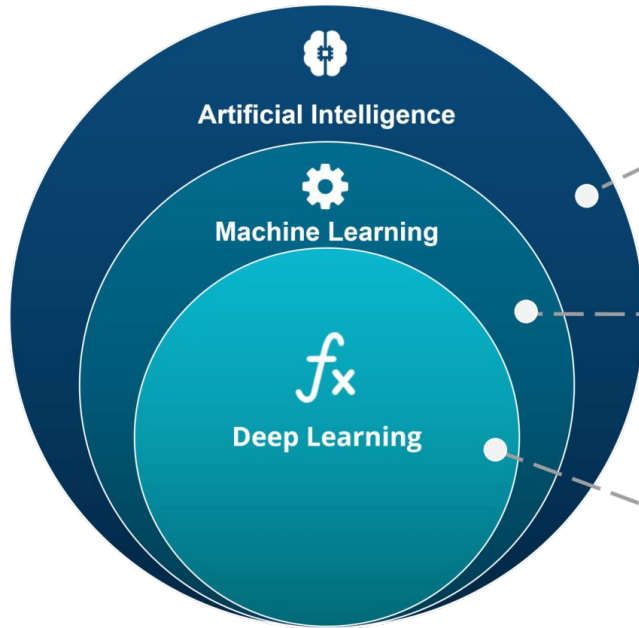


MACHINE LEARNING

# Algoritmos Regressores



Prof. Altino Dantas



## ARTIFICIAL INTELLIGENCE

A technique which enables machines to mimic human behaviour

## MACHINE LEARNING

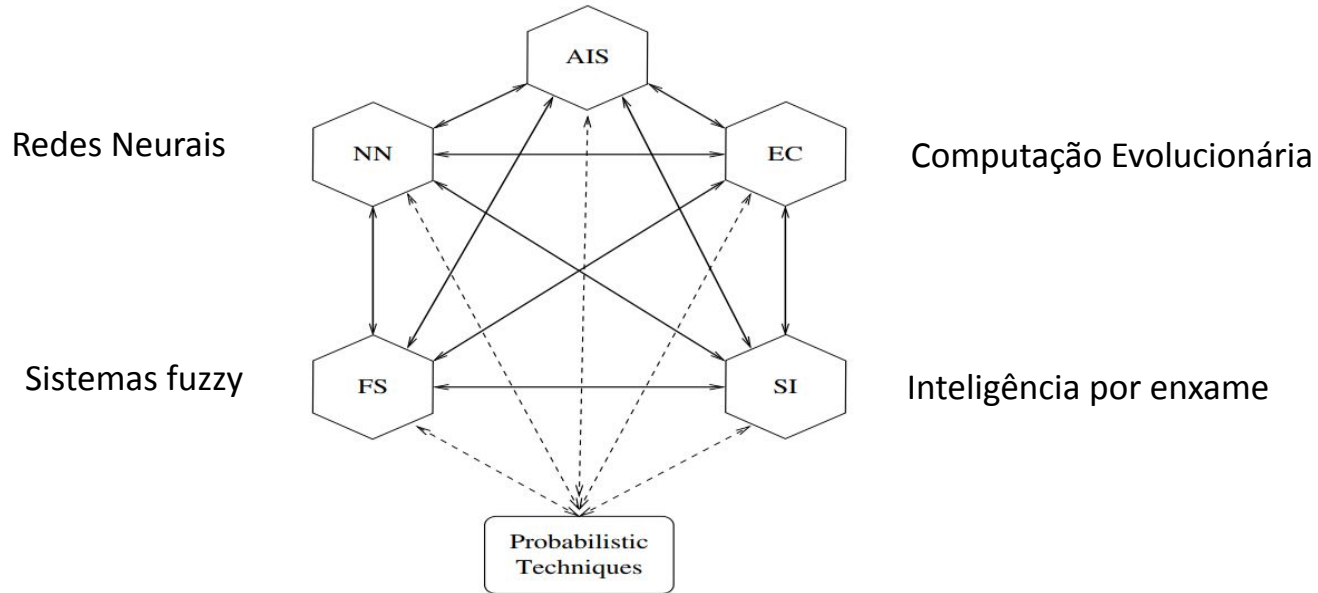
Subset of AI technique which use statistical methods to enable machines to improve with experience

## DEEP LEARNING

Subset of ML which make the computation of multi-layer neural network feasible

# Paradigmas de Inteligência Computacional

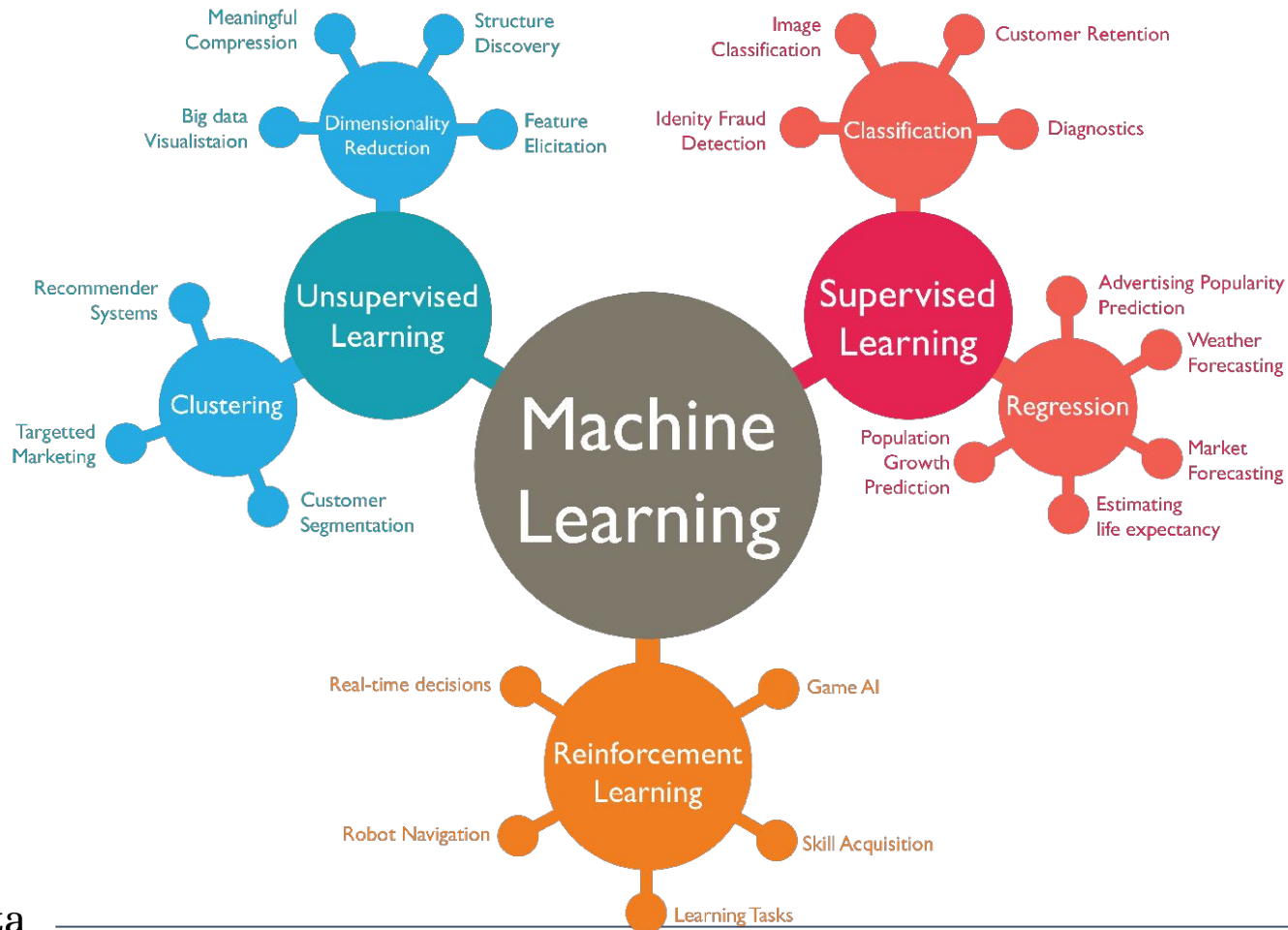
Sistemas Imunológicos Inteligentes



# Aprendizado de máquina

- É um **subconjunto de técnicas** de IA ou aplicação de IA para resolução de problemas;
- Algoritmos que **se aprimoram** de forma automática a partir de exemplos;
- São modelos matemáticos ajustados com **amostras de dados** de um certo contexto, chamados de dados de treinamento;
- A partir dos dados conhecidos, são capazes de fazer **previsões** sobre dados desconhecidos, **não** sendo **explicitamente** programados para tal;
- A partir de dados conhecidos, são capazes de encontrar relações desconhecidas (reconhecer padrões).





# Quanto ao tipo de aprendizado

- **Supervisionado**
  - Os dados são conhecidos e anotados (saídas conhecidas);
  - Diferentes partições de dados: treino, validação, teste.
- **Não-supervisionado**
  - Os dados são conhecidos mais não são anotados;
  - Tipicamente usado para agrupamento ou redução de dimensionalidade;
- **Reforço**
  - Os dados são ambientais e o aprendizado gerado pela avaliação dos resultados das ações tomadas.

# Quanto a tarefas | **supervisionado**

- **Classificação:**

- Inferência de uma **categoria** para um conjunto de características;
- Múltiplas classes (discretas);
- Ex.: classificar um animal por uma foto, onde cada pixel é um característica;

- **Regressão**

- Inferências de um valor contínuo para um conjuntos de características;
- Única “classe” com valor contínuo;
- Quantidade de leitos de UTI que serão ocupados por pacientes de covid-19, em uma certa data, no Estado de Goiás, considerando diversas variáveis epidemiológicas.



O modelo é construído aqui

Dados de treino



Dados de validação

O modelo pode sofrer  
algumas validações  
prévias aqui

O modelo é testado aqui

Dados de teste

O modelo é  
operacionalizado aqui

Dados novos  
Ambiente produtivo

[Hackinganalytics.com](https://hackinganalytics.com)





# Forma de definição dos conjuntos

- Divisão Treino/teste
- Validação cruzada *k-Fold*
- Validação cruzada por *Leave-one-out*

# Quanto a tarefas | **não-supervisionado**

- **Clusterização:**
  - Agrupamento de dados por características comuns;
  - Ex.: Sistema de recomendação.
- **Redução de dimensionalidade**
  - Transformação de um espaço de alta dimensionalidade para um espaço com menos dimensões, preservando característica dos dados;
  - Ex.: Selecionar características mais relevantes para então usar um classificador.

# Não confundir

- **Redução de dimensionalidade:** gerar um novo dataset com menos dimensões do que o original;
- **Seleção de características:** Estratégias para definir as features que mais contribuem para a tarefa de machine que será realizada;
- **Extração de características:** Combinar informações das *features* originais e gerar novas *features*.

# Conceitos fundamentais

**Conjunto de treino:** faz parte do conjunto de dados, que é usado para treinar um modelo de aprendizado de máquina.

**Conjunto de testes:** faz parte do conjunto de dados e possui a mesma estrutura do conjunto de treinamento e testa o desempenho do modelo de aprendizado de máquina.

**Característica (features):** é a variável preditora ou uma variável independente no conjunto de dados.

**Variável alvo:** é a variável dependente no conjunto de dados cujo valor é previsto pelo modelo de aprendizado de máquina.

**Overfitting:** Overfitting é a condição que leva à versespecialização do modelo. Isso ocorre no caso de um conjunto de dados complexo.

**Regularização:** esta é a técnica usada para simplificar o modelo e é um remédio para *overfitting*.

# Aplicações



# Aplicações - por temas

- Aproximação universal de função;
- Controle de processo;
- Reconhecimento / classificação de padrões;
- Agrupamento de dados (clusterização);
- Sistemas de previsão;
- Otimização de sistemas;
- Memórias associativas.

# Aplicações - por problemas

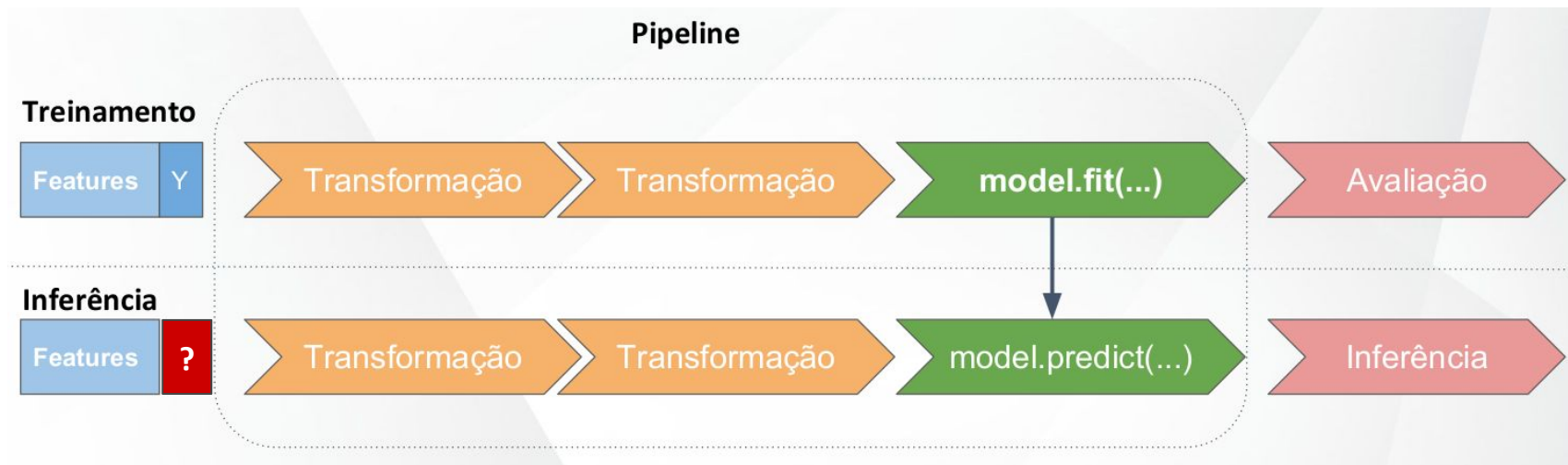
- Diagnóstico de doenças
- Reconhecimento de fala
- Mineração de dados
- Composição de músicas
- Processamento de imagens
- Previsão
- Controle de robô
- Aprovação de crédito
- Classificação
- Reconhecimento de padrões
- Estratégias de planejamento de jogos
- Compactação etc.

# Ferramentas para esta disciplina





# Pipeline de Aprendizado de Máquina



Obrigado.