

MAC5921 – Deep Learning

Aula 01 – 10/08/2023

Nina S. T. Hirata

Importante: e-disciplinas

Esta disciplina foi elaborada tendo em mente:

- **Aula ideal:** Construção coletiva de conhecimento
- **Avaliação:** participação ATIVA
- **Viés do prof.:** dados do tipo imagens
- Há pré-requisitos – mais ao final





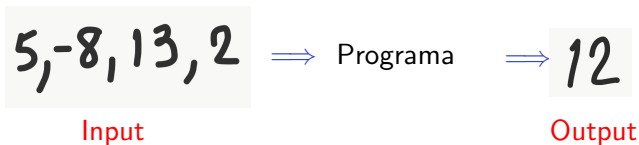


Situar DL

E para isso vamos começar falando de ML

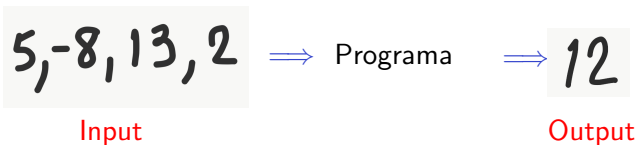
Processamento de dados pode ser caracterizado em termos de **input** e **output**

Exemplo: Calcular a soma de uma certa quantidade de números



Todos aqui sabem escrever um programa que resolve esse problema

Exemplo: Calcular a soma de uma certa quantidade de números



Todos sabem escrever um programa que reconhece caracteres? Alguém sabe?

 \Rightarrow 'a'

 \Rightarrow 'X'

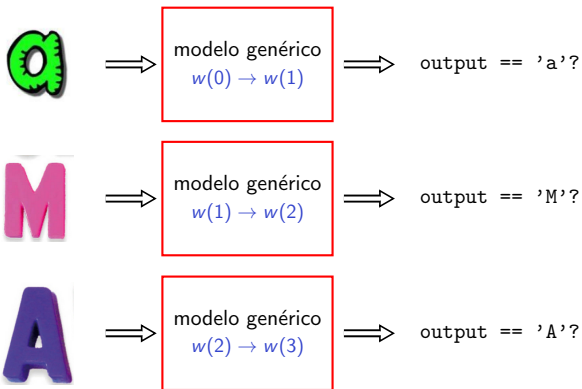
 \Rightarrow 'A'

 \Rightarrow 'd'

 \Rightarrow 'b'

 \Rightarrow 'M'

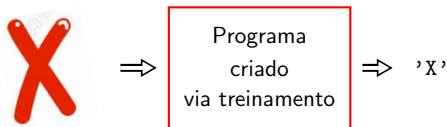
Vamos fazer o computador trabalhar?



O modelo genérico é parametrizado por w

$w(t)$, $t = 0, 1, 2, \dots$: valores dos parâmetros ao longo de iterações

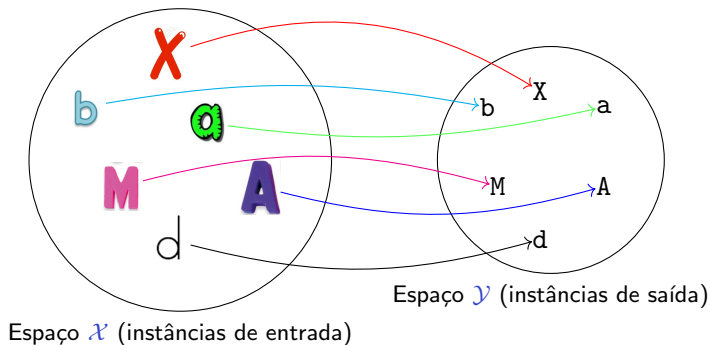
Depois de treinado, o modelo genérico está preparado para gerar o output



Treinamento de modelos

- Precisamos de exemplos de dados entrada-saída
- Faz sentido para problemas nos quais, dada a entrada, dizer qual é a saída é “simples”, enquanto dizer como mapear a entrada para a saída não é
- O modelo genérico deve ser expressivo suficiente (capaz de representar o mapeamento)

ML – perspectiva probabilístico-estatística



$P(\mathbf{x})$: distribuição de probabilidade sobre \mathcal{X}

$P(\mathbf{y})$: distribuição de probabilidade sobre \mathcal{Y}

$P(\mathbf{x}, \mathbf{y})$: distribuição conjunta sobre $\mathcal{X} \times \mathcal{Y}$

Supondo que você conhece $P(\mathbf{x}, \mathbf{y})$, e dado que observou uma certa instância $x \in \mathcal{X}$, qual é a saída mais provável correspondente a x ?

Supondo que você conhece $P(\mathbf{x}, \mathbf{y})$, e dado que observou uma certa instância $\mathbf{x} \in \mathcal{X}$, qual é a saída mais provável correspondente a \mathbf{x} ?

Temos $P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}|\mathbf{y}) P(\mathbf{y}) = P(\mathbf{y}|\mathbf{x}) P(\mathbf{x})$

Bayes' Theorem

$$P(\mathbf{y} | \mathbf{x}) = \frac{P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})}{P(\mathbf{x})}$$

Bayes' Theorem

$$P(\mathbf{y} | \mathbf{x}) = \frac{P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})}{P(\mathbf{x})}$$

Se conhecemos as distribuições, a aposta vencedora é:

$$y^* = \arg \max_{y \in \mathcal{Y}} \{P(\mathbf{y} = y | \mathbf{x} = x)\}$$

$$P(\mathbf{y} | \mathbf{x}) = \frac{P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})}{P(\mathbf{x})}$$

$$P(\mathbf{y} | \mathbf{x}) = \frac{P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})}{P(\mathbf{x})}$$

Discriminativa

Abordagens gerativas × discriminativas

$$P(\mathbf{y} | \mathbf{x}) = \frac{P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})}{P(\mathbf{x})}$$

Discriminativa

Gerativa

Abordagens gerativas × discriminativas

$$P(\mathbf{y} | \mathbf{x}) = \frac{P(\mathbf{x} | \mathbf{y}) P(\mathbf{y})}{P(\mathbf{x})}$$

Discriminativa

Gerativa

Probabilistic machine learning trabalha $P(\mathbf{x}, \mathbf{y})$

GAN – generative adversarial network

Rede neural: estima $P(\mathbf{y} | \mathbf{x})$

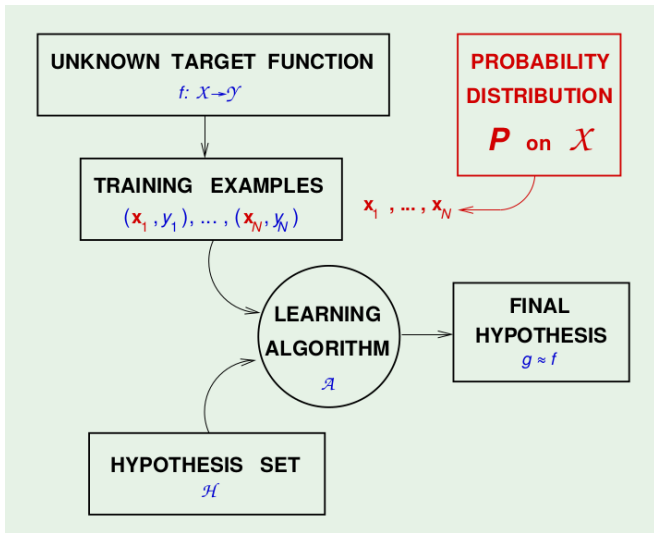
Tipos de aprendizado

Supervisionado: amostras do espaço $\mathcal{X} \times \mathcal{Y}$

Não-Supervisionado: amostras do espaço \mathcal{X} apenas

Por reforço: estado, ação, recompensa

The learning diagram



VC generalization bound

With probability $1 - \delta$ ($\delta > 0$)

$$E_{out} \leq E_{in} + \underbrace{\sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}}_{\Omega}$$

d_{VC}	E_{in}	Ω
small	large	small
↓	↑	↓
large	small	large

N : tamanho do conjunto de treinamento

E_{in} : erro no conjunto de treinamento

E_{out} : erro real, desconhecido

$m_{\mathcal{H}}$: relacionado à complexidade de \mathcal{H} (dimensão VC)

O que aconteceu até chegarmos a DL?

no data
rules

```
if age > 40:
    if is_home_owner:
        print("give a credit")
    else:
        if income > 5000:
            print("give a credit")
        else:
            print("to refuse")
else:
    if education == "university":
        print("...")
    else:
        print("...")
```


O que aconteceu até chegarmos a DL?

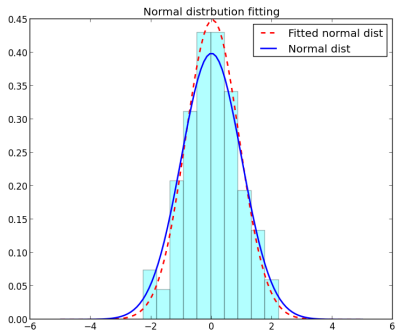
no data

rules

some data

parametric estimation

specific model-based



O que aconteceu até chegarmos a DL?

no data

rules

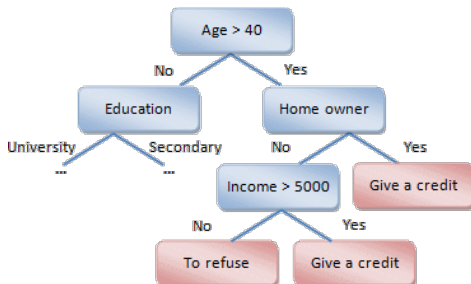
some data

parametric estimation

specific model-based

more data

model induction
(ML algorithms)



O que aconteceu até chegarmos a DL?

no data

rules

some data

parametric estimation

specific model-based

more data

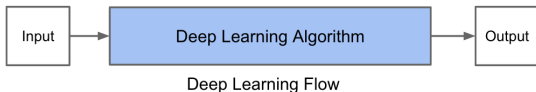
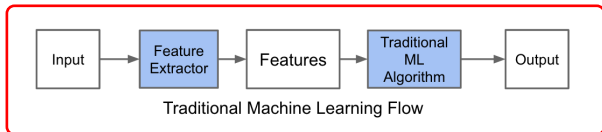
model induction

(ML algorithms)

much more data

feature engineering

ensembles



O que aconteceu até chegarmos a DL?

no data

rules

some data

parametric estimation

specific model-based

more data

model induction

(ML algorithms)

much more data

feature engineering

ensembles

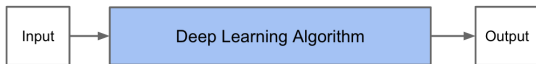
Big Data

representation learning

(deep learning, end-to-end)



Traditional Machine Learning Flow



Deep Learning Flow

O que aconteceu até chegarmos a DL?

no data

rules

some data

parametric estimation

specific model-based

more data

model induction

(ML algorithms)

much more data

feature engineering

ensembles

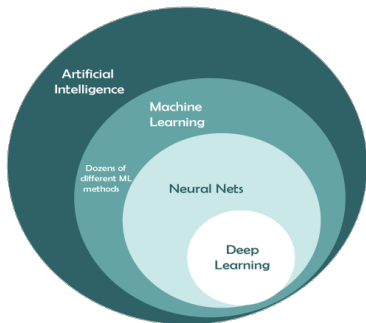
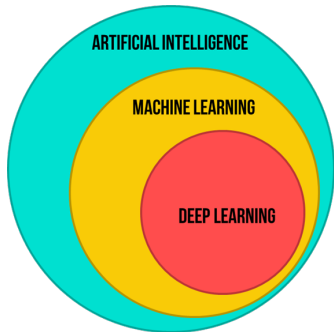
Big Data

representation learning

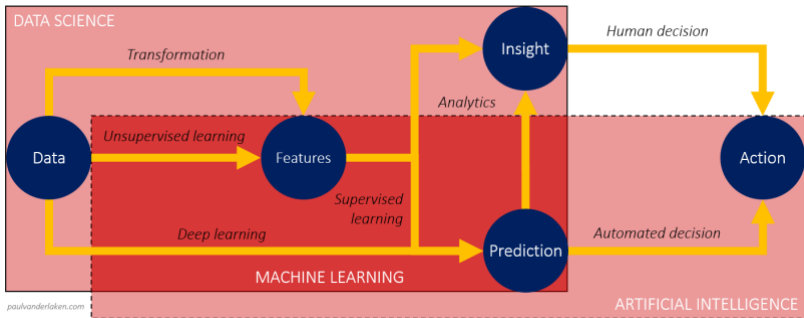
(deep learning, end-to-end)

Data-driven

Qual a diferença entre IA, ML e DL ???

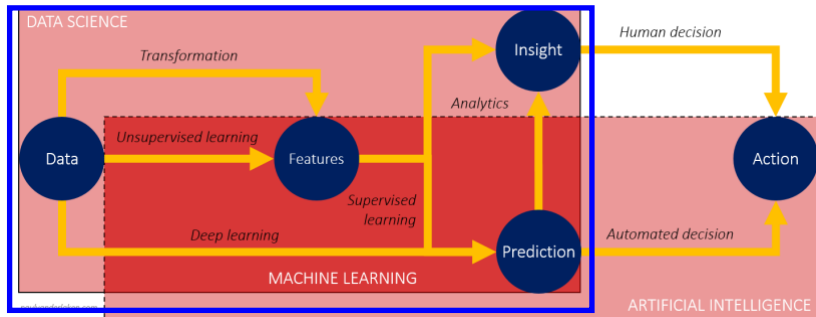


Qual a diferença entre IA, ML, DL, etc ???



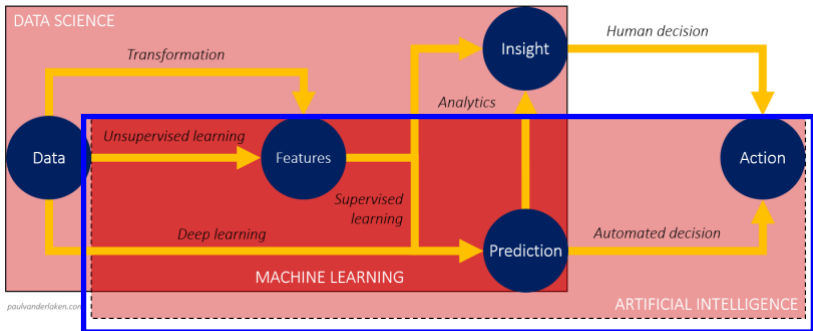
(Author: Paul van der Laken)

Qual a diferença entre IA, ML, DL, etc ???



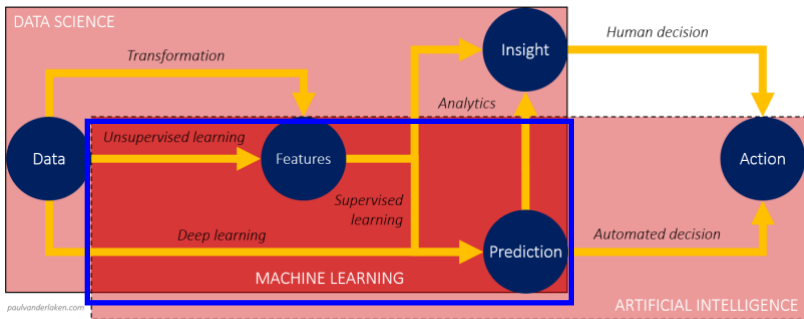
(Author: Paul van der Laken)

Qual a diferença entre IA, ML, DL, etc ???



(Author: Paul van der Laken)

Qual a diferença entre IA, ML, DL, etc ???



(Author: Paul van der Laken)

Por que DL decolou?

Big Data: muitos, muitos, muitos dados

Poder de computação: Memória, CPU, GPU

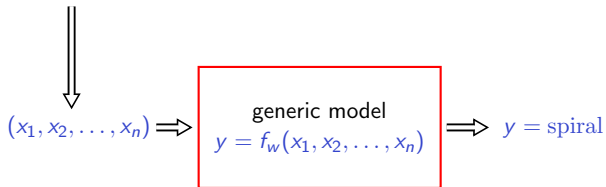
Free software: Caffe, TensorFlow, PyTorch

Machine Learning tradicional

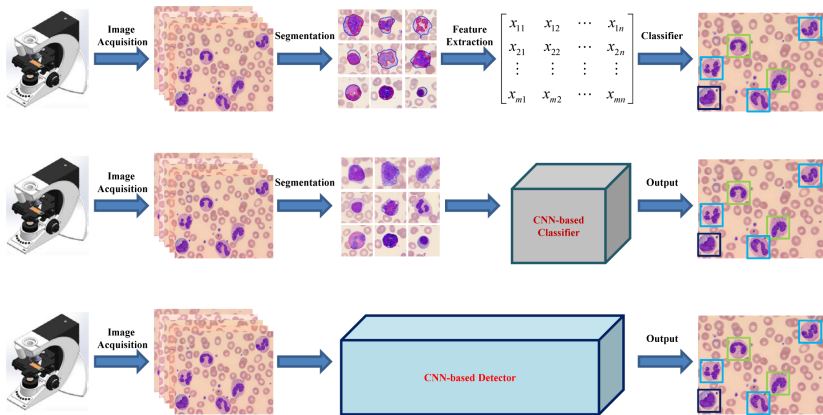


Características do objeto

- concentration
- asymmetry
- smoothness
- entropy
- spirality, etc

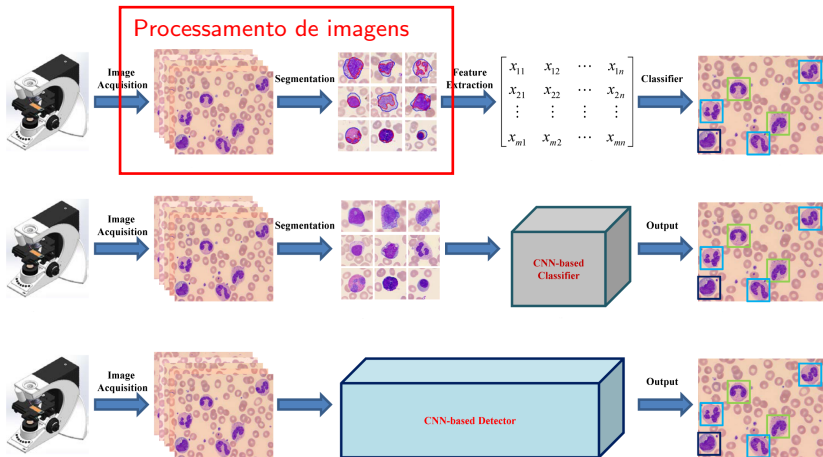


DL – vãos mais altos



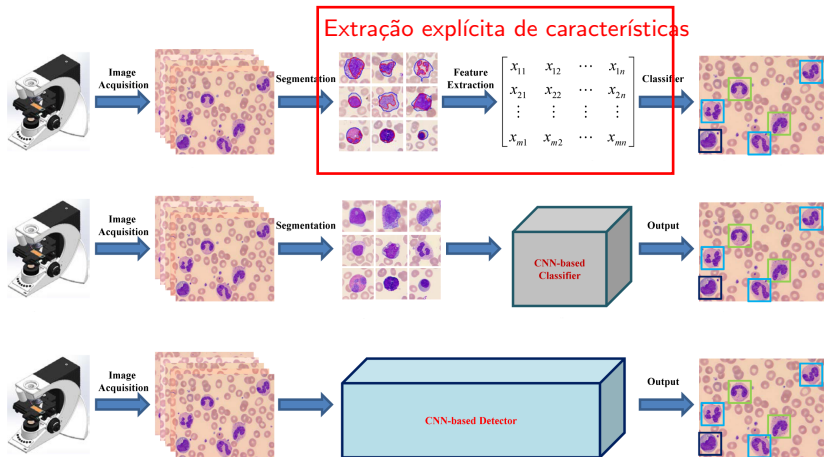
Source: adapted from <https://doi.org/10.1371/journal.pone.0218808>

DL – vãos mais altos



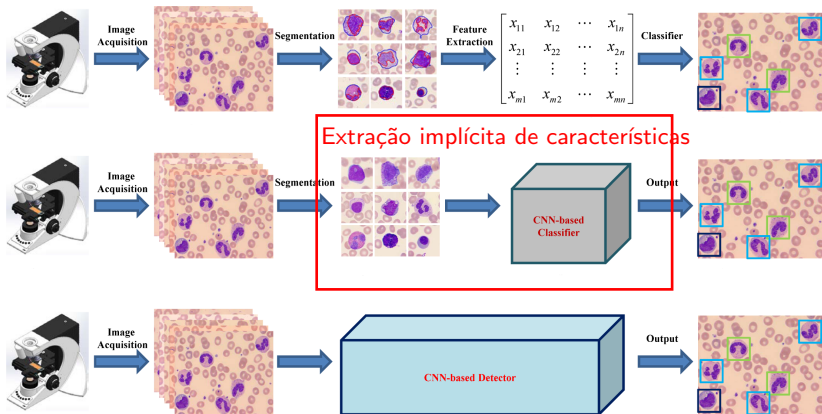
Source: adapted from <https://doi.org/10.1371/journal.pone.0218808>

DL – vãos mais altos



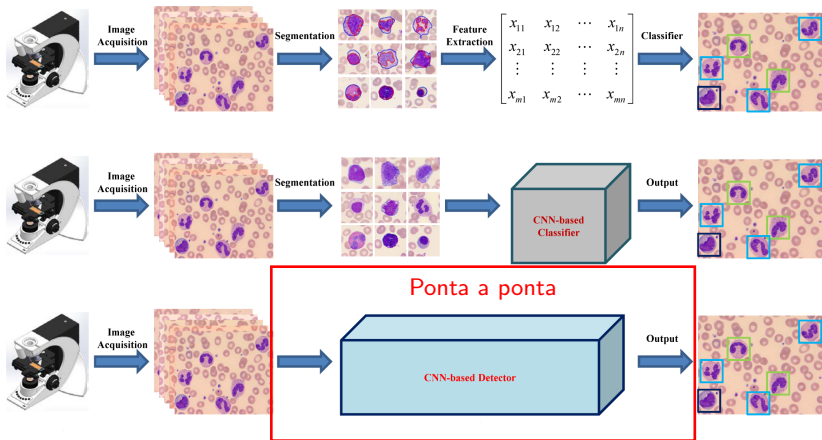
Source: adapted from <https://doi.org/10.1371/journal.pone.0218808>

DL – vãos mais altos



Source: adapted from <https://doi.org/10.1371/journal.pone.0218808>

DL – vãos mais altos



Source: adapted from <https://doi.org/10.1371/journal.pone.0218808>

GPT / ChatGPT

<https://thispersondoesnotexist.com/>

Segment Anything: <https://segment-anything.com/>

DALL·E 2: <https://openai.com/dall-e-2>

Ajude a completar este slide

O que é esperado que você já saiba

Problema de classificação

Problema de regressão

Função de perda

Minimização de função por gradiente descendente

Alguns algoritmos de aprendizado de máquina

Alguma noção de redes neurais e do algoritmo de backpropagation

Para que servem os conjuntos de treinamento, validação e teste

Overfitting / detecção de overfitting

Regularização

Seleção de modelos

Prática com Python e Numpy