# Curso: Aprendizagem de Máquina em Inteligência Artificial

Disciplina: Aprendizado Não Supervisionado

Prof. Marcelo Novaes de Rezende



#### Apresentação do professor e alunos

Marcelo Novaes de Rezende tem graduação em Engenharia Civil e Computação, além de Mestrado e Doutorado (todos os cursos na USP, em São Carlos). Desde os anos 80 trabalha com (e gosta muito de) computação científica (aquela cheia de matrizes e vetores). Tem carreira acadêmica, como desenvolvedor, executivo e empreendedor em tecnologia. Fundou a Linked para ministrar cursos de assuntos que gosta e participar de projetos desafiadores em Data Science.



Marcelo Novaes de Rezende
PROFESSOR/DESENVOLVEDOR

#### Apresentação do curso

Disciplina: Aprendizado Não Supervisionado

_							
	SEMA NA ENCONT RO	DATA	CONTEÚDO TEÓRICO E PRÁTICO				
	01	20/10/20	Teoria : Revisão Aprendizado de máquina, normalização				
			Prática : Exemplos de aprendizado de máquina e normalização em Python				
	02	27/10/20	Teoria: Aprendizado não supervisionado de máquina e aplicações, K-Means 1				
			Prática : K-Means em Excel				
	03	03/11/20 Teoria : K-Means Parte 2					
			Prática : K-Means em Python, qualidade de cluster				
	04	10/11/20	Teoria: KNN, regularização, dropout e batch norm				
			Prática : KNN				
	05	17/11/20	Teoria : PCA e ICA Parte 1				
			Prática : Autovalores e autovetores em Python, correlação, PCA				
	06	24/11/20	Teoria : PCA e ICA Parte 2				
			Prática: PCA e ICA exemplos em Python				
	07	01/12/20	Teoria :Bias e Variância, Regras de Associação				
			Prática : visualização de Bias e Variância				
	08	08/12/20	Revisão teoria				
			Revisão Prática				

Pós-Graduação em Aprendizagem de Máquina em Inteligência Artificial— Disciplina Aprendizado Não Supervisionado

#### Apresentação do curso

#### **BIBLIOGRAFIA BÁSICA**

GÉRON, Aurélien. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Sebastopol: O'Reilly, 2017.

KELLEHER, John D.; MAC NAMEE, Brian; D'ARCY, Aoife. Fundamentals of Machine Learning for Predictive Data

Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge: MIT Press, 2015.

TAN, Pang-Ning; STEINBACH, Michael; KARPATNE, Anuj; KUMAR, Vipin. Introduction to Data Mining. 2 ed. New

York: Pearson, 2019.

#### **BIBLIOGRAFIA COMPLEMENTAR**

CASTRO, Leandro de; FERRARI, Daniel G. Introdução à Mineração de Dados:Conceitos Básicos, Algoritmos e

Aplicações. São Paulo: Saraiva, 2018.

SILVA, Leandro Augusto da; PERES, Sarajane Marques; BOSCARIOLI, Clodis.

Introdução à Mineração de Dados. Rio de Janeiro: Elsevier, 2016.

VANDERPLAS, Jake. Python Data Science Handbook. Sebastopol: O'Reilly,

2017.

#### Critério de avaliação

### Avaliação

Trabalhos em grupo, a partir da aula 02

# Revisão Machine Learning e tipos de aprendizado

- •IA,
- Machine Learning e
- Deep Learning



# **ARTIFICIAL INTELLIGENCE (IA)**



"Campo de estudo que procura entender e emular comportamento inteligente em termos de processos computacionais" (Schalkoff, 1990)



#### **COMPORTAMENTOS INTELIGENTES:**

- ✓ RACIOCINAR
- ✓ RECONHECER PADRÕES
- ✓ APRENDER
- ✓ FALAR



## POR QUE IA AGORA?



ABUNDÂNCIA DE DADOS

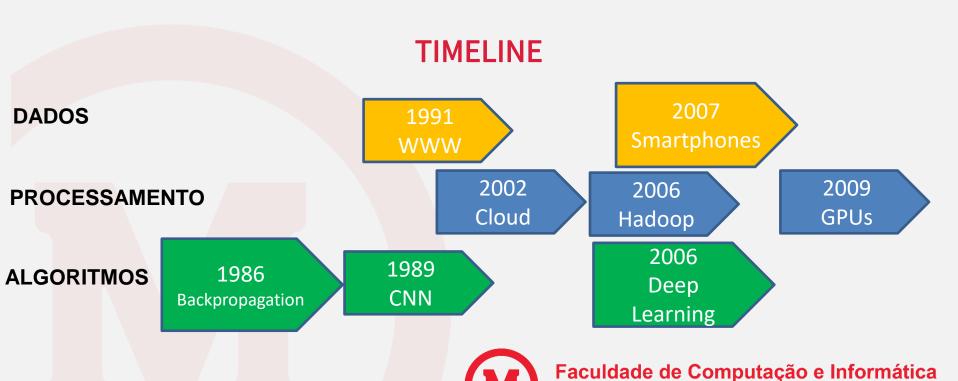
+

CAPACIDADE DE PROCESSAMENTO

+

**MELHORES ALGORITMOS** 





**Mackenzie** 

# IA X Machine Learning



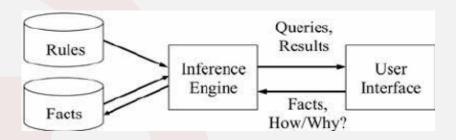
Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel 1959





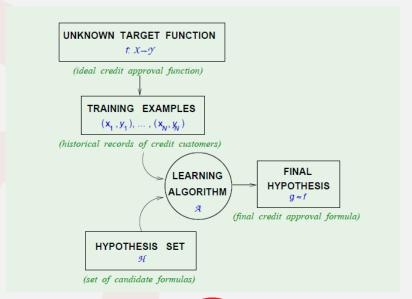
## Sistema baseado em regras.



É IA, mas **não** é Machine Learning, pois a inteligência foi explicitamente definida por regras!



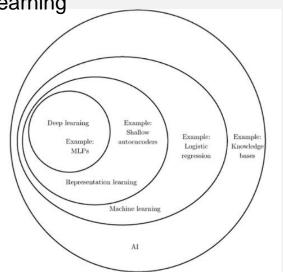
#### Machine Learning Summary



FONTE: ABU-MOSTAFA et al. Learning from Data

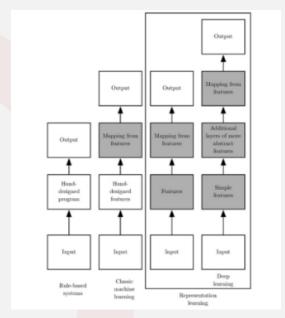


IA, Machine Learning, Deep Learning



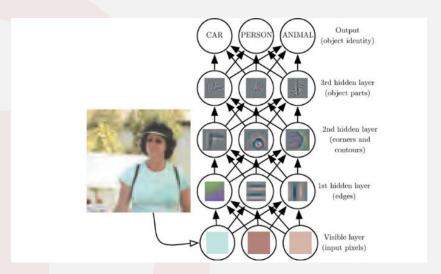


IA, Machine Learning, Deep Learning





#### Classificação com MLP





Impacto das diferentes representações (features)

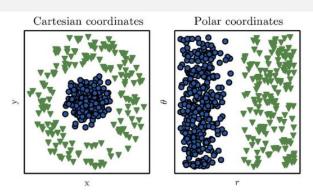


Figure 1.1: Example of different representations: suppose we want to separate two categories of data by drawing a line between them in a scatterplot. In the plot on the left, we represent some data using Cartesian coordinates, and the task is impossible. In the plot on the right, we represent the data with polar coordinates and the task becomes simple to solve with a vertical line. (Figure produced in collaboration with David Warde-Farley.)



Exemplos de utilização de ML/DL



## Utilizações de Machine Learning

Carros Autônomos Diagnósticos Médicos

Investimentos em bolsa de valores

Detecção de fraudes

Processamento de linguagem natural (NLP)

Recomendações

Reconhecimento Facial/de padrões Avaliação de crédito

Automação de atividades humanas em processos



Faculdade de Computação e Informática **Mackenzie** 

Brincando com carro autônomo (para casa)

<a href="https://www.youtube.com/watch?v=ppFyPUx9RIU">https://www.youtube.com/watch?v=ppFyPUx9RIU</a>



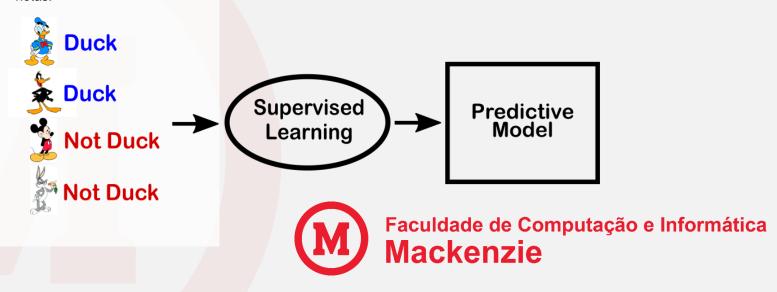
- APRENDIZADO SUPERVISIONADO
- APRENDIZADO NÃO SUPERVISIONADO
- APRENDIZADO POR REFORÇO



#### APRENDIZADO SUPERVISIONADO:

Sabemos a resposta correta (output) para cada entrada no data set (input). Acreditamos haver relação entre inputs e outputs.

Exemplo: Temos as notas de ensino médio dos alunos e sabemos se cada um deles foi ou não aprovado nos vestibulares. Podemos treinar a máquina mostrando instâncias de aprovados ou não aprovados para que, depois, possamos prever a aprovação ou não de um aluno nos vestibulares em função de suas notas.

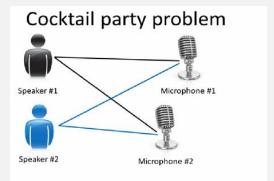


#### APRENDIZADO NÃO SUPERVISIONADO:

Para o data set procuramos derivar alguma estrutura entre os dados. Porém, o resultado não pode ser avaliado como "correto" ou "errado". Os dados não são **rotulados**.

Exemplo: Temos as notas de ensino médio dos alunos e queremos agrupá-los por semelhança (?).

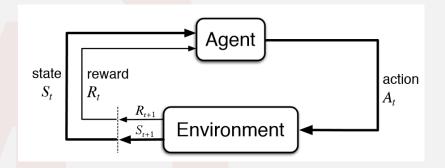
Exemplo clássico de aprendizado não supervisionado: separar as vozes que falam simultaneamente





#### **APRENDIZADO POR REFORÇO:**

Um agente interage com um ambiente. Ele está em um estado s e toma uma ação A, que o leva a um novo estado e a uma recompensa. O objetivo é maximizar a soma das recompensas. O aprendizado leva ao mapeamento otimizado entre estados e ações.





## Regressão x Classificação

#### REGRESSÃO

Queremos um output contínuo com base no input.

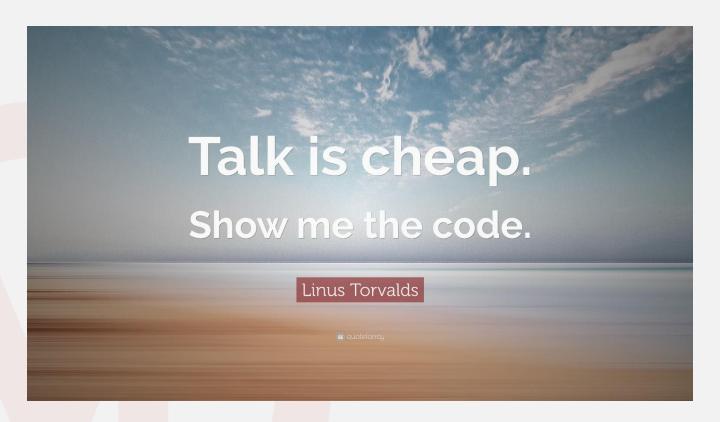
Exemplo : queremos prever o valor de uma ação.

#### **CLASSIFIC**AÇÃO

Queremos apenas mapear um input para determinada classe.

Exemplo : queremos apenas saber se a ação vai subir ou descer.









Inicialmente, vamos revisar as estruturas do Python com o notebook python.ipynb

Agora vamos trabalhar com Datasets usando o Pandas (pandas.ipynb)

Trabalhar com a biblioteca numpy é essencial na computação científica (numpy.ipynb)

Agora, um exemplo de aprendizado supervisionado e não supervisionado com a biblioteca sklearn (sklearn.ipynb)



### **Standardization**

Para aplicarmos algoritmos de Data Science, muitas vezes é importante efetuarmos "feature scaling", uma maneira de trazer as diversas features para escalas próximas. Normalmente, diferenças numéricas de ordem de grandeza entre as features atrapalham os algoritmos que envolvem otimização (minimizar o erro, por exemplo).

Uma das maneiras possíveis, de fazer o "feature scaling" é por "standardization (Z-score normalization)":

Exemplo

$$Xs = (x - x_{medio})/\sigma$$
 Xs é a feature X após standardization,

$$x_{medio}$$
 é o (1/m)  $\sum x$ 

$$\sigma$$
  $\acute{e}$   $\sqrt{\Sigma(x-xmedio)^2/m}$ 

		-0,414746544 -0,721966206 -0,875576037 0,19969278 0,660522273 2,350230415				
Feature sem standardization			Feature c/standardization			
2	m	10	-0,875576037	m=10		
3	Média	7,7	-0,721966206	Média=	0	
4	Desv. Padrão c/n=10	6,51	-0,568356375	Desvio Desv. Padrão c/n=10	1	
5			-0,414746544			
3			-0,721966206			
2			-0,875576037			
9			0,19969278			
12			0,660522273			
23			2,350230415			
14			0,967741935			

Existe também min-max scaling....como é?

Observe que após standardization, a média da feature é zero e o desvio padrão é 1



Praticar standardization e min max scaler com sklearn Standardization.ipynb



Buscando outliers no dataset (algoritmo hiper básico)

Vamos partir da distância de cada amostra ao centróide e analisar possíveis outliers

Outlier.ipynb



## Até a próxima aula

## **OBRIGADO!**

Prof Marcelo Rezende email rezendemn@gmail.com