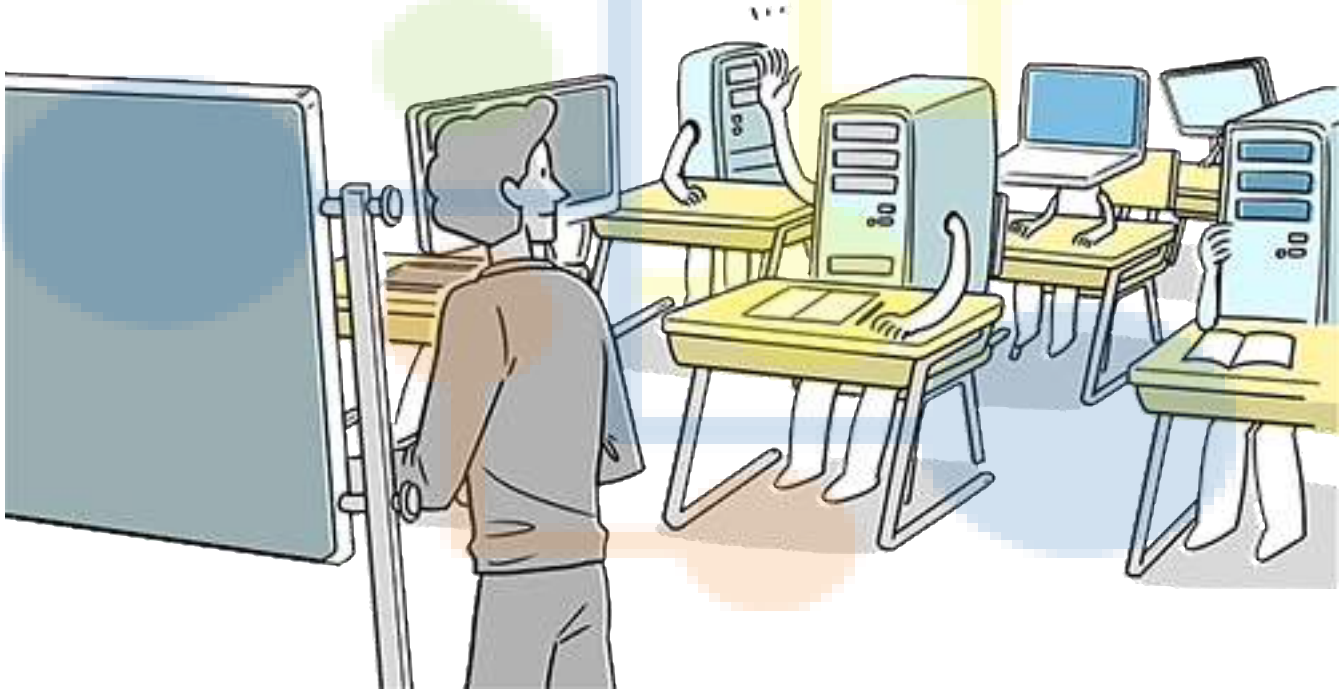
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Big Data Analytics com R e Microsoft Azure Machine Learning

Machine Learning em Linguagem R

Seja Bem-Vindo(a)!

Machine Learning em Linguagem R



Machine Learning em Linguagem R

O que veremos neste capítulo:

- Definição de Machine Learning
- Frameworks de Machine Learning
- Processo de Aprendizagem
- Treinamento, Validação e Teste
- Modelos Preditivos
- Algoritmos de Machine Learning
- Regressão e Classificação Através de Projetos
- Lista de Exercícios com a Construção de Modelos

Machine Learning em Linguagem R

Projetos Inteiros de Regressão e Classificação

Prevendo despesas hospitalares
Prevendo a ocorrência de câncer

Machine Learning em Linguagem R

Embora tenhamos aqui uma grande quantidade de conteúdo, Machine Learning ainda será estudado em muito mais detalhes nos demais cursos da Formação Cientista de Dados.

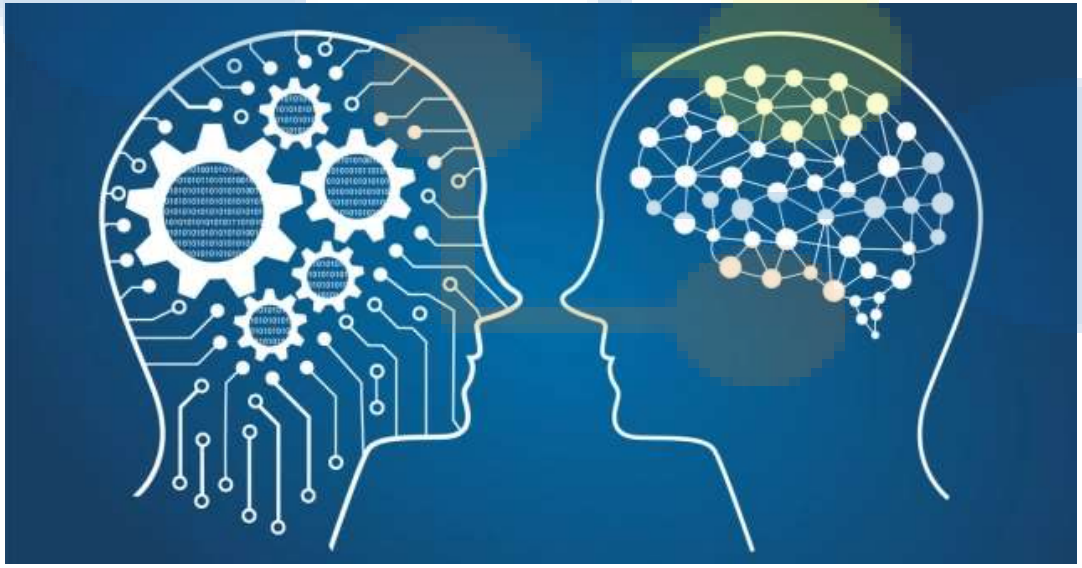
Big Data Analytics com R e Microsoft Azure Machine Learning

Introdução ao Aprendizado de Máquina (Machine Learning)

Seja Bem-Vindo(a)!

Introdução ao Aprendizado de Máquina (Machine Learning)

O termo Machine Learning (ou aprendizado de máquina em português) possui atualmente as mais variadas definições, especialmente depois de tantos filmes sobre robôs e Inteligência Artificial, que transformaram Machine Learning em algo que realmente não é.



Introdução ao Aprendizado de Máquina (Machine Learning)

O que é Machine Learning?
(Aprendizado de Máquina)

Introdução ao Aprendizado de Máquina (Machine Learning)

Machine Learning é o método de análise de dados que automatiza a construção de modelos analíticos.

Introdução ao Aprendizado de Máquina (Machine Learning)

E como as máquinas aprendem?

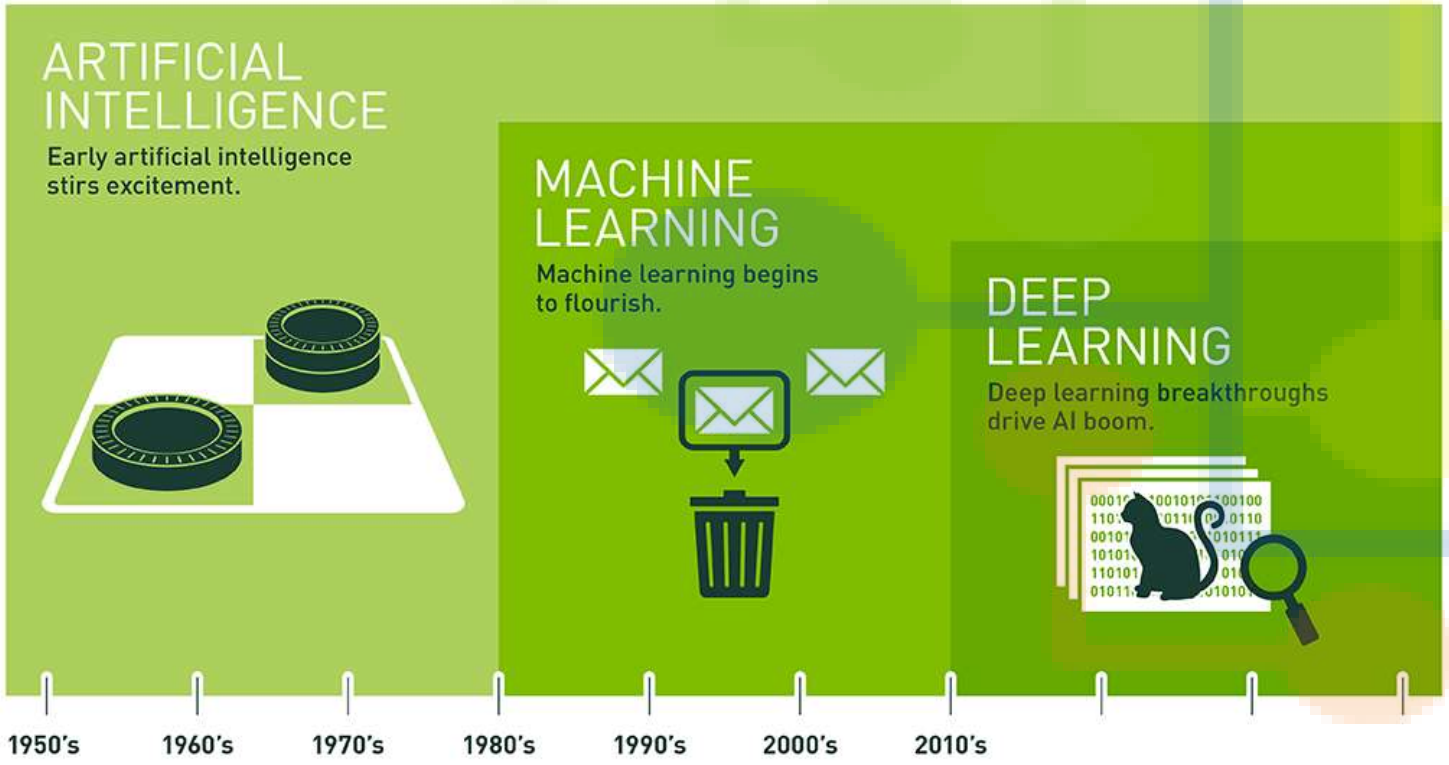
Introdução ao Aprendizado de Máquina (Machine Learning)

Machine Learning pode realizar análises preditivas mais rápido que qualquer humano seria capaz de fazer!

Introdução ao Aprendizado de Máquina (Machine Learning)

Então Machine Learning e IA são conceitos diferentes?

Introdução ao Aprendizado de Máquina (Machine Learning)



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Machine Learning é um subconjunto da Inteligência Artificial

Introdução ao Aprendizado de Máquina (Machine Learning)

Inteligência Artificial inclui Machine Learning, mas Machine Learning por si só não define Inteligência Artificial.

Introdução ao Aprendizado de Máquina (Machine Learning)

Inteligência Artificial é baseada em Machine Learning e Machine Learning é essencialmente diferente de Estatística.

Introdução ao Aprendizado de Máquina (Machine Learning)

Técnica	Estatística	Machine Learning
Entrada de Dados	Os parâmetros interpretam fenômenos da vida real e trabalham a magnitude.	Os dados são randomizados e transformados para aumentar a acurácia de análises preditivas.
Tratamento de Dados	Modelos são usados para previsões em amostras pequenas.	Trabalha com Big Data na forma de redes e grafos. Os dados são divididos em dados de treino e dados de teste.
Resultado	Captura a variabilidade e a incerteza dos parâmetros.	Probabilidade é usada para comparações e para buscar as melhores decisões.
Distribuicão dos Dados	Assumimos uma distribuição bem definida dos dados.	A distribuição dos dados é desconhecida ou ignorada antes do processo de aprendizagem.
Objetivos	Assumimos um determinado resultado e então tentamos prová-lo.	Os algoritmos aprendem a partir dos dados.

Introdução ao Aprendizado de Máquina (Machine Learning)

Machine Learning se baseia em alguns importantes conceitos da Matemática, Estatística e Ciência da Computação:

- Manipulação de Matrizes
- Teoria da Probabilidade e Inferência Estatística
- Programação
- Armazenamento e Processamento de Dados

O Que São Algoritmos?

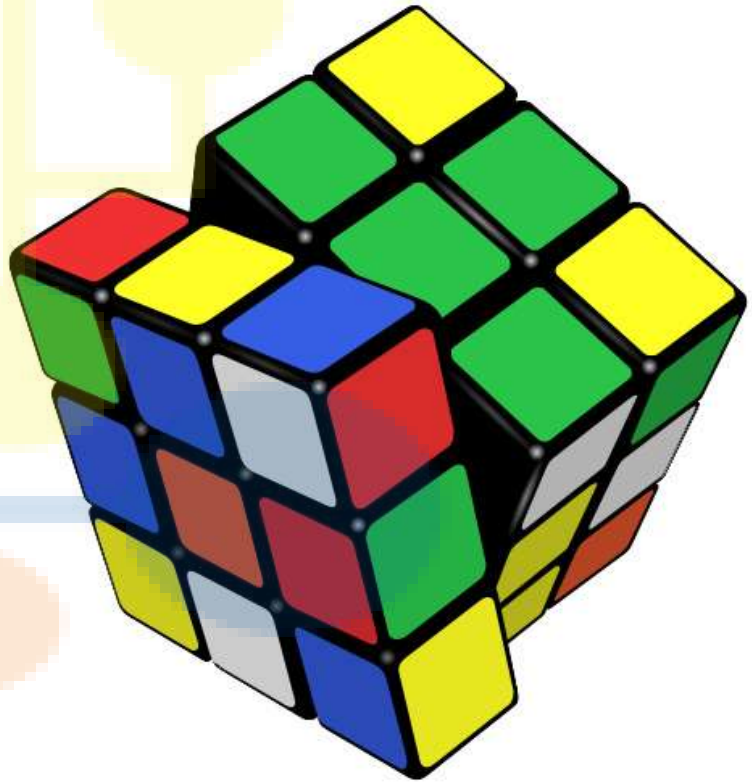
Seja Bem-Vindo(a)!

O Que São Algoritmos?

Machine Learning usa algoritmos para analisar grandes conjuntos de dados!

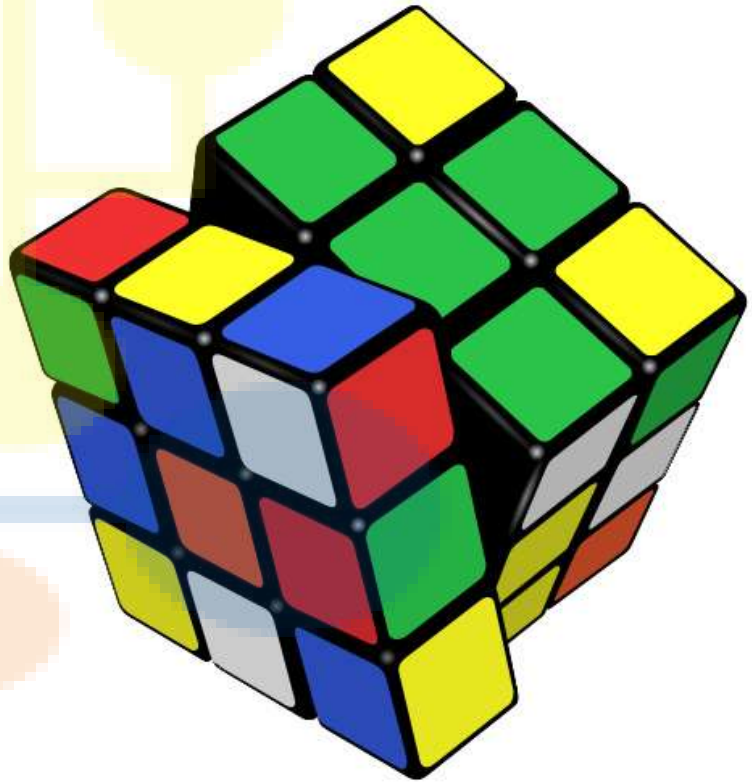
O Que São Algoritmos?

Ok entendi, mas o que são algoritmos?



O Que São Algoritmos?

Algoritmos são procedimentos ou fórmulas usados para resolver problemas.



O Que São Algoritmos?

Algoritmos são procedimentos ou fórmulas usados para resolver problemas.

O Que São Algoritmos?

O tipo de problema a ser resolvido, determina o tipo de algoritmo a ser utilizado.

O Que São Algoritmos?

Algoritmos - Exemplo

Algoritmo: Sacar dinheiro

INÍCIO

1. **Ir até o caixa eletrônico.**
2. **Colocar o cartão.**
3. **Digitar a senha.**
4. **Solicitar o saldo.**
5. **Se o saldo for maior ou igual à quantia desejada, sacar a quantia desejada; caso contrário sacar o valor do saldo.**
6. **Retirar dinheiro e cartão.**

FIM.

5

O Que São Algoritmos?

Falhas são mais comuns que sucesso em processos de Machine Learning.

Machine Learning Frameworks

Seja Bem-Vindo(a)!

Machine Learning Frameworks

Para criar modelos de Machine Learning você tem duas opções:

Machine Learning Frameworks

- Um framework é um conjunto de softwares que produzem um resultado específico. Um framework nos permite focar mais no problema de negócio e menos na parte de codificação.

- Frameworks de Machine Learning permitem que você trabalhe em um problema, sem ter que saber muito sobre programação (embora seja altamente recomendável que você conheça bem sobre programação).

- O framework cuida da gestão de infraestrutura, enquanto você pode focar mais na parte inteligente da sua aplicação.

Machine Learning Frameworks

E por que usar Machine Learning Frameworks?

Machine Learning Frameworks

Principais Machine Learning Frameworks

Machine Learning Frameworks

Linguagem R
(Pacote caret)

Machine Learning Frameworks

Microsoft Azure Machine Learning

Machine Learning Frameworks

Scikit-Learn
(Linguagem Python)

Machine Learning Frameworks

Apache Spark MLlib

Machine Learning Frameworks

Google Tensor Flow

TensorFlow

Machine Learning Frameworks

Keras

Caffe

CNTK

Mxnet

Machine Learning Frameworks

Tipos de Aprendizagem em Machine Learning

Seja Bem-Vindo(a)!

Tipos de Aprendizagem em Machine Learning

O Processo de Aprendizagem ocorre de diferentes formas e podemos dividir os algoritmos de Machine Learning em 3 grupos principais:

Tipos de Aprendizagem em Machine Learning

Aprendizagem Supervisionada

Aprendizagem Não Supervisionada

Aprendizagem Por Reforço

Tipos de Aprendizagem em Machine Learning

Aprendizagem Supervisionada

É o termo usado sempre que o algoritmo é "treinado" sobre um conjunto de dados históricos contendo entradas e saídas.

Baseado no treinamento com os dados históricos, o modelo pode tomar decisões precisas quando recebe novos dados.

Tipos de Aprendizagem em Machine Learning

Aprendizagem Não Supervisionada

A aprendizagem não supervisionada ocorre quando um algoritmo aprende com exemplos simples, sem qualquer resposta associada, deixando a cargo do algoritmo determinar os padrões de dados por conta própria.

Tipos de Aprendizagem em Machine Learning

Aprendizagem Por Reforço

O conceito de Aprendizagem Por Reforço (Reinforcement Learning) é como aprender por tentativa e erro. Os erros ajudam a aprender, porque eles têm uma grande penalidade associada a eles (custo, perda de tempo e assim por diante), ensinando que um determinado curso de ação tem menor probabilidade de êxito do que outros.

Tipos de Aprendizagem em Machine Learning

Aprendizagem Supervisionada

A aprendizagem supervisionada ocorre quando um algoritmo aprende a partir de dados históricos de exemplo, com entradas (inputs) e possíveis saídas (outputs), que podem consistir em valores quantitativos ou qualitativos, a fim de prever a resposta correta quando recebe novos dados.

Tipos de Aprendizagem em Machine Learning

O Processo de Aprendizagem em Machine Learning

Seja Bem-Vindo(a)!

O Processo de Aprendizagem em Machine Learning

O Processo de Aprendizagem

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

O Processo de Aprendizagem em Machine Learning

Processo de Aprendizagem

O Processo de Aprendizagem em Machine Learning

Um algoritmo de ML, como um algoritmo de classificação por exemplo, funciona da mesma forma. Ele constrói suas capacidades cognitivas através da criação de uma formulação matemática que inclui todas as características dadas sobre um determinado fenômeno.

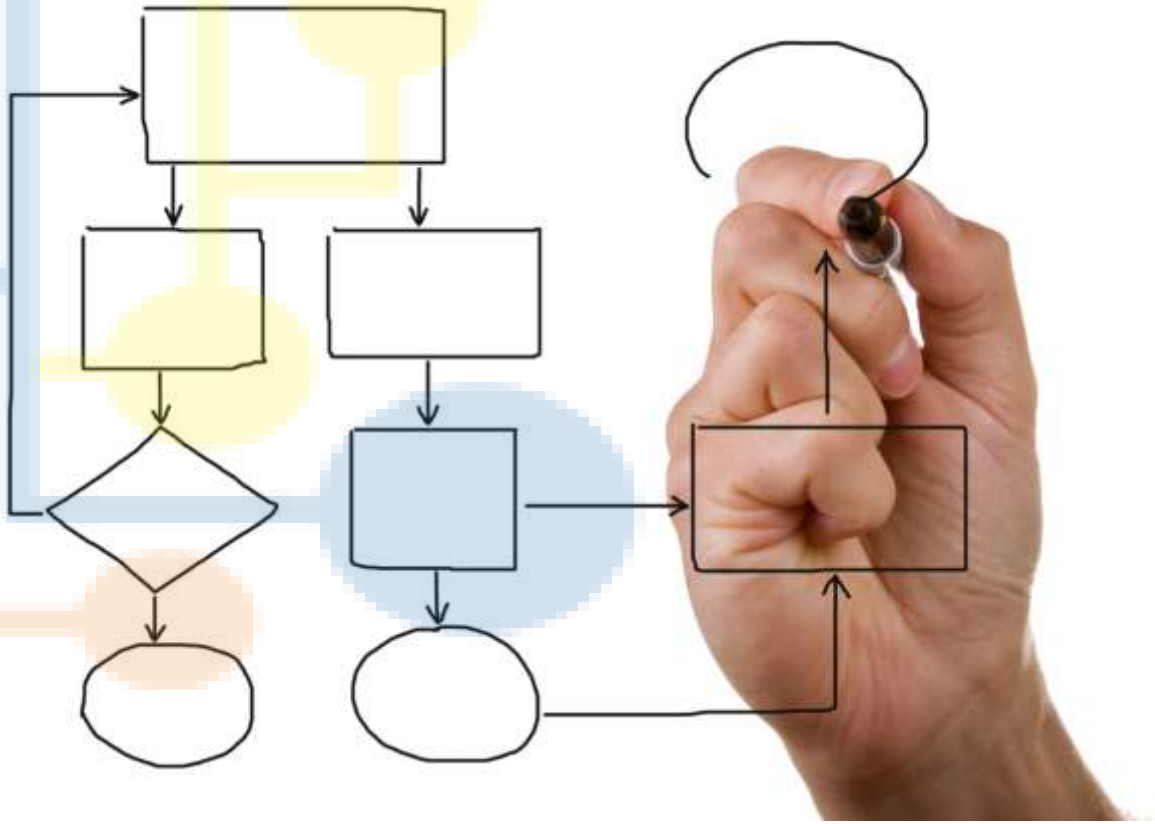
O Processo de Aprendizagem em Machine Learning

O Processo de Aprendizagem ocorre de diferentes formas e podemos dividir os algoritmos de Machine Learning em 3 grupos principais:

Aprendizagem Supervisionada, Aprendizagem Não Supervisionada e Aprendizagem Por Reforço.

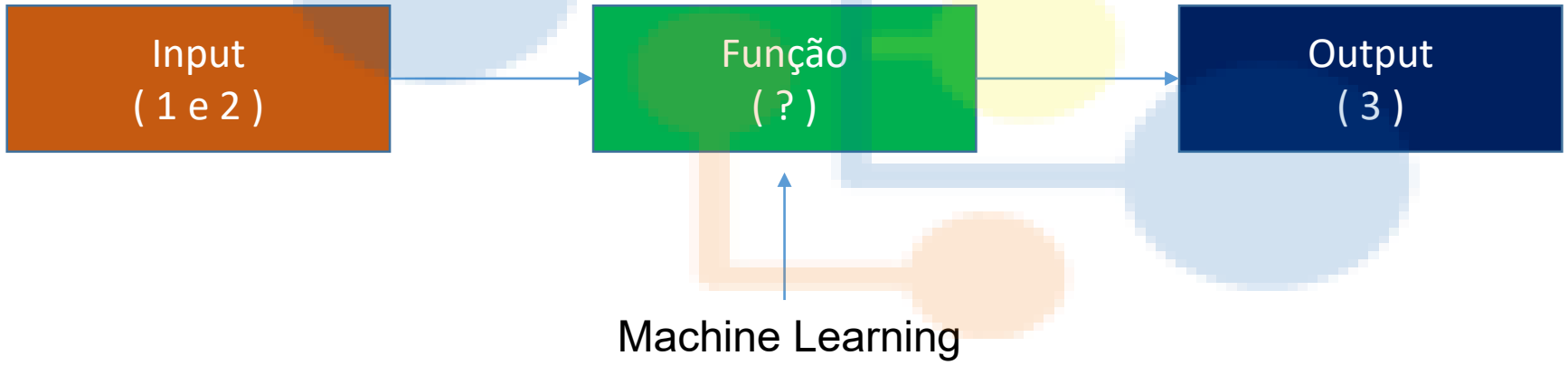
O Processo de Aprendizagem em Machine Learning

Do ponto de vista matemático, você pode expressar o processo de representação no aprendizado de máquina utilizando o mapeamento equivalente.



O Processo de Aprendizagem em Machine Learning

Processo de Aprendizagem



O Processo de Aprendizagem em Detalhes

Seja Bem-Vindo(a)!

O Processo de Aprendizagem em Detalhes

Um componente chave do processo de aprendizagem é a generalização!

O Processo de Aprendizagem em Detalhes

E para poder generalizar a função que melhor resolve o problema, os algoritmos de Machine Learning se baseiam em 3 componentes:

O Processo de Aprendizagem em Detalhes

O Processo de Aprendizagem em Detalhes

Os algoritmos de aprendizagem possuem diversos parâmetros internos (valores separados em vetores e matrizes).

O Processo de Aprendizagem em Detalhes

Esses parâmetros funcionam como uma espécide de memória para o algoritmo, permitindo que o mapeamento ocorra e as características analisadas sejam conectadas.

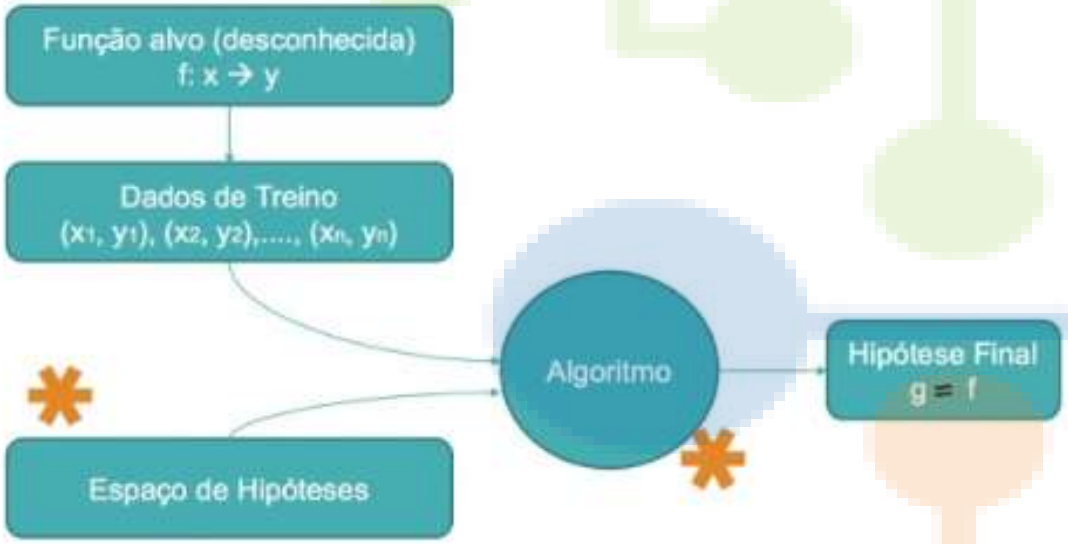
O Processo de Aprendizagem em Detalhes

As dimensões e tipos de parâmetros internos delimitam o tipo de funções-alvo que um algoritmo pode aprender. O engine de otimização no algoritmo muda os valores iniciais dos parâmetros durante a aprendizagem para representar função-alvo.

O Processo de Aprendizagem em Detalhes

Otimização

O Processo de Aprendizagem em Detalhes



- Espaço de Hipóteses

$$\mathcal{H} = \{h\} \qquad g \in \mathcal{H}$$

- Algoritmo de Aprendizagem

Espaço de Hipóteses + **Algoritmo de Aprendizagem** = **Modelo de Aprendizagem**

Redes Neurais | Back Propagation

O Processo de Aprendizagem em Detalhes

Falso Positivo

O Processo de Aprendizagem em Detalhes

Big Data é uma grande mistura de dados. Um bom algoritmo de Machine Learning deve ser capaz de distinguir os sinais e mapear as funções alvo de forma eficiente.

O Processo de Aprendizagem em Detalhes

Cost Function

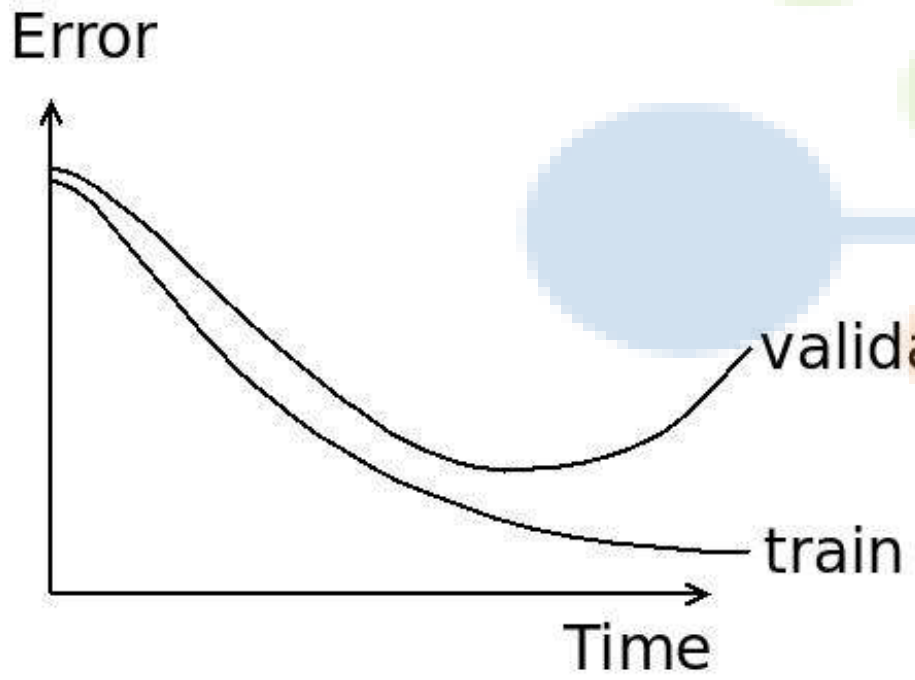
Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters: θ_0, θ_1

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

O Processo de Aprendizagem em Detalhes



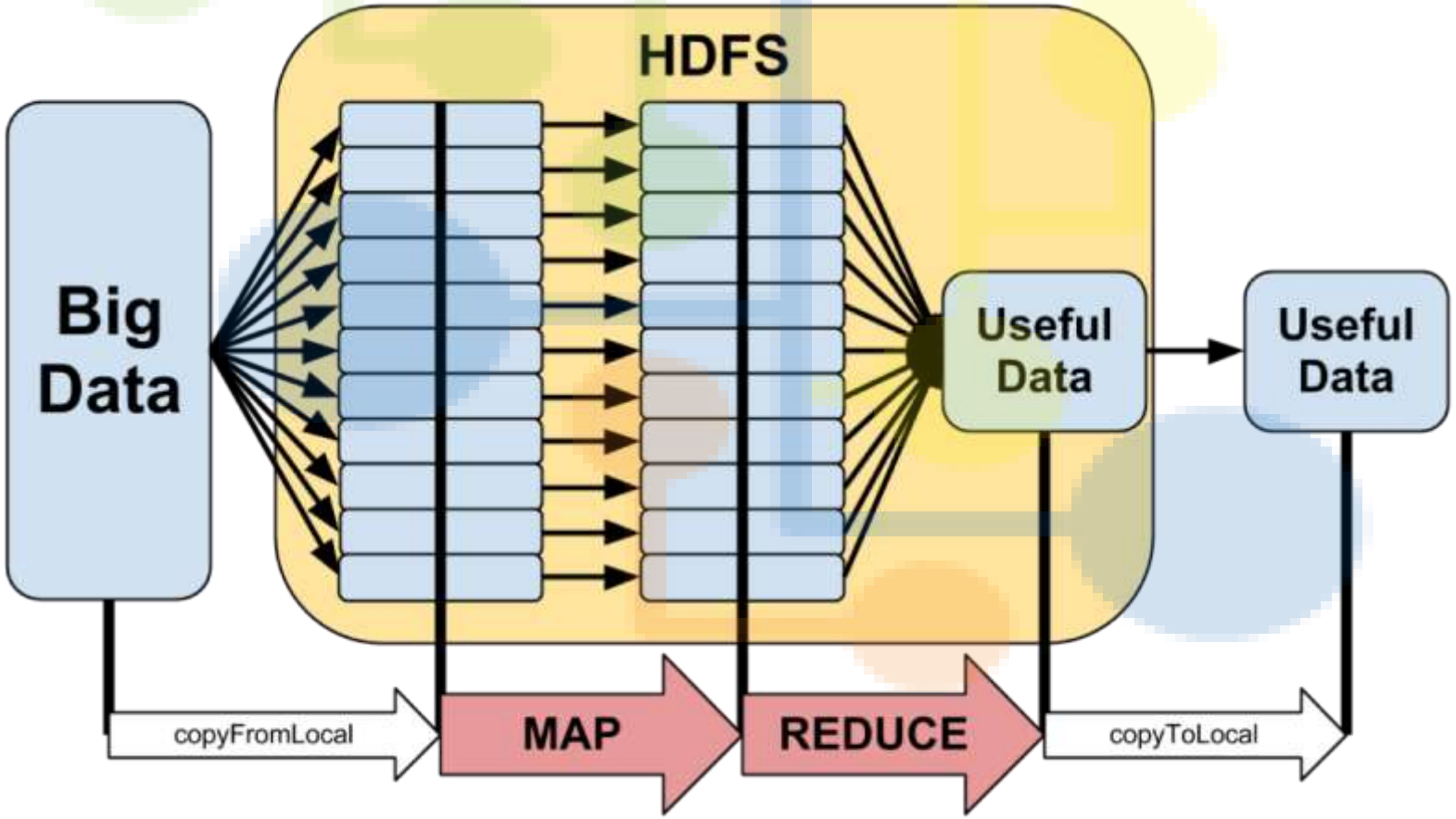
Definindo o Erro

Cost Function → Nível de erro

O Processo de Aprendizagem em Detalhes

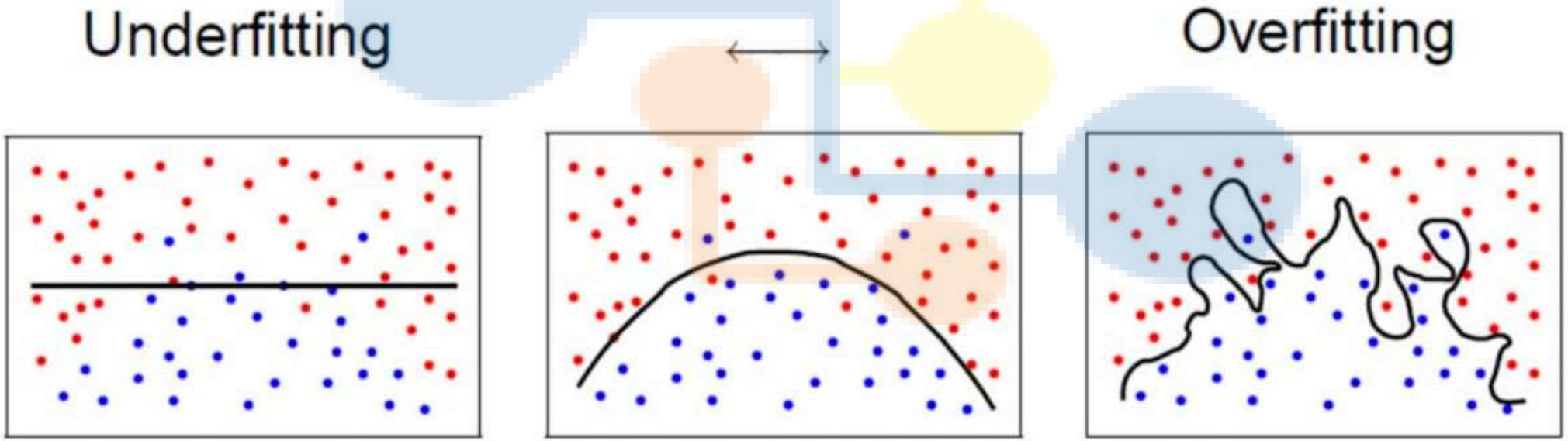
As técnicas de aprendizagem de máquina baseadas em algoritmos estatísticos utilizam Cálculo e Álgebra Linear e os dados precisam estar carregados em memória.

O Processo de Aprendizagem em Detalhes



O Processo de Aprendizagem em Detalhes

O modelo pode aprender demais (overfitting) ou aprender de menos (underfitting).



O Processo de Aprendizagem em Detalhes

Para atingir o equilíbrio e criar grandes soluções de Machine Learning, você terá que fazer escolhas.

Simplicidade

Complexidade

Ponto Ideal

O Processo de Aprendizagem em Detalhes

Para visualizar se os seus algoritmos de Machine Learning estão sofrendo algum tipo de força tendenciosa, você pode usar um gráfico chamado *Curva de Aprendizagem.*

O Processo de Aprendizagem em Detalhes

Para usar uma curva de aprendizagem, você precisa:

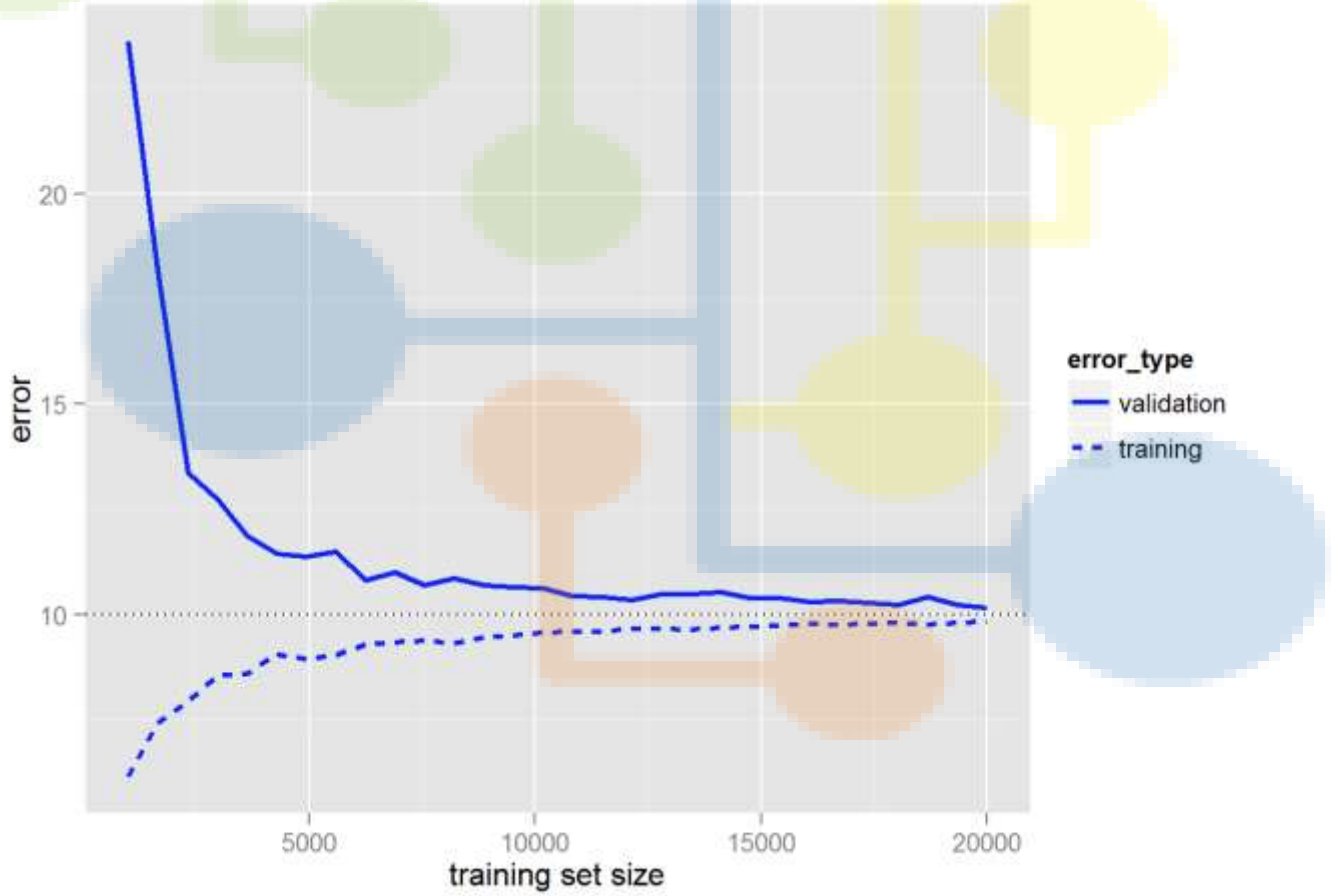
1- Dividir seus dados em amostras, chamadas dados de treino e dados de teste (uma divisão 70/30 funciona bem). Dados de validação podem ser usados durante o treinamento.

2- Criar porções dos seus dados de treino, com tamanhos diferentes a cada passagem de treino.

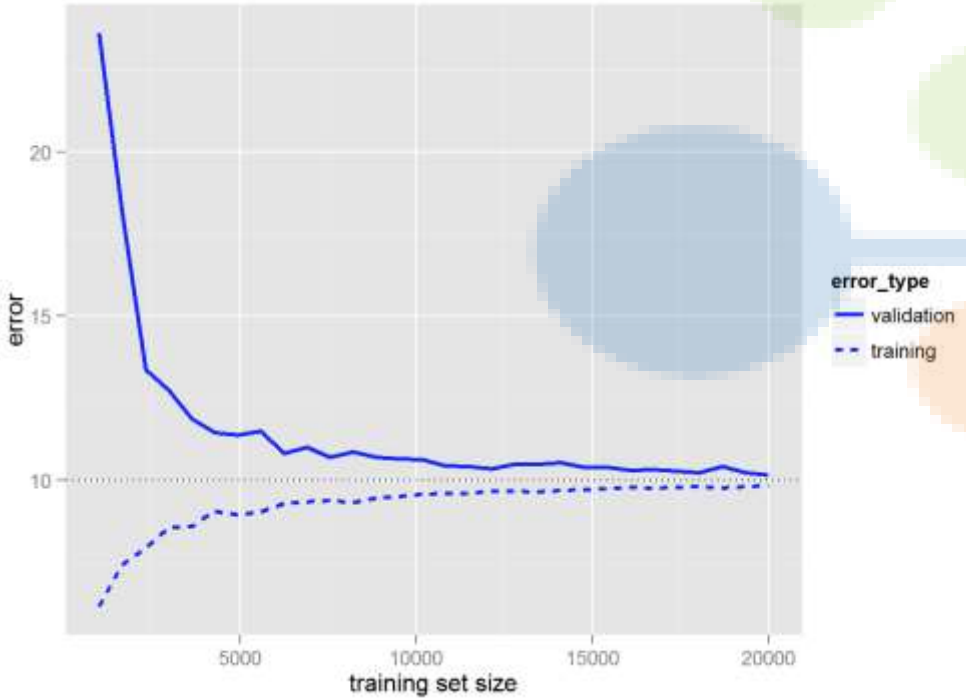
3- Treinar seus modelos com os diferentes subsets. Registrar a performance.

4- Gerar um gráfico com os resultados. Atenção aos intervalos de confiança e ao desvio padrão.

O Processo de Aprendizagem em Detalhes



O Processo de Aprendizagem em Detalhes



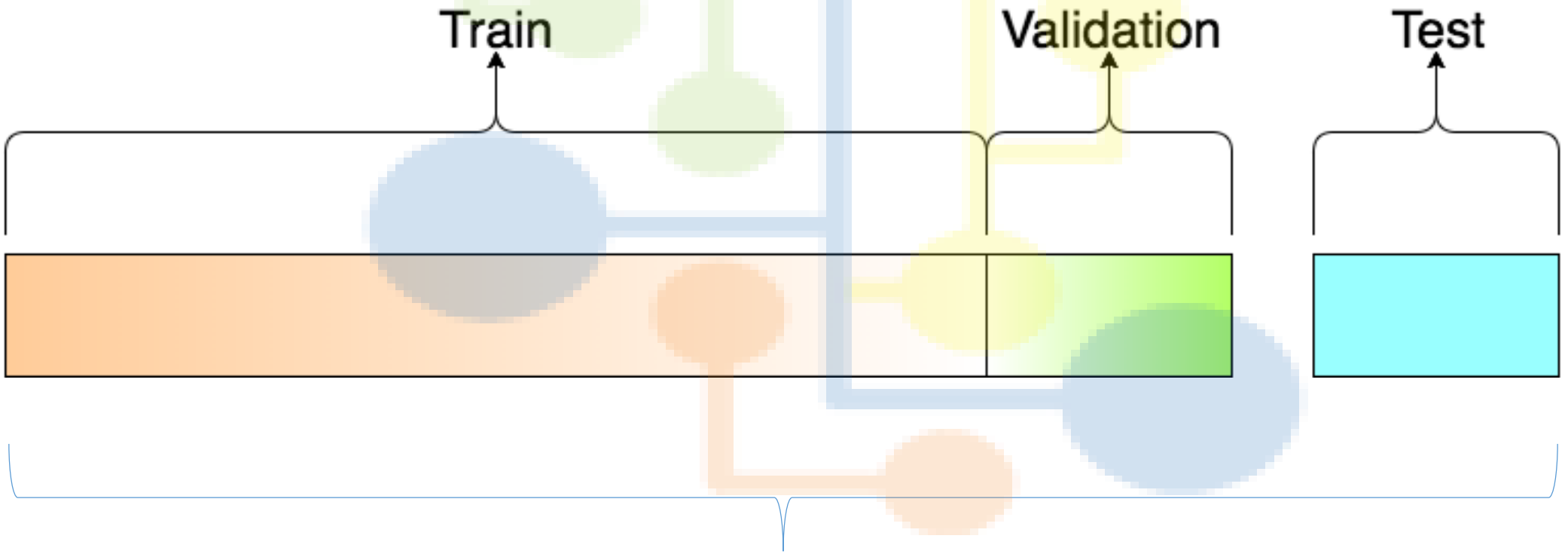
Podemos criar curvas de aprendizagem em R de diversas formas, usando os pacotes mlr, caret ou mesmo o ggplot2.

Big Data Analytics com R e Microsoft Azure Machine Learning

Treinamento, Validação e Teste

Seja Bem-Vindo(a)!

Treinamento, Validação e Teste



Treinamento, Validação e Teste

Treinamento, Validação e Teste

75 a 70% - dados de treino
25 a 30% - dados de teste

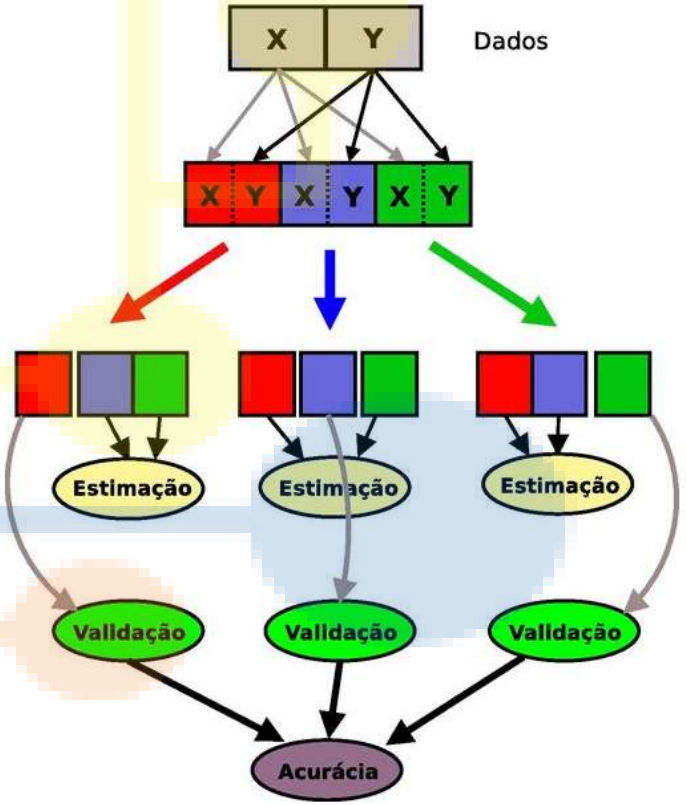
Treinamento, Validação e Teste

Treinamento, Validação e Teste

70% - dados de treino
20% - dados de validação
10% - dados teste

Treinamento, Validação e Teste

Treinamento, Validação e Teste



Treinamento, Validação e Teste

Treinamento, Validação e Teste

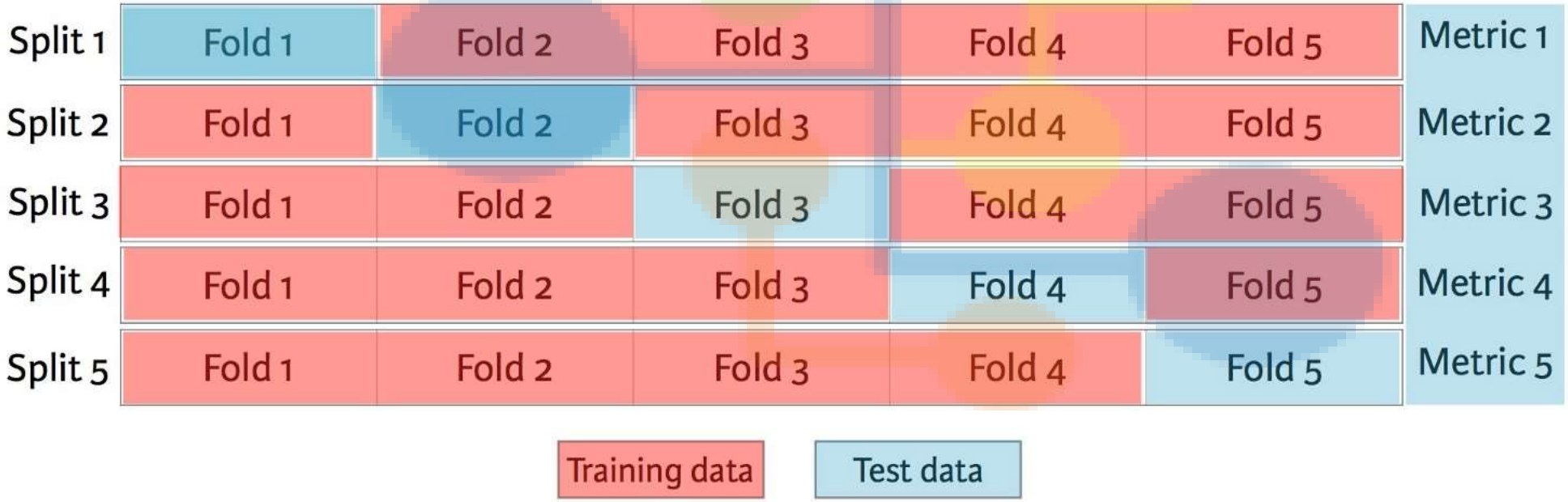
n > 10.000

Treinamento, Validação e Teste

Cross-Validation

Treinamento, Validação e Teste

Cross-Validation



Treinamento, Validação e Teste

Cross-Validation

O conceito central das técnicas de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, e posteriormente, utiliza-se alguns destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) e o restante dos subconjuntos (dados de validação ou de teste) são empregados na validação do modelo.

O Que é um Modelo de Machine Learning?

Seja Bem-Vindo(a)!

O Que é um Modelo de Machine Learning?

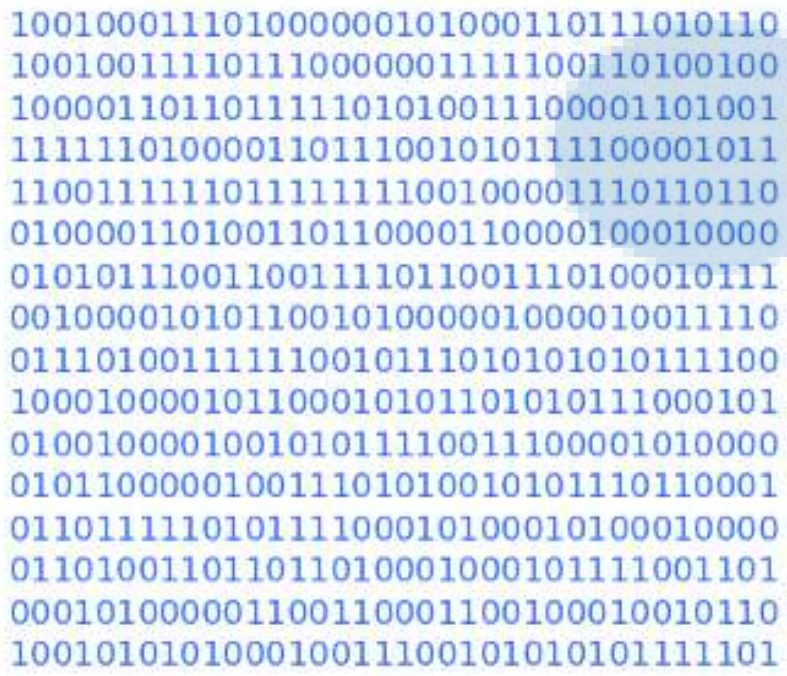
Como já vimos, a aprendizagem de máquina é um subcampo da Inteligência Artificial que evoluiu a partir do estudo de reconhecimento de padrões e teoria da aprendizagem computacional.

O Que é um Modelo de Machine Learning?

Machine Learning é um campo de estudo que dá ao computador a capacidade de aprender, sem ser programado de forma explícita.

O Que é um Modelo de Machine Learning?

Dados



Algoritmo

Modelo

$$f(x)$$

O Que é um Modelo de Machine Learning?

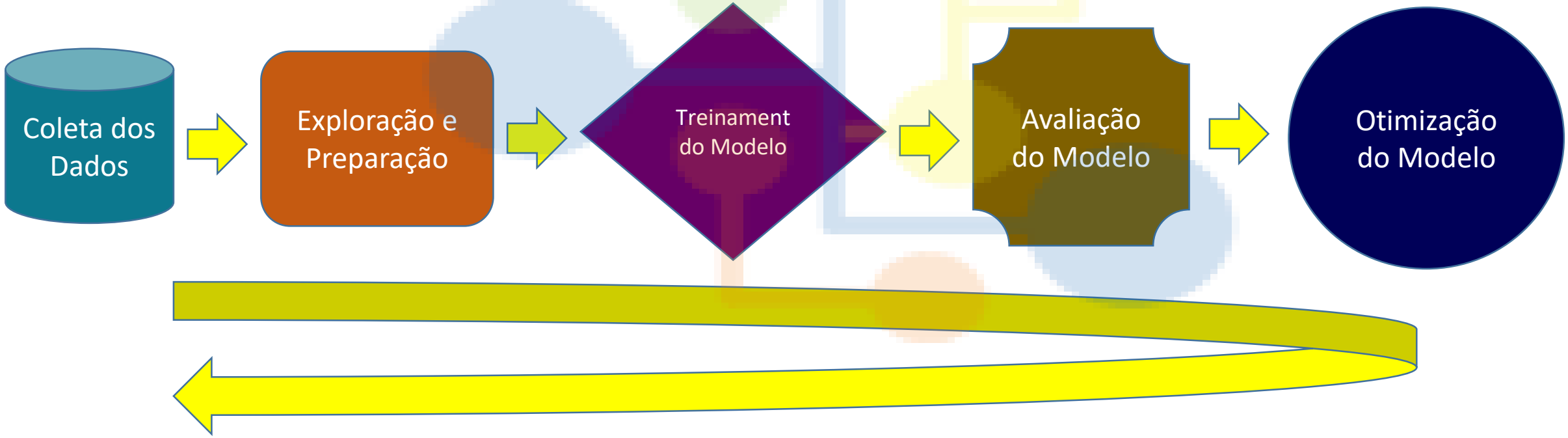
Modelo

Existem muitos tipos diferentes de modelos. Você pode já estar familiarizado com alguns. Os exemplos incluem:

- Equações matemáticas
- Diagramas relacionais
- Agrupamentos de dados, conhecidos como clusters

O Que é um Modelo de Machine Learning?

Criação do Modelo



O Que é um Modelo de Machine Learning?

Este é um trabalho iterativo e assim como um surfista está sempre em busca da onda perfeita, seu trabalho como Cientista de Dados é buscar sempre o melhor modelo possível para suas previsões.

O Que é um Modelo de Machine Learning?

Machine Learning na Prática

O Que é um Modelo de Machine Learning?

Lembre-se: um modelo de Machine Learning será usado para resolver um problema específico!

O Que é um Modelo de Machine Learning?

Não caia na tentação de querer aplicar seu modelo a tudo que você vê pela frente.

Cada problema de negócio, cada conjunto de dados, pode requerer um modelo diferente.

Big Data Analytics com R e Microsoft Azure Machine Learning

Algoritmos de Machine Learning

Seja Bem-Vindo(a)!

Algoritmos de Machine Learning

Aprendizagem Supervisionada

- Classificação
- Regressão

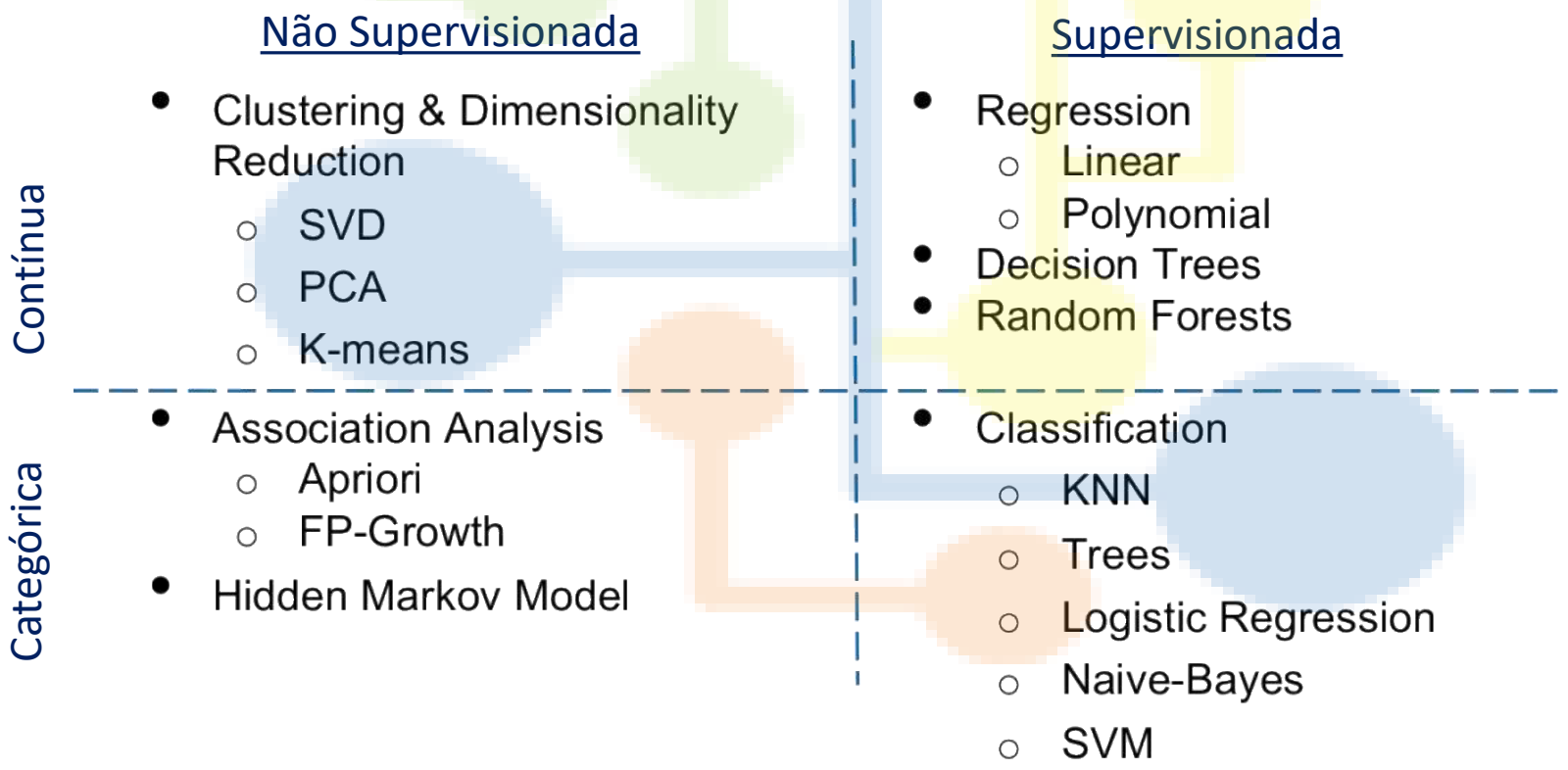
Aprendizagem Não Supervisionada

- Clustering
- Segmentação
- Redução de Dimensionalidade

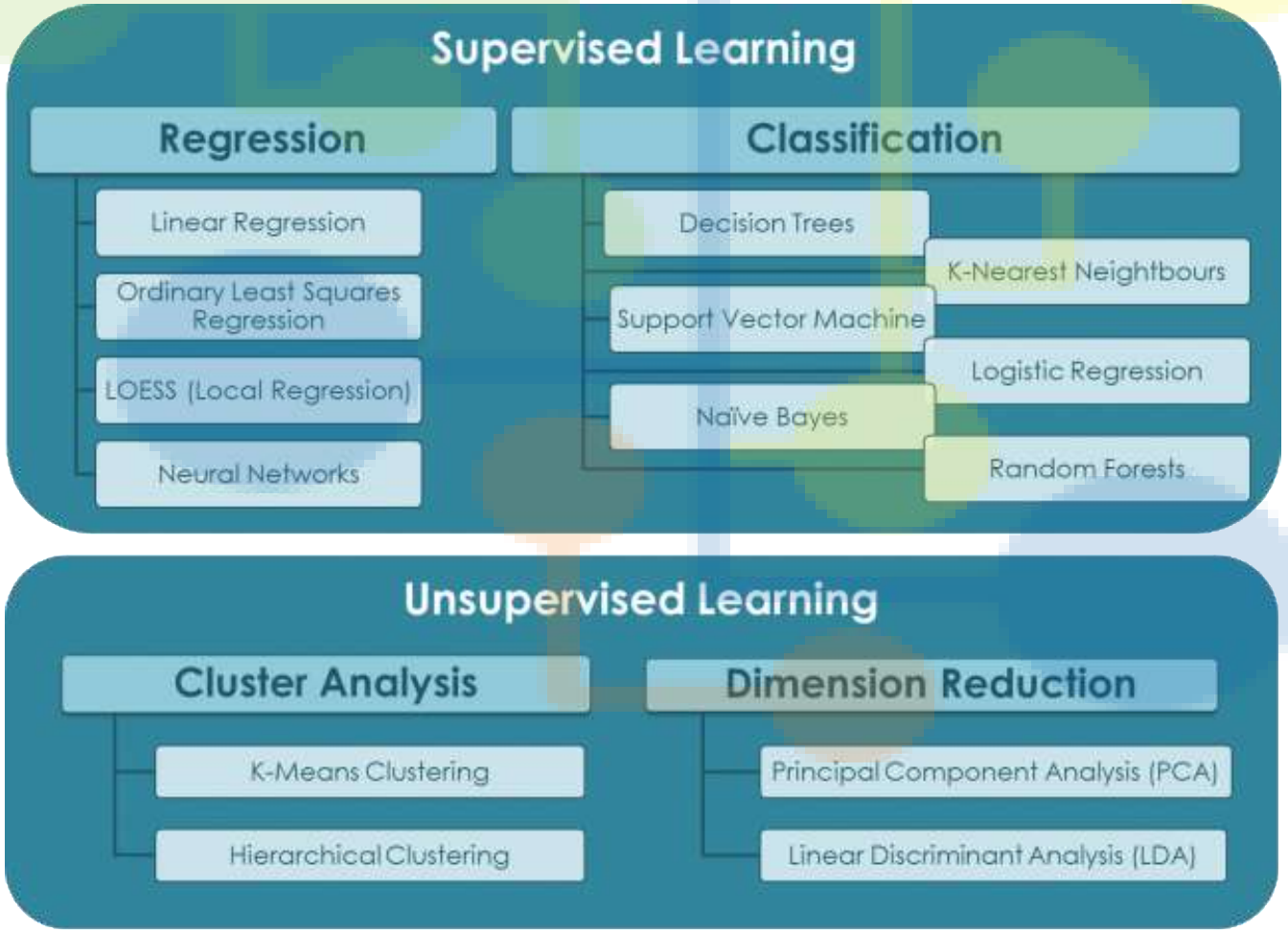
Aprendizagem por Reforço

- Sistemas de Recomendação
- Sistemas de Recompensa
- Processo de Decisão

Algoritmos de Machine Learning



Algoritmos de Machine Learning



Algoritmos de Machine Learning

Algoritmos de Machine Learning

Há tantas algoritmos disponíveis com tantos métodos diferentes, que somente o processo de escolha de qual deve ser usado, já vai consumir bastante do seu tempo como Cientista de Dados.

Algoritmos de Machine Learning

Podemos categorizar os algoritmos de Machine Learning em 2 grupos principais:

- Estilo de Aprendizagem
- Similaridade (Funcionamento)

Algoritmos de Machine Learning

Algoritmos de Machine Learning

Algoritmos de Regressão

- Ordinary Least Squares Regression (OLSR)
- Linear Regression
- Logistic Regression
- Stepwise Regression
- Multivariate Adaptive Regression Splines (MARS)
- Locally Estimated Scatterplot Smoothing (LOESS)

Algoritmos de Machine Learning

Algoritmos Regulatórios

- Ridge Regression
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Elastic Net
- Least-Angle Regression (LARS)

Algoritmos de Machine Learning

Algoritmos Baseados em Instância (Instance-based)

- k-Nearest Neighbour (kNN)
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)
- Locally Weighted Learning (LWL)

Algoritmos de Machine Learning

Algoritmos de Árvore de Decisão

- Classification and Regression Tree (CART)
- Conditional Decision Trees
- Iterative Dichotomiser 3 (ID3)
- C4.5 and C5.0 (different versions of a powerful approach)
- Chi-squared Automatic Interaction Detection (CHAID)
- Decision Stump
- M5

Algoritmos de Machine Learning

Algoritmos Bayesianos

- Naive Bayes
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Averaged One-Dependence Estimators (AODE)
- Bayesian Belief Network (BBN)
- Bayesian Network (BN)

Algoritmos de Machine Learning

Algoritmos de Clustering

- k-Means
- k-Medians
- Expectation Maximisation (EM)
- Hierarchical Clustering

Algoritmos de Machine Learning

Algoritmos Baseados em Regras de Associação

- Apriori algorithm
- Eclat algorithm

Algoritmos de Machine Learning

Redes Neurais Artificiais

- Perceptron
- Back-Propagation
- Hopfield Network
- Radial Basis Function Network (RBFN)

Algoritmos de Machine Learning

Deep Learning

- Deep Boltzmann Machine (DBM)
- Deep Belief Networks (DBN)
- Convolutional Neural Network (CNN)
- Stacked Auto-Encoders

Algoritmos de Machine Learning

Algoritmos de Redução de Dimensionalidade

- Principal Component Analysis (PCA)
- Principal Component Regression (PCR)
- Partial Least Squares Regression (PLSR)
- Multidimensional Scaling (MDS)
- Projection Pursuit
- Linear Discriminant Analysis (LDA)
- Mixture Discriminant Analysis (MDA)
- Quadratic Discriminant Analysis (QDA)
- Flexible Discriminant Analysis (FDA)

Algoritmos de Machine Learning

Algoritmos Ensemble

- Boosting
- Bootstrapped Aggregation (Bagging)
- AdaBoost
- Stacked Generalization (blending)
- Gradient Boosting Machines (GBM)
- Gradient Boosted Regression Trees (GBRT)
- Random Forest

Algoritmos de Machine Learning

Outros Algoritmos

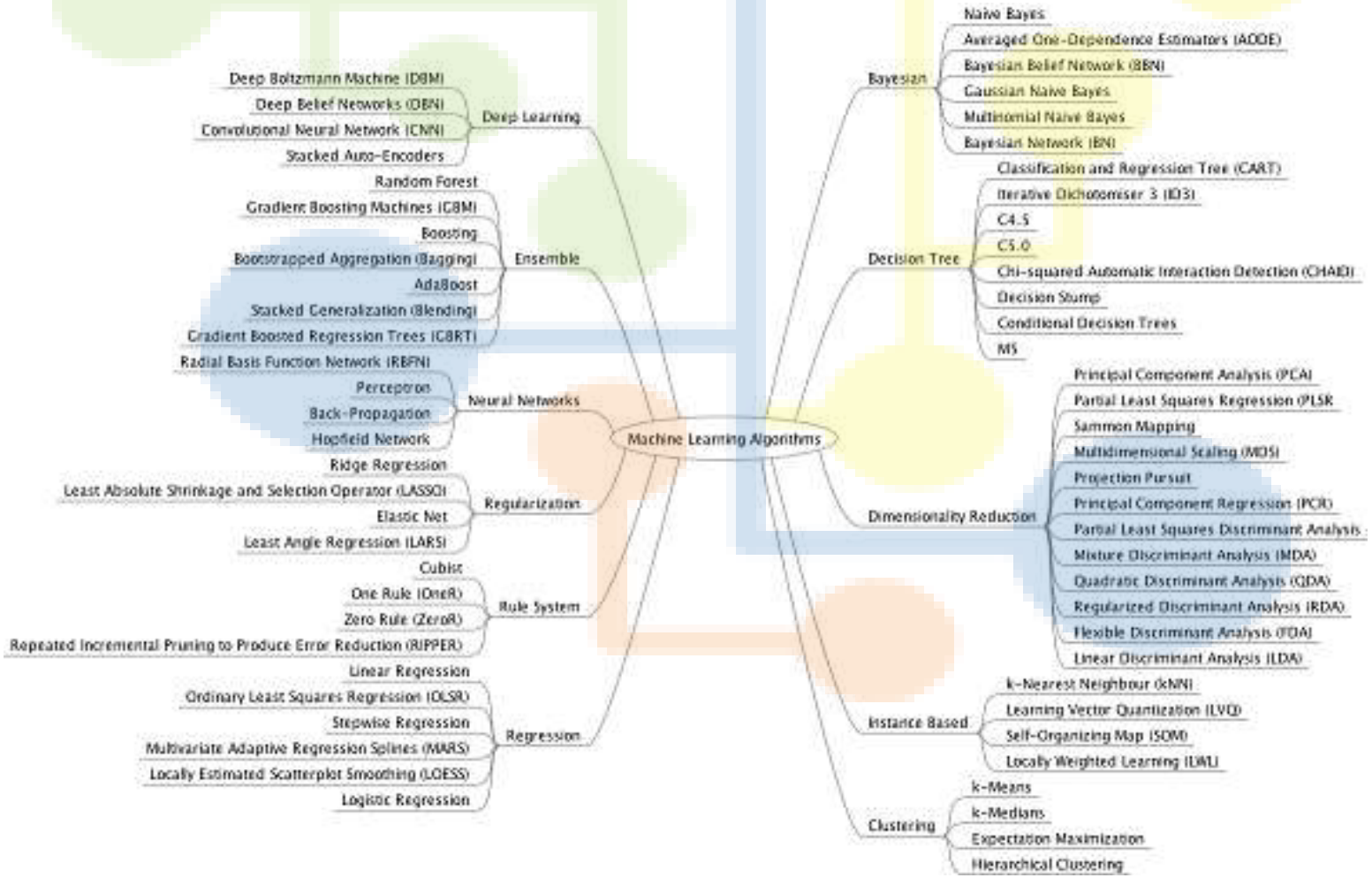
- Support Vector Machines
- Computer Vision (CV)
- Natural Language Processing (NLP)
- Recommender Systems
- Graphical Models

Algoritmos de Machine Learning

Sim, eu sei...muita coisa não??????

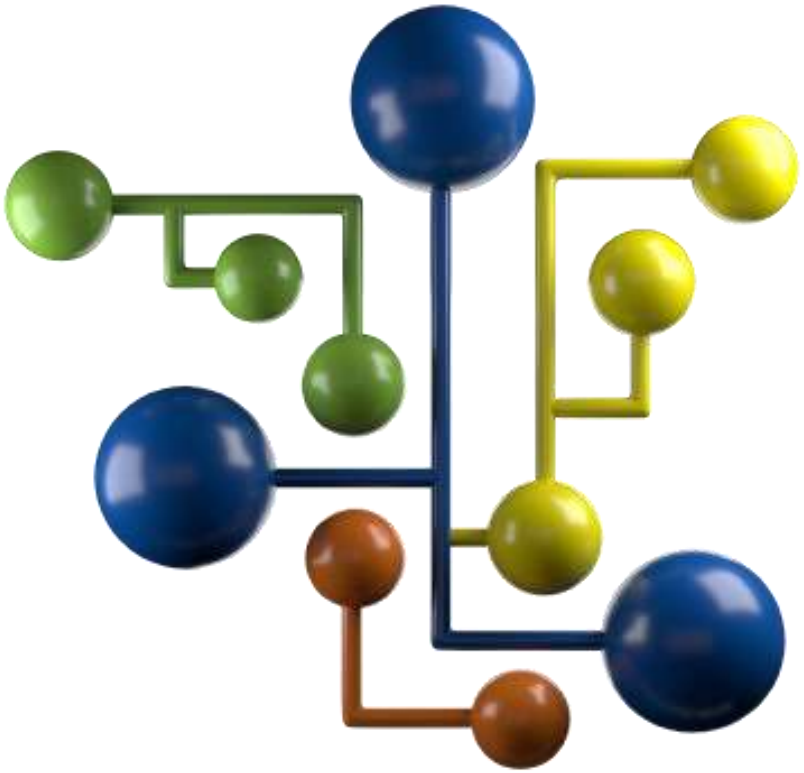
Mas espere, ainda não acabou!

Algoritmos de Machine Learning



Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Muito Obrigado por Participar!



Tenha uma Excelente Jornada de Aprendizagem.

Equipe Data Science Academy