

Machine Learning

Daniela Maria Uez
dani.uez@gmail.com

18 JULHO 2019

Machine Learning (ML)

Aprendizagem de máquina (AM)

É um campo de estudo que dá aos computadores a capacidade de aprender sem estarem programados explicitamente para isso

[Arthur L. Samuel. 1959]

Estudo dos algoritmos que melhoram sua performance **P** em uma tarefa **T** através da experiência **E**

[Tom Mitchell , 1997]

Também é chamado **Aprendizado Indutivo**

O que é aprender?

- Adquirir conhecimento
- Adquirir habilidade prática
- Ter melhor compreensão de algo

- Aprender é **generalizar** → **indução**

Inferência dedutiva e indutiva (I)



Dedução: do geral para o particular

- Todos os pássaros têm penas
- O pica-pau é um pássaro
- Logo, o pica-pau tem penas.

Inferência dedutiva e indutiva (II)

Indução: do particular para o geral

- Os cisnes que eu vejo na TV são brancos
- Logo, todos os cisnes são brancos

É uma boa aproximação.



Inferência dedutiva e indutiva (II)

Indução: do particular para o geral

- Os cisnes que eu vejo na TV são brancos
- Logo, todos os cisnes são brancos

É uma boa aproximação.

Desde que os exemplos escolhidos sejam suficientes.

A inferência indutiva é um dos principais métodos usados para derivar conhecimento novo e prever eventos futuros



Tipos de AM

- **Aprendizagem supervisionada:** um “professor” guia o aprendizado
Ex: detectar spams
- **Aprendizagem não-supervisionada:** não há interferência humana
Ex: recomendar livros com base no que já foi comprado
- **Aprendizagem por reforço:** alguém diz se o resultado ficou bom ou não, mas não como melhorá-lo
Ex: andar no labirinto

Terminologia

- **Componentes da Entrada**

- Conceitos: “coisas” que podem ser aprendidas
- Instâncias: exemplos individuais e independentes de um conceito
- Atributos: medidas de características de uma instância

Terminologia

- **Tipos de Aprendizado de Máquina (Objetivos da Mineração de Dados)**
- **Aprendizado supervisionado (Atividades de Predição)**
- **Classificação: previsão de classes discretas pré-definidas**
- **Regressão: previsão de um valor numérico contínuo**
- **Aprendizado não-supervisionado (Atividades de Descrição)**
- **Agrupamentos: agrupar instâncias similares em aglomerados**
- **Regras de associação (Atividades de Descrição)**
- **Detecção de associações entre atributos**
- **Mais geral que a Classificação: qualquer associação entre atributos, não apenas com uma classe específica**

Aprendizagem supervisionada

- Utiliza um conjunto de treino com exemplos rotulados (positivo ou negativo)
- O objetivo é treinar um modelo que possa prever rótulos desconhecidos em outros exemplos (o conjunto de teste)
- Exemplos:
 - **Classificação:** rótulos finitos e não ordenados - reconhecimento de padrões
 - Reconhecimento de face, voz, escrita
 - Diagnóstico médico, score de crédito
 - **Regressão:** mostram como uma variável evolui em relação a outras
 - Análise temporal (finanças, meteorologia)

Aprendizagem supervisionada

- Principais algoritmos:
 - Regression
 - Decision Tree
 - Random Forest
 - KNN
 - Logistic Regression

Aprendizagem não-supervisionada

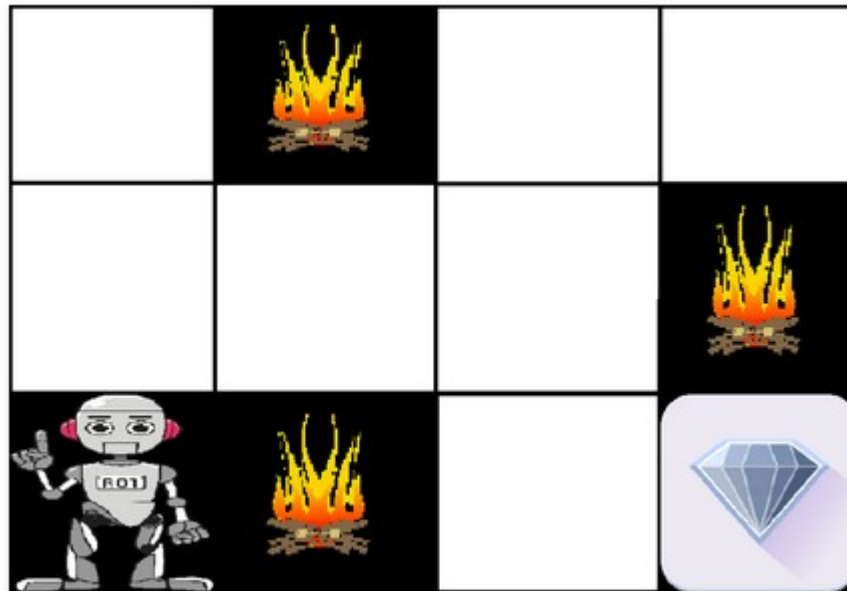
- O objetivo é observar similaridades entre os objetos e incluí-los em grupos apropriados
- Anomalias: objetos que diferem largamente de todos os grupos
- Exemplos:
 - Análise do carrinho de compras: propabilidade do cliente que comprou **x** comprar **y**
 - Detecção de anomalias
 - Visualização de dados

Aprendizagem não-supervisionada

- Principais algoritmos:
 - Redes neurais
 - K-Means Clustering
 - Transfer Learning
 - K-Nearest Neighbors
 - VP Tree

Aprendizagem por reforço

- Não há conjunto de treinamento, rotulado ou não
- Agentes procura agir de forma a maximizar a recompensa
-



Aprendizagem por reforço

- Exmplos:
 - Jogo de gamão, damas
 - Movimentação de robôs, carros autônomos

Referências

- Bases de Dados -

- Censo do Legislativo - <http://dados.gov.br/dataset/censo-do-legislativo>
- IBGE - https://downloads.ibge.gov.br/downloads_estatisticas.htm
- 33 Brilliant And Free Data Sources Anyone Can Use - <https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/#39b6b626b54d>
- Portal da transparência – [dados do governo federal](#)

Referências

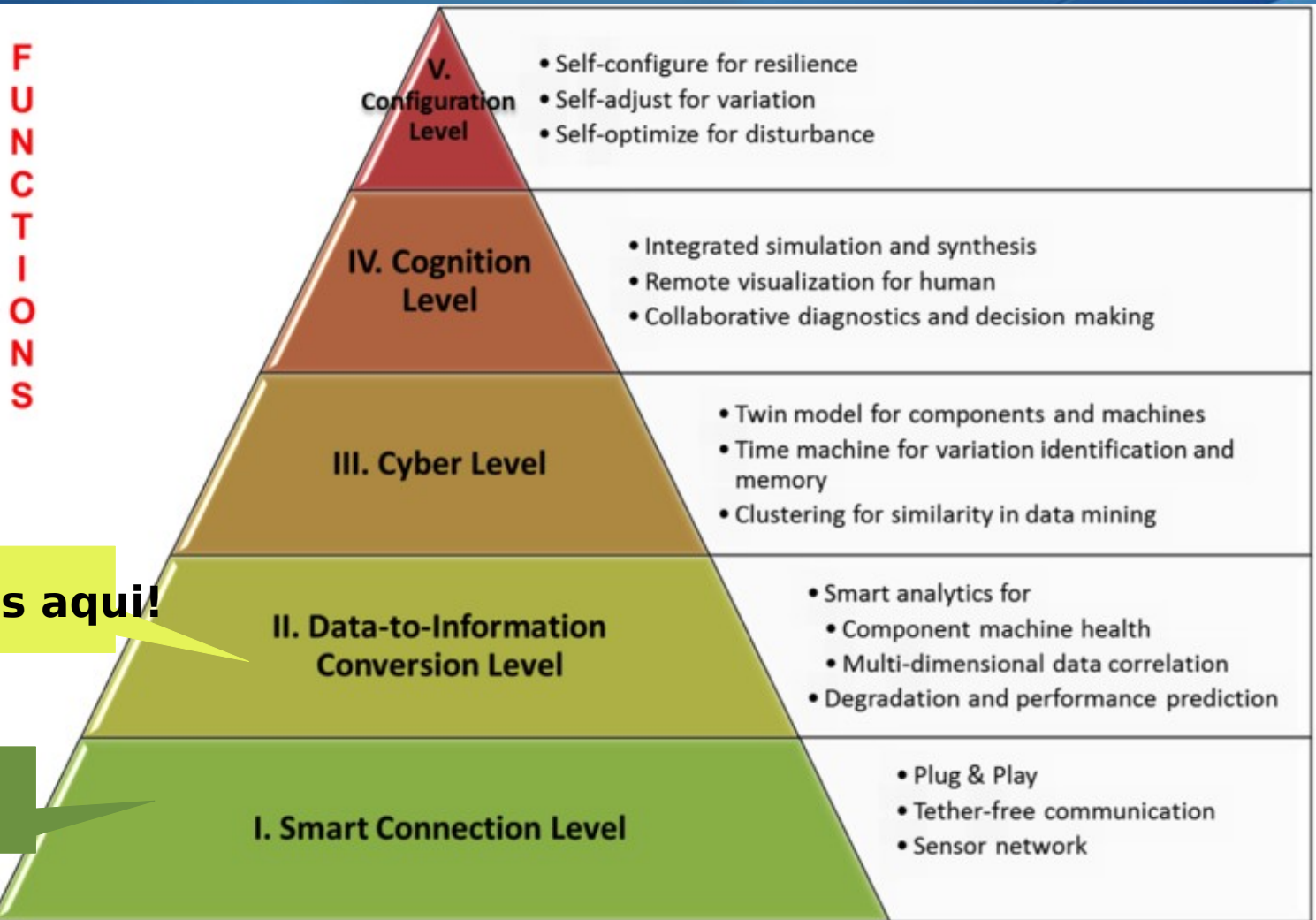
Orange - <https://orange.biolab.si/>

Tutorial sobre Orange:

<https://orange3.readthedocs.io/projects/orange-development/tutorial.html>

Dados da aula: www.uez.com.br/ucs

Arquitetura 5C



Estamos aqui!

IIoT