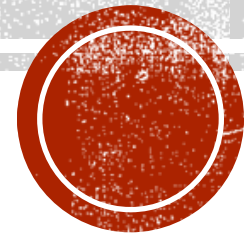


MACHINE LEARNING



Doutoranda: Natália Silva Morosini

Disciplina: LGN5831

Piracicaba

2019

O QUE É?

- Tarefa (capacidade) computacional;
- Correlações entre dados anotados;
- Relacionamento entre *inputs* (quantidades observáveis) e *outputs*;
- Algoritmos computacionais eficientes;
- Criação de modelos estatísticos – inferência de anotações para novos dados;
- Requer grande quantidade de dados.

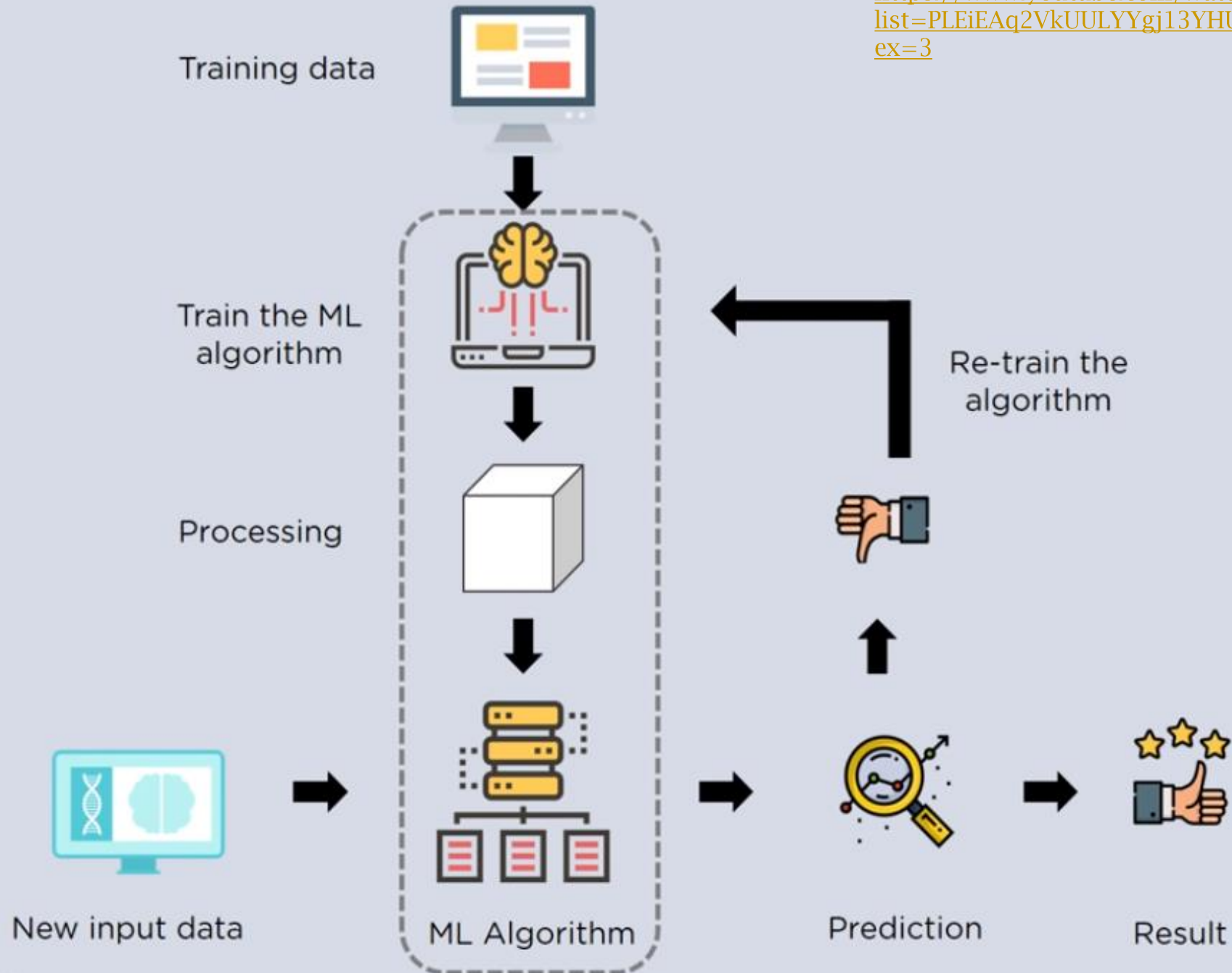


CARACTERÍSTICAS - CONCEITOS

- TRAINING DATA = dados usados no training (*input*);
- FEATURES = quantidades observáveis imputadas em um algoritmo;
- ALGORITMO = classifica os *features* em outputs;
- TRAINING = processo automático de construção de modelos.



<https://www.youtube.com/watch?v=HgBpFaATdoA&list=PLEiEAq2VkUULYYgj13YHUWmRePqiu8Ddy&index=3>



CATEGORIAS - SUBCLASSES

■ As técnicas de *machine learning* podem ser classificadas em 4 categorias principais:

- SUPERVISED LEARNING

- UNSUPERVISED LEARNING

- SEMI-SUPERVISED LEARNING

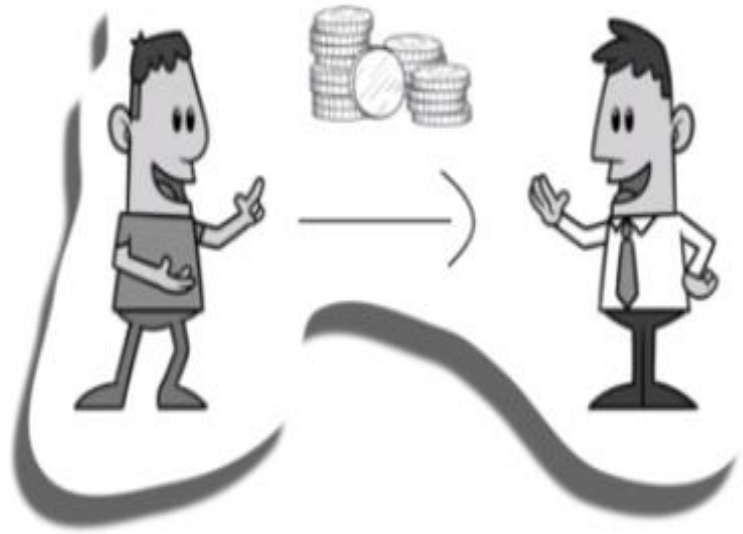
- REINFORCEMENT LEARNING



SUPERVISED LEARNING

- Valor da variável *output* tem que ser conhecido para cada amostra;
- **Objetivo de prever um *output* ou um destino conhecido;**
- Classificação;
- Resultado: Training data – par de valores *input e output*,
- Utiliza algoritmos:
 - Regressão linear ;
 - K Nearest Neighbours;
 - Naive Bayes;
 - Decision Tree.





3 GRAMS



7 GRAMS

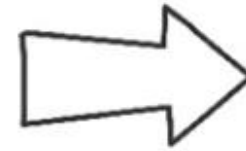
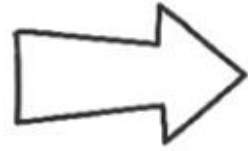


4 GRAMS

WEIGHT = FEATURE

CURRENCY = LABEL





1 RUPEE

LABELED DATA

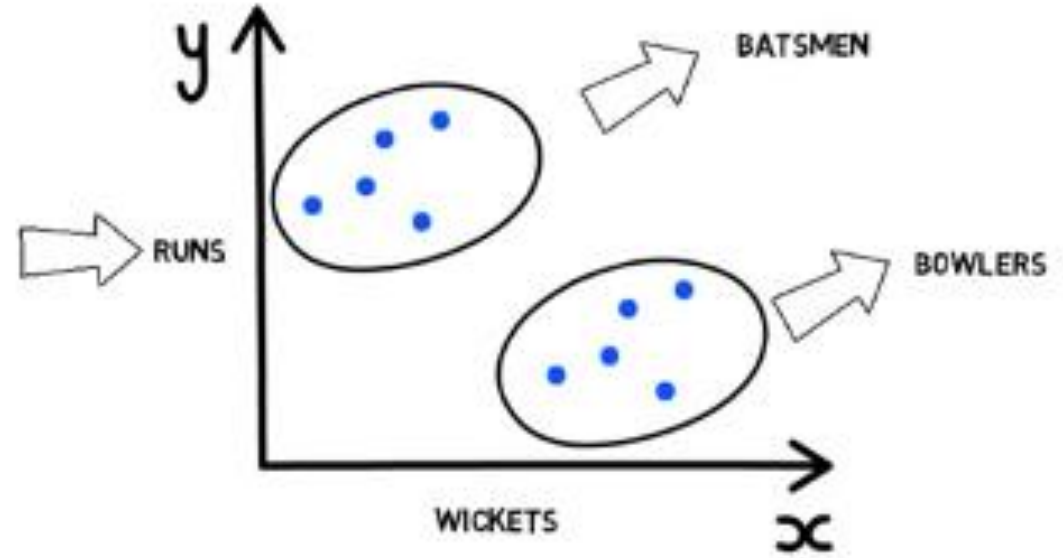
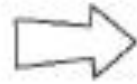


UNSUPERVISED LEARNING

- Não há outputs para serem preditos;
- Requer apenas valores *'features'*;
- **Objetivo: encontrar padrões dentro dos dados;**
- Algoritmos descobrem uma estrutura oculta nos *training datas*:
 - K-mers;
 - Hierarchical;
 - Apriori;
 - Genetic Algorithms.



NAMES	SCORES	WICKETS



NO LABELED DATA



SEMI-SUPERVISED LEARNING

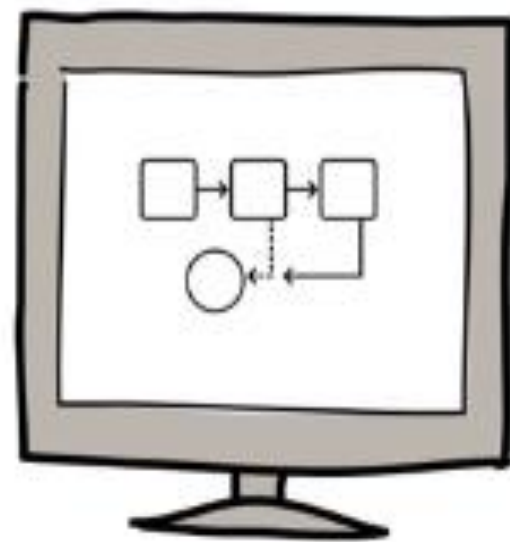
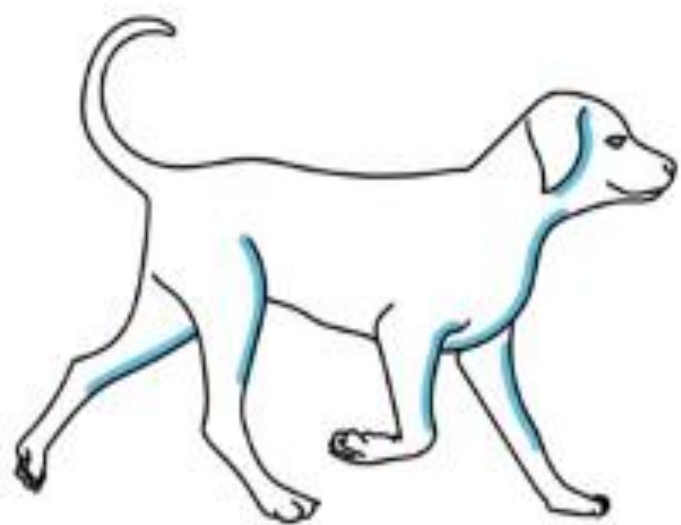
- Combinam um pequeno conjunto de *training datas* supervisionados com um amplo conjunto dos não supervisionados;
- Rótulos para apenas um pequeno subconjunto de dados.



REINFORCEMENT LEARNING

- Não há rótulos;
- A forma de supervisão vem de um sinal de recompensa que informa ao agente o quão bem ele está indo, mas não diz qual deveria ter sido a ação correta;
- *Better decisions*





CAT



DOG

PROBLEMAS

- Classificar os *inputs* em uma das principais categorias;
- *Training* concluído: prevê a classe dos dados a partir dos *features*.
- Classificadores:
 - Multi-class;
 - Binary classification;
 - One-class density estimation (anomaly detection).



DESIGN

- Funcionamento do algoritmo nos dados
- Objetivo do algoritmo do *machine learning*: gerar o output correto para os pontos de dados;
- A capacidade do modelo de prever o *output* correto = generalização;
- Generalização: combinação entre a complexidade do modelo e da função subjacente aos dados.



APLICAÇÃO

SCENARIO - 1

Facebook
Face Recognition



SCENARIO - 2

Netflix Movie
Recommendation



SCENARIO - 3

Fraud
Detection





HEY SIRI

THANKS

OKAY.
I WILL REMIND YOU



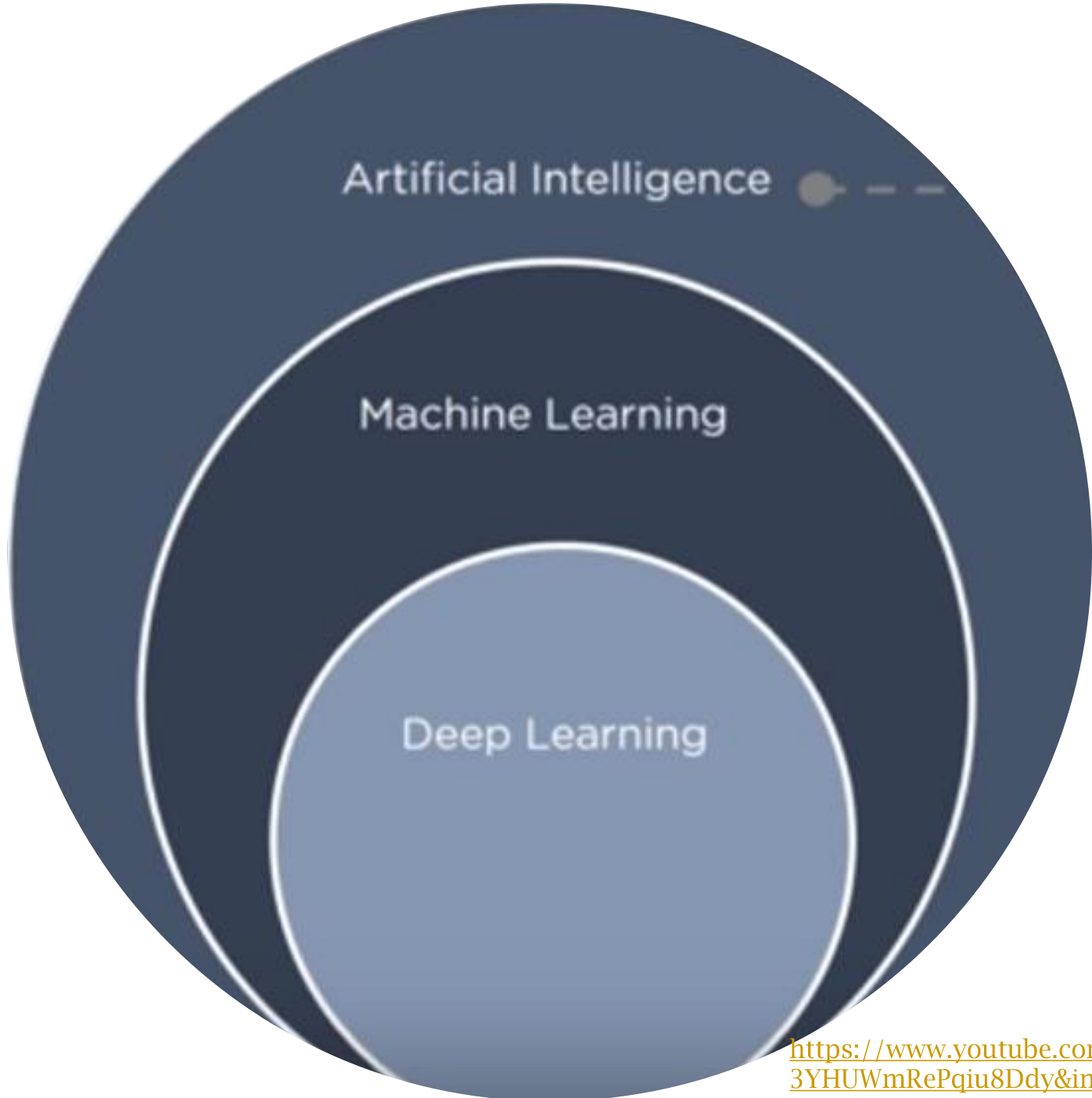
HEALTHCARE



APLICAÇÃO

- Epigenética;
- Epimutações;
- Metilação de histonas;
- Transição entre estados de heterocromatina e eucromatina;
- MOTIFS;
- Farmacogenética.





<https://www.youtube.com/watch?v=HgBpFaATdoA&list=PLEiEAq2VkUULYYgi13YHUWmRePqiu8Ddy&index=3>



**OBRIG
ADA!**

