolvimentos seriam realizados por Timmis, Neal, de Castro, Von Zuben et al. Um fato que merece menção: os dois últimos têm sua história ligada à esta universidade.

1987 – 1993: “Segundo inverno”. Cortes de financiamento / fechamento de empresas de AI.

1993 – 2011: Conquistas importantes.

- Alguns fatores: aumento exponencial (lei de Moore) da capacidade de computação; desenvolvimento da teoria de agentes; métodos probabilísticos ganham fôlego (e.g. métodos bayesianos); avanços teóricos importantes (e.g. consolidação das SVMs).

1995: Kennedy e Eberhart – *Particle Swarm Optimization*.

1997: *Deep Blue* derrota Garry Kasparov, então campeão mundial de xadrez, num *match* de seis partidas.

2005: Um robô de Stanford vence a DARPA Grand Challenge. Ele dirigiu autonomamente por 131 milhas.

2011 – dias atuais: Era do “*big data*” e do aprendizado profundo (*deep learning*).

2011: Watson (IBM) vence um desafio de *Jeopardy!* contra dois *experts* humanos.

2012: Krizhevsky, Sutskever, Hinton: uma rede convolucional profunda ganha a competição ILSVRC na base de dados ImageNet por ampla margem. É um marco de uma revolução que segue em frente.

2016: AlphaGo derrota Lee Sedol, um dos principais mestres de Go do planeta num *match* de cinco partidas.

5. Dedicatória

O curso é dedicado ao Dr. Alan Caio Rodrigues Marques, brilhante jovem pesquisador que, infelizmente, faleceu em agosto de 2020.



6. Referências bibliográficas

ALPAYDIN, E. **Introduction to Machine Learning**. MIT Press. 3rd edition. 2014.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. Springer, 2006.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference and Prediction**. Springer. 2nd edition, 2006.

HAYKIN, S. **Adaptive Filter Theory**. Pearson, 5th edition, 2013.

HAYKIN, S. **Neural Networks and Learning Machines**. Prentice Hall, 3rd edition, 2008.

I. GOODFELLOW, Y. BENGIO, A. COURVILLE, **Deep Learning**, MIT Press, 2016.

WIKIPEDIA