



Data Science
Academy

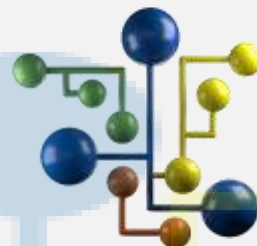
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Machine Learning



Data Science
Academy

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



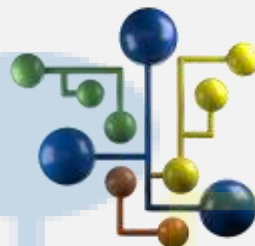
**Data Science
Academy**

Seja muito bem-vindo(a)!



Data Science
Academy

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



**Data Science
Academy**

Algoritmos de Machine Learning e Modelos Preditivos

An abstract background diagram consisting of several colored circles (blue, green, yellow, orange) connected by lines of the same color, forming a network-like structure. The circles vary in size and are distributed across the slide, with some lines forming loops and others connecting different clusters.

Dados + Análise = Valor



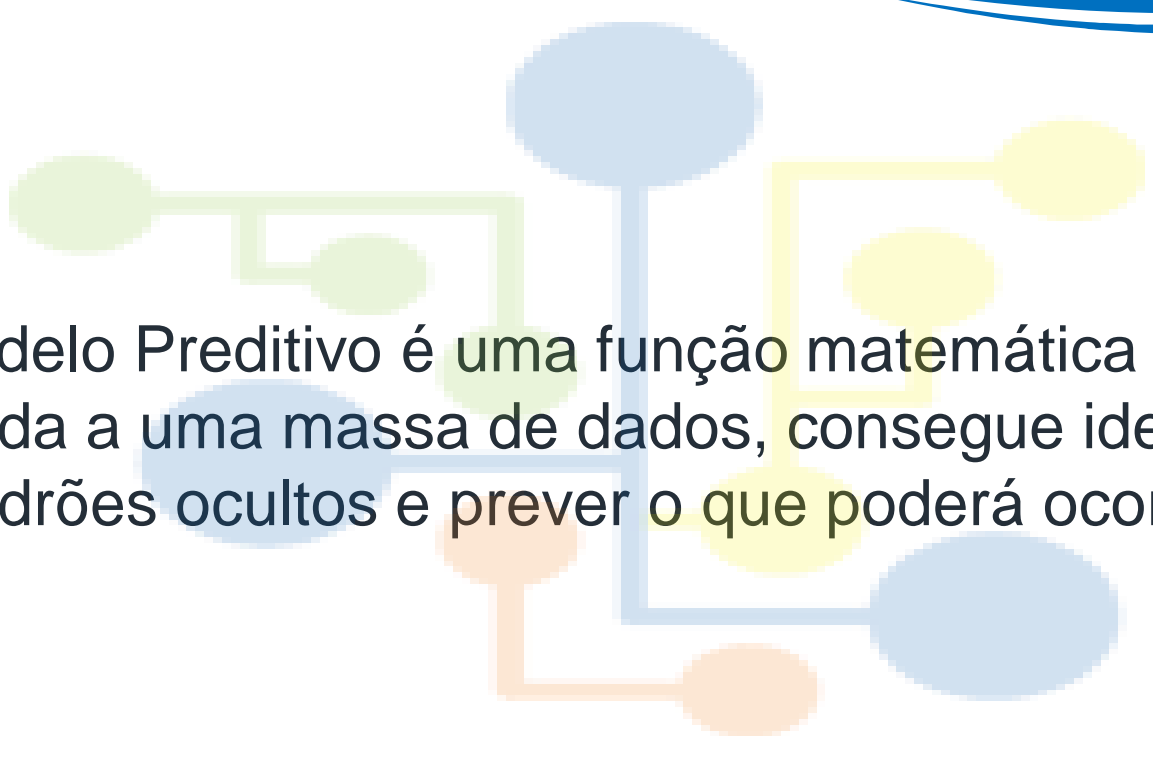
Modelos Descritivos

- Quantos clientes perdemos nos últimos 3 meses?
- As fraudes aumentaram ou diminuiram no último ano?



Modelos Preditivos

- Quantos clientes podemos conquistar nos próximos 3 meses?
- As fraudes aumentarão ou diminuirão no próximo ano?

A faint, stylized diagram of a neural network is visible in the background. It consists of several interconnected nodes (circles) in blue, green, yellow, and orange, connected by lines representing weights or connections.

Modelo Preditivo é uma função matemática que, aplicada a uma massa de dados, consegue identificar padrões ocultos e prever o que poderá ocorrer.



Aprendizagem
Supervisionada

Aprendizagem Não
Supervisionada

Métodos
Baseados em
Instância

Métodos
Probabilísticos

Métodos
Baseados em
Procura

Métodos
Baseados em
Otimização

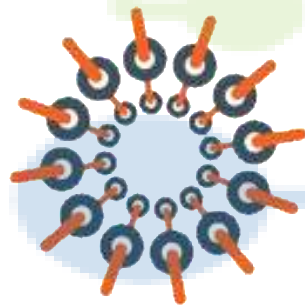


A construção de bons modelos preditivos implica o domínio de um conjunto de metodologias e conceitos sem os quais a qualidade poderá ser afetada.



O objetivo da análise preditiva é ir além das estatísticas e mostrar, através dos dados coletados, uma melhor visão do que vai acontecer no futuro. Assim é possível coletar insights que levarão a decisões melhores.



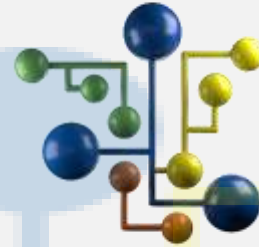


DataDriven Business



Data Science
Academy

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



**Data Science
Academy**

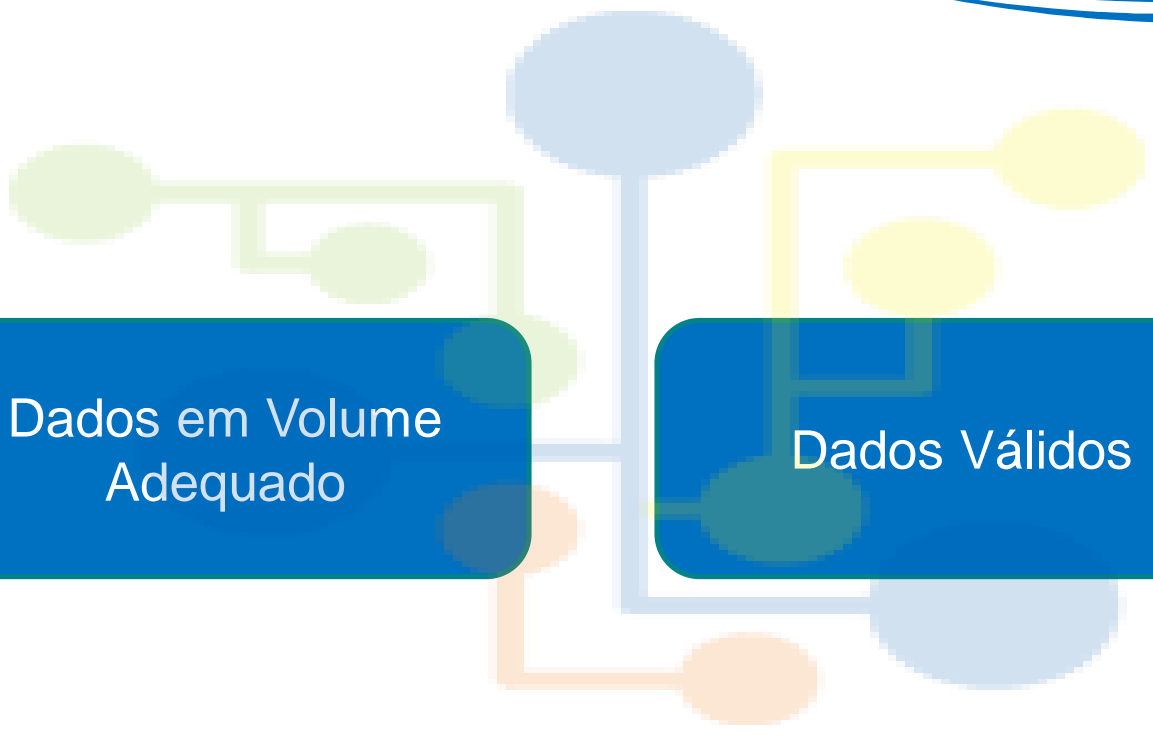
A Importância do Modelo Preditivo



Suponhamos uma operadora de telefonia móvel. Um dos principais problemas de empresas deste segmento é a taxa de perda de clientes ou churn rate.



Agregando ao modelo regras de negócio, como agrupar clientes por rentabilidade, a operadora pode fazer ofertas diferenciadas para evitar a desconexão.

An abstract background diagram consisting of several interconnected nodes and lines. The nodes are colored in shades of blue, green, yellow, and orange. The lines are colored in shades of blue, green, and yellow. The diagram is centered on the slide and partially obscured by two blue boxes.

Dados em Volume
Adequado

Dados Válidos





Criar iniciativas de Big Data Analytics, não é simplesmente adquirir tecnologias.





Identifique com a maior precisão possível o problema de negócio.

Quanto mais precisa a pergunta, mais precisa será a resposta e, portanto, maior o valor da resposta.



Mas não superestime o valor da predição.

Mesmo em uma sociedade cada vez mais data-driven, a intuição muitas vezes é necessária.

An abstract background diagram consisting of several interconnected nodes (circles) in blue, green, yellow, and orange, connected by lines of the same color. The nodes are arranged in a non-linear, branching pattern.

Tenha dados em volume e qualidade adequados.

Sem qualidade, o volume não tem valor.



E não subestime o desafio da implementação.

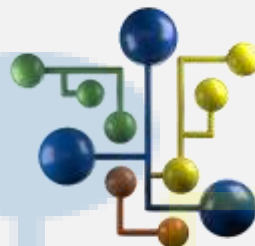
Não basta ter apenas a tecnologia, é necessário expertise (conhecimento do negócio, tecnologia, modelagem) para fazer a coisa acontecer.





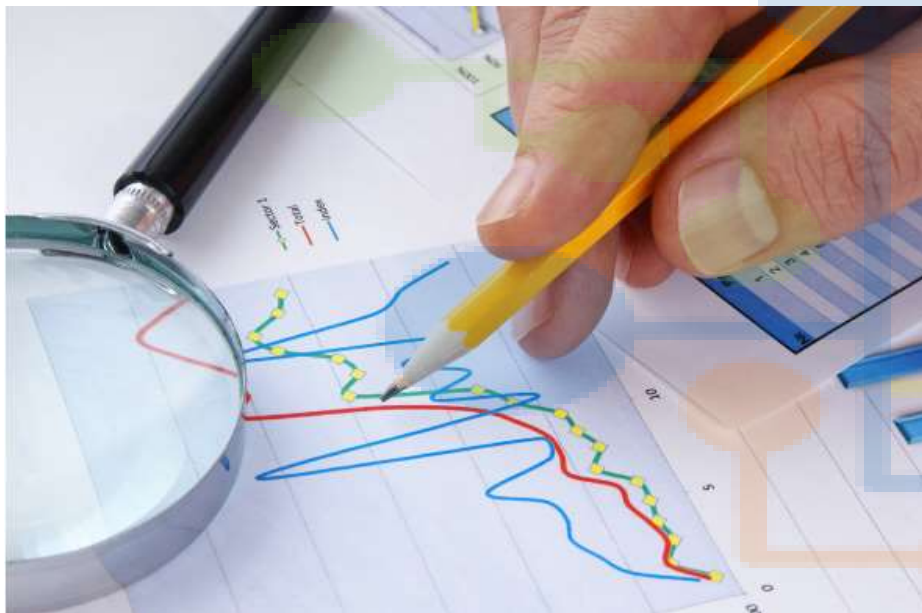
Data Science
Academy

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



**Data Science
Academy**

O que é um Modelo Preditivo?



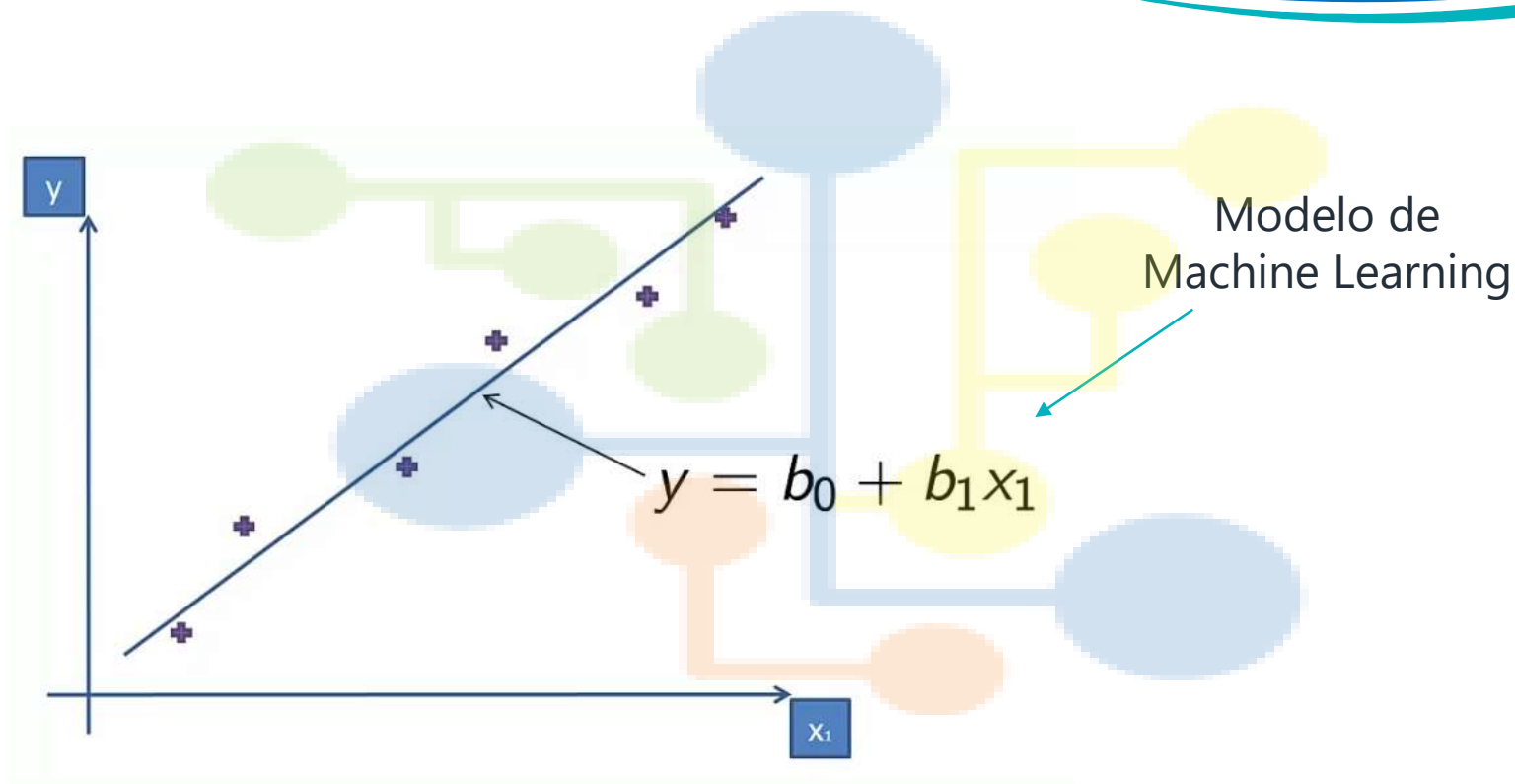
Modelo Preditivo

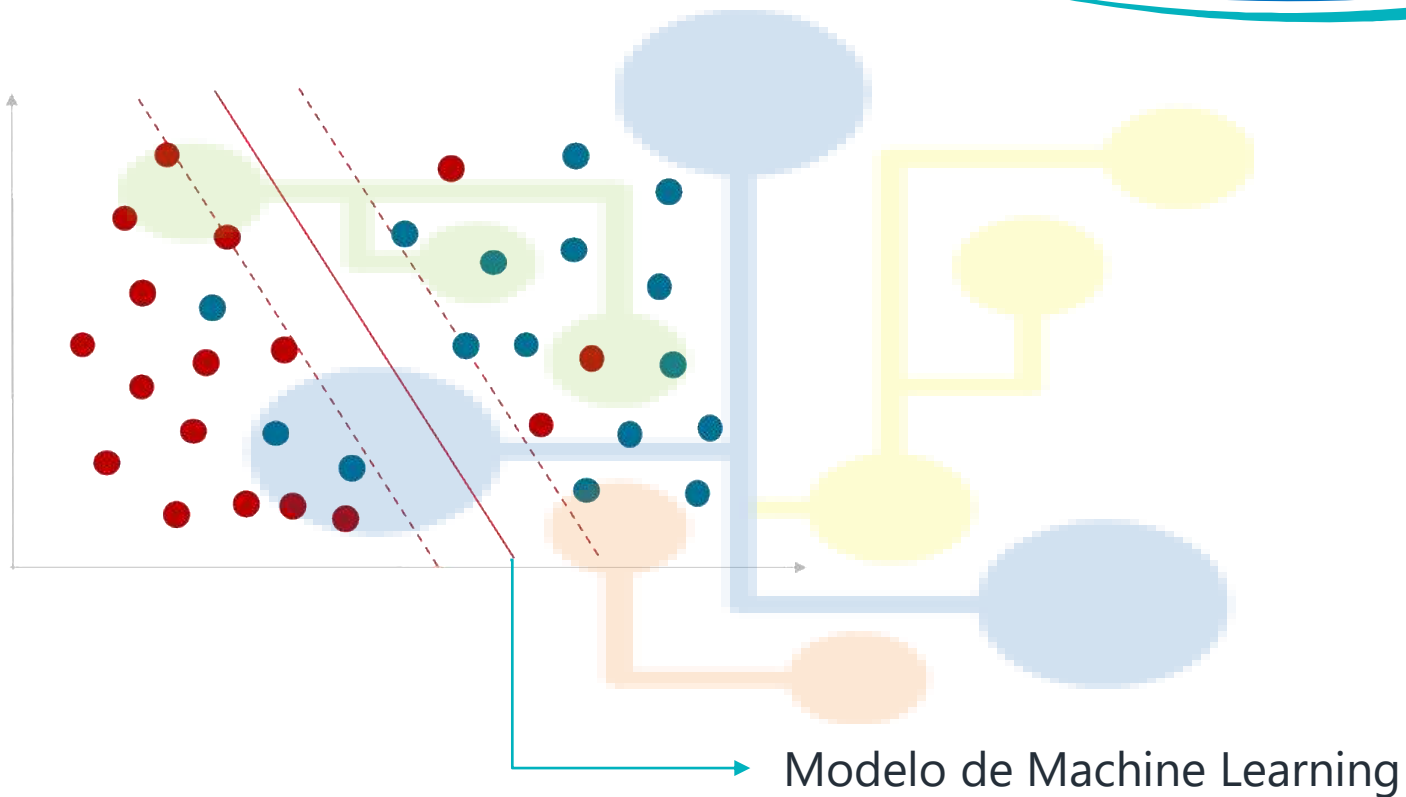


Modelo Preditivo é uma função matemática, aproximada, encontrada através do treinamento com dados e que permite fazer previsões.



O modelo pode empregar uma equação linear simples, ou pode ser uma rede neural complexa.





An abstract background diagram consisting of several interconnected nodes (circles) in blue, green, yellow, and orange, connected by lines of the same color, creating a network-like structure.

Como o Modelo Preditivo é Construído?



Classificação

Variáveis Preditoras

Espécie	Tamanho (Petal)	Largura (Petal)	Tamanho (Sepal)	Largura (Sepal)
Setosa	5.1	3.5	1.4	0.2
Setosa	4.9	3.0	1.4	0.2
Versicolor	7.0	3.2	4.7	1.4
Versicolor	6.4	3.2	4.5	1.5
Virgínica	6.3	3.3	6.0	2.5
Virgínica	5.8	2.7	5.1	1.9

Classe (Variável Target ou Variável Alvo ou Label)



O objetivo do aprendizado de máquina é aprender a aproximação da função f que melhor representa a relação entre os atributos de entrada (chamadas variáveis preditoras) com a variável target.



Regressão

Variáveis Predictoras

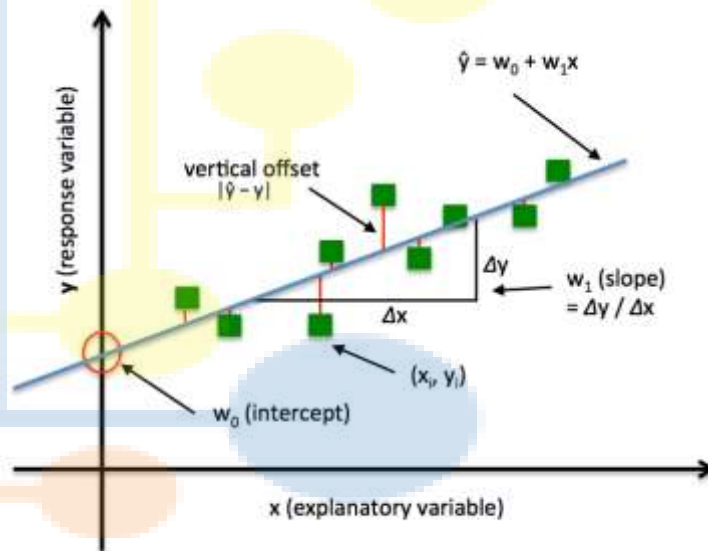
Salário (R\$)	Idade	Anos de Estudo	Salário Médio dos Pais	Número de Filhos
8.000	28	15	9.000	4
7.450	24	17	6.700	5
12.000	32	20	11.200	3
23.500	35	22	20.300	2
25.900	38	25	22.100	1

Classe (Variável Target ou Variável Alvo ou Label)



Regressão

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$





Principais Métodos de Aprendizagem

Métodos Baseados em Instância

Métodos Probabilísticos

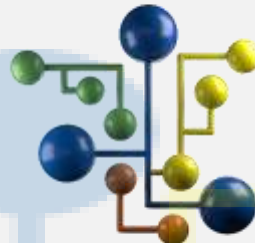
Métodos Baseados em Procura

Métodos Baseados em Otimização



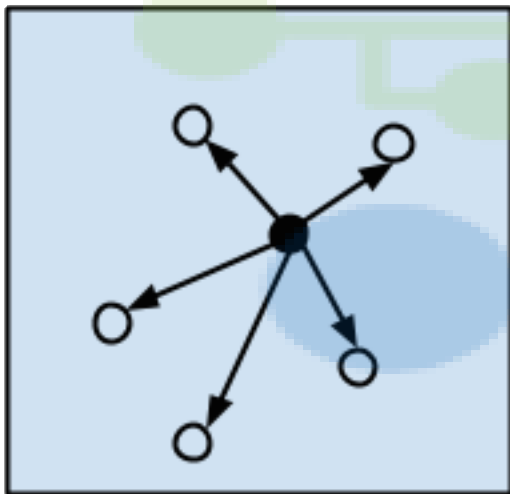
Data Science
Academy

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



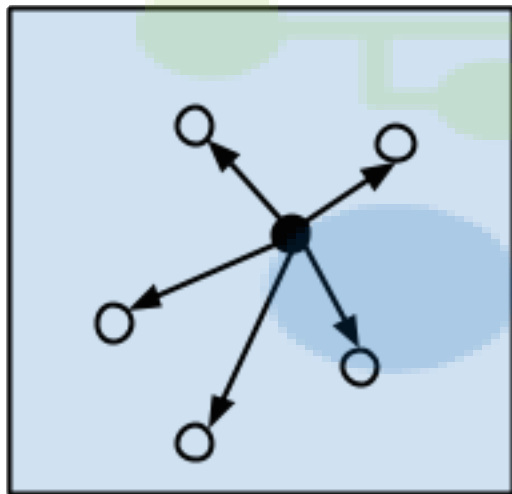
**Data Science
Academy**

Aprendizagem Baseada em Instância

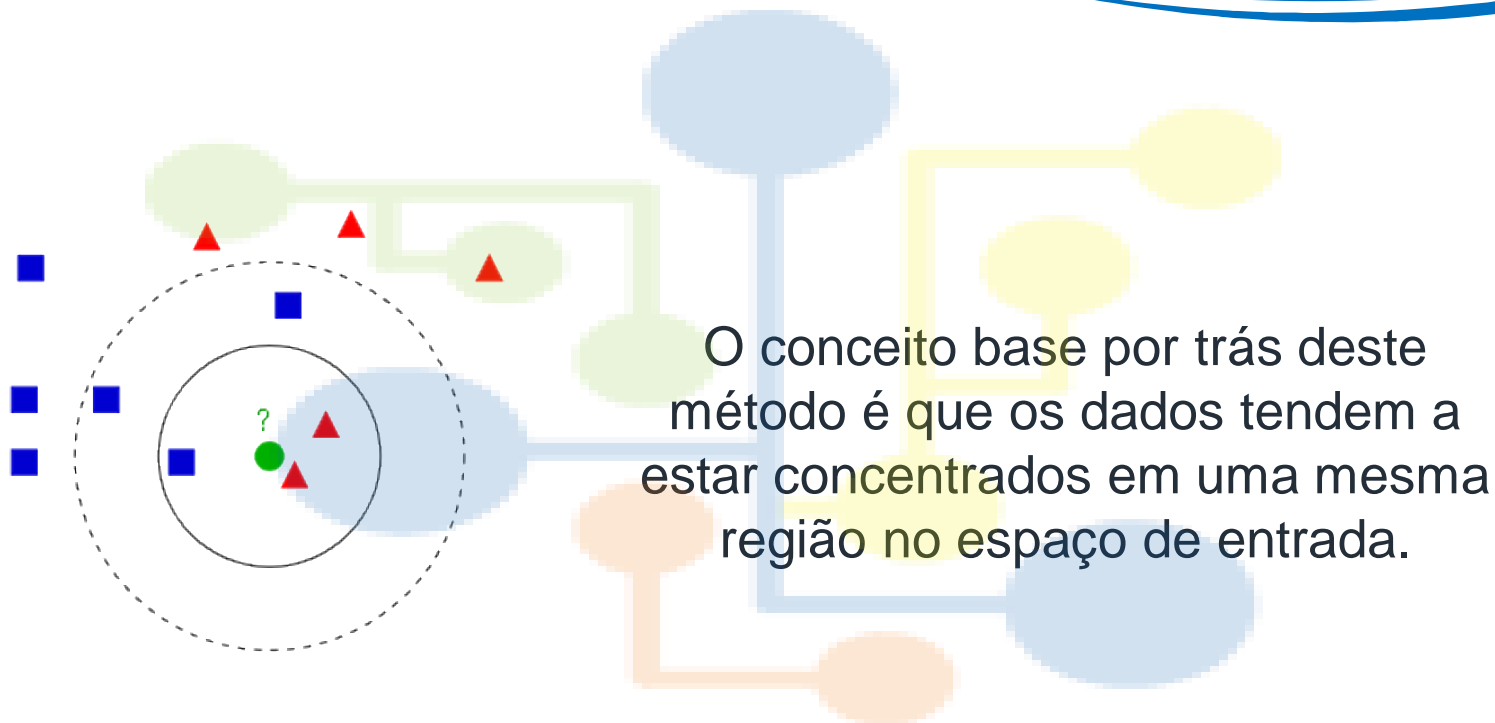


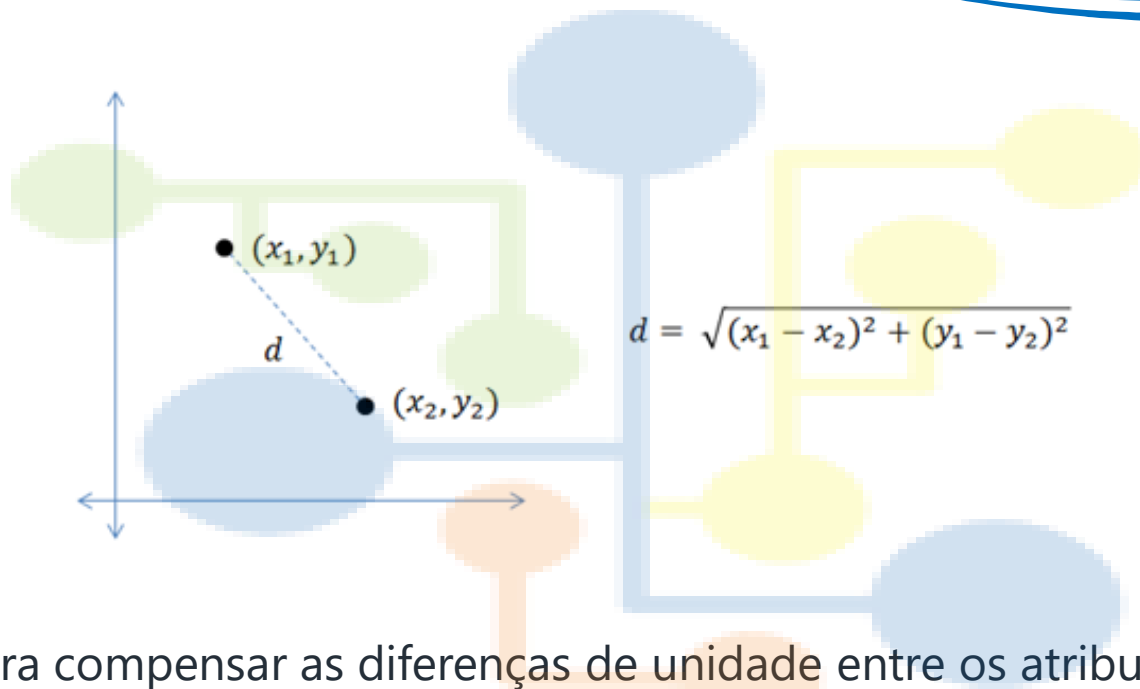
Métodos de Aprendizagem Baseados em Instâncias





A aprendizagem consiste
somente em armazenar os
exemplos de treinamento!





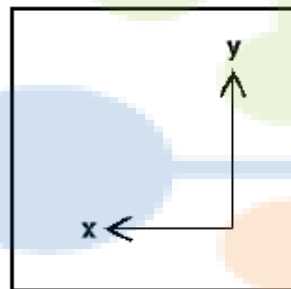
Para compensar as diferenças de unidade entre os atributos contínuos, temos que normalizá-los para ficar no intervalo $[0,1]$.



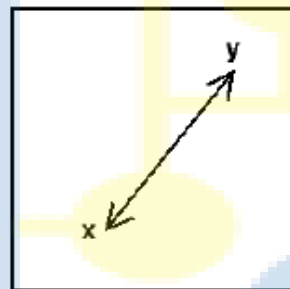
Métodos de aprendizagem baseados em instâncias assumem que as instâncias podem ser representadas como pontos em um espaço Euclidiano.



Outras Medidas de Distância



Manhattan

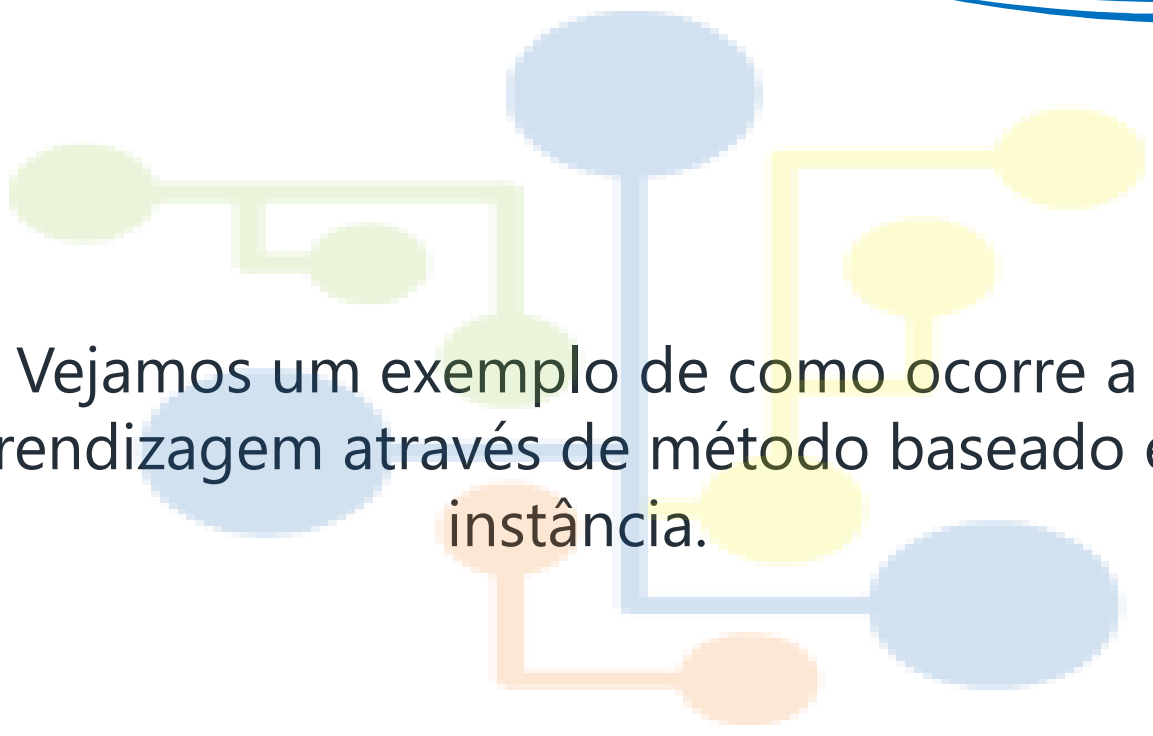


Euclidean

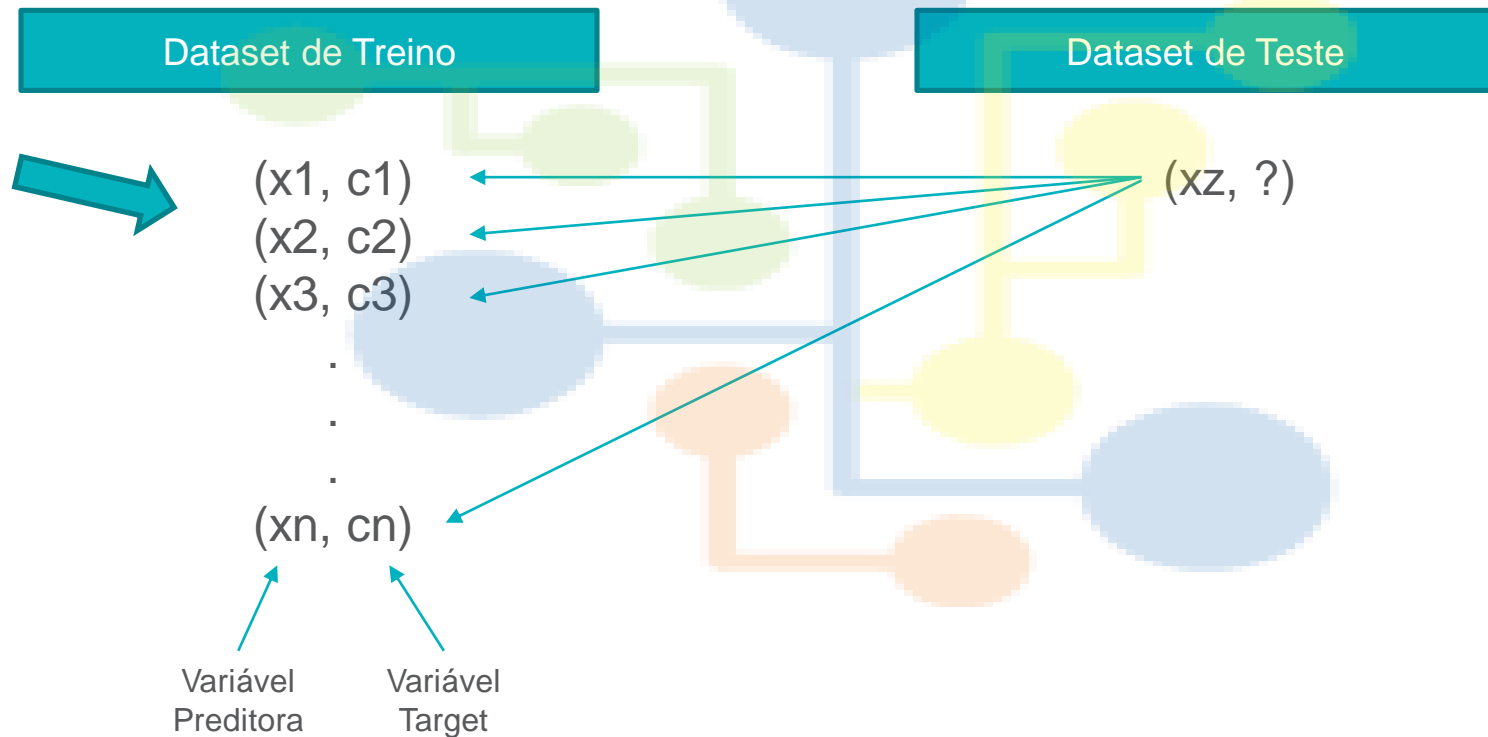


Outras Medidas de Distância

- Correlação de Pearson – Coeficiente de correlação usado em estatística. Muito usado em bioinformática.
- Similaridade de Cosseno – Cosseno do ângulo entre os vetores. Usado para classificação de textos e outros dados de alta dimensão.
- Distância de edição – Usado para medir distância entre strings. Usado em classificação de textos e bioinformática.

A decorative background diagram consisting of several colored circles (blue, green, yellow, orange) connected by lines, forming a network-like structure. The circles are of varying sizes and are connected by lines of the same color, creating a complex, interconnected web.

Vejamos um exemplo de como ocorre a aprendizagem através de método baseado em instância.





Dataset de Treino

Dataset de Teste

(x_1, c_1)

(x_2, c_2)

(x_3, c_3)

.

.

.

(x_n, c_n)

$(x_z, ?)$

Variável
Preditora

Variável
Target



Dataset de Treino

Dataset de Teste

(x_1, c_1)

(x_2, c_2)

(x_3, c_3)

.

.

.

(x_n, c_n)

d_1

d_2

d_3

d_n

$(x_z, ?)$

Variável
Preditora

Variável
Target

Qual a similaridade entre x_1 e x_z ?
Distância Euclidiana



Dataset de Treino

Dataset de Teste

(x_1, c_1)

(x_2, c_2)

(x_3, c_3)

.

.

.

(x_n, c_n)

d_1

d_2

d_3

d_n

$(x_z, ?)$

Variável
Preditora

Variável
Target

Qual a similaridade entre x_1 e x_z ?
Distância Euclidiana



Dataset de Treino

Dataset de Teste

(x_1, c_1)

$(x_2, \mathbf{c_2})$

(x_3, c_3)

.

.

.

(x_n, c_n)

d_1

d_2

d_3

d_n

$(x_z, \mathbf{c_2})$

Variável
Preditora

Variável
Target

Qual a similaridade entre x_1 e x_z ?
Distância Euclidiana



Métodos Baseados em Instância

Constroem aproximações da função alvo para cada instância de teste diferente.



Métodos Baseados em Instância

Os métodos de aprendizagem baseados em instâncias são métodos não paramétricos.



Métodos Baseados em Instância

Uma desvantagem é o alto custo para classificação.
Toda computação ocorre no momento da classificação.



Outras Características:

- Ao contrário das outras abordagens, não ocorre a construção de um modelo de classificação explícito.
- Novos exemplos são classificados com base na comparação direta e similaridade aos exemplos de treinamento.
- Treinamento pode ser fácil, apenas memoriza exemplos.
- Teste pode ser intenso computacionalmente pois requer comparação com todos os exemplos de treinamento.
- Métodos baseados em instância favorecem a similaridade global e não a simplicidade do conceito.

A diagram illustrating the K-Nearest Neighbors (KNN) algorithm. It features several colored circles (blue, green, yellow, orange) connected by lines of the same color, forming a network. The circles represent data points or objects, and the lines represent their relationships or distances. The text 'KNN' is centered over the diagram.

KNN

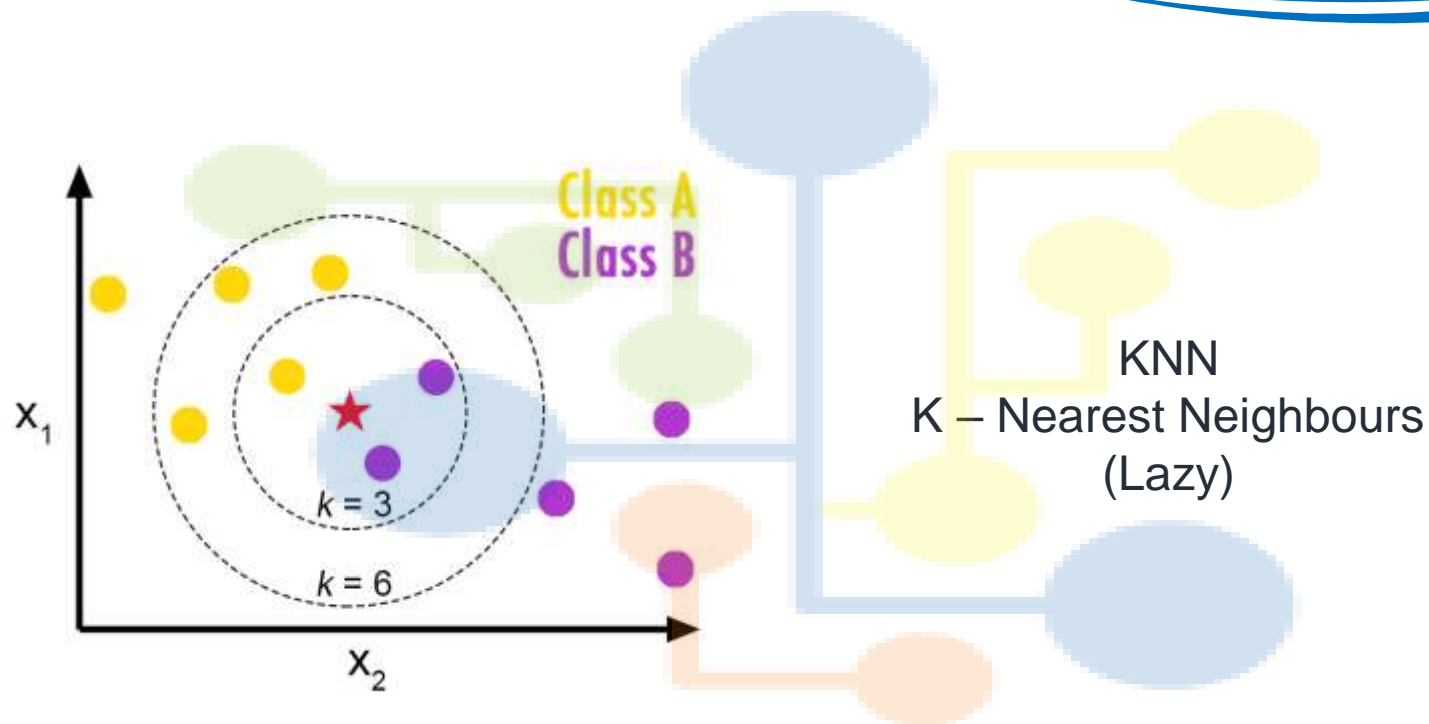
K – Nearest Neighbours

Objetos relacionados ao mesmo conceito são semelhantes entre si.



Regra do KNN

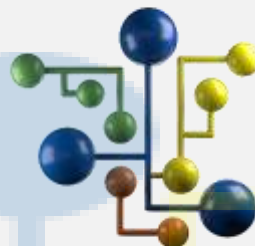
Classificar xz atribuindo a ele o rótulo representado mais frequentemente dentre as k amostras mais próximas e utilizando um esquema de votação.





Data Science
Academy

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



**Data Science
Academy**

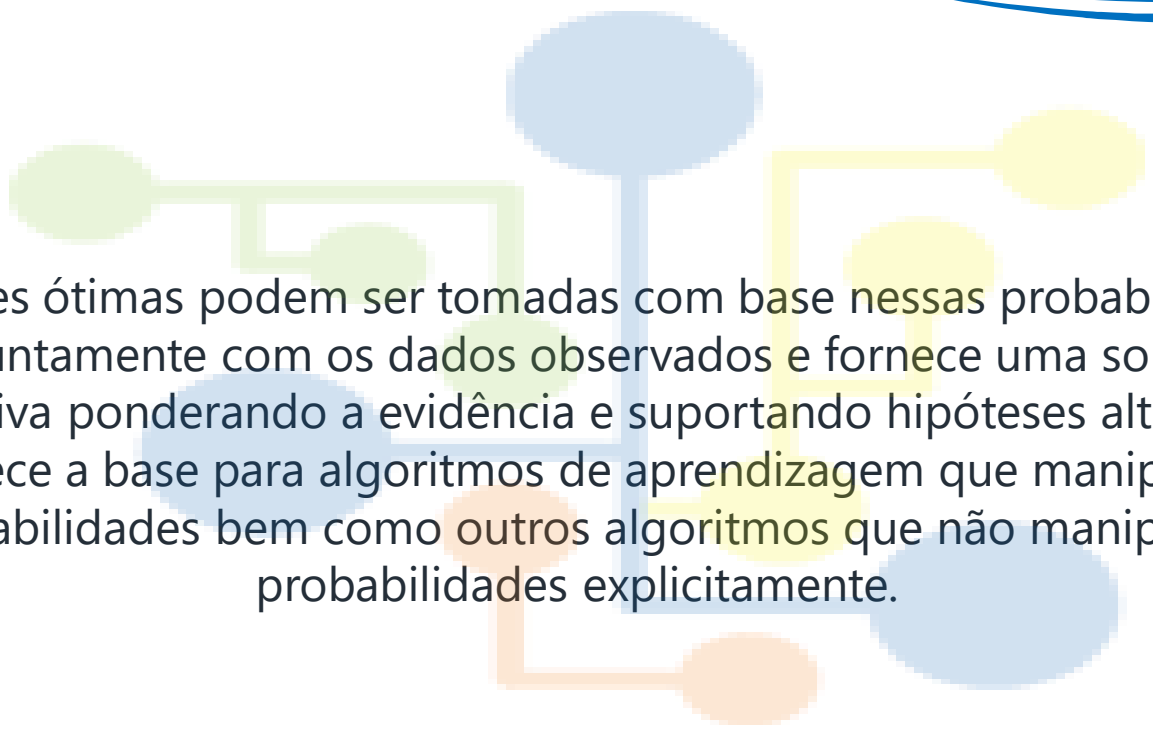
Aprendizagem Baseada em Métodos Probabilísticos



Métodos Probabilísticos

Os métodos probabilísticos bayesianos assumem que a probabilidade de um evento A, que pode ser uma classe, dado em um evento B, não depende apenas da relação entre A e B, mas também da probabilidade de observar A independentemente de B.

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

A decorative background diagram consisting of several colored circles (blue, green, yellow, orange) connected by lines, forming a network-like structure.

Decisões ótimas podem ser tomadas com base nessas probabilidades conjuntamente com os dados observados e fornece uma solução quantitativa ponderando a evidência e suportando hipóteses alternativas. Fornece a base para algoritmos de aprendizagem que manipulam probabilidades bem como outros algoritmos que não manipulam probabilidades explicitamente.



Os Métodos Probabilísticos são relevantes por dois motivos:

1. Fornecem algoritmos de aprendizagem práticos:

- Aprendizagem Naïve Bayes
- Aprendizagem de Redes Bayesianas
- Combinam conhecimento a priori com os dados observados



Os Métodos Probabilísticos são relevantes por dois motivos:

2. Fornecem uma estrutura conceitual útil:

Fornecem a “norma de ouro” (regra do menor erro possível) para avaliar outros algoritmos de aprendizagem.



Cada exemplo de treinamento pode decrementar ou incrementar a probabilidade de uma hipótese ser correta.



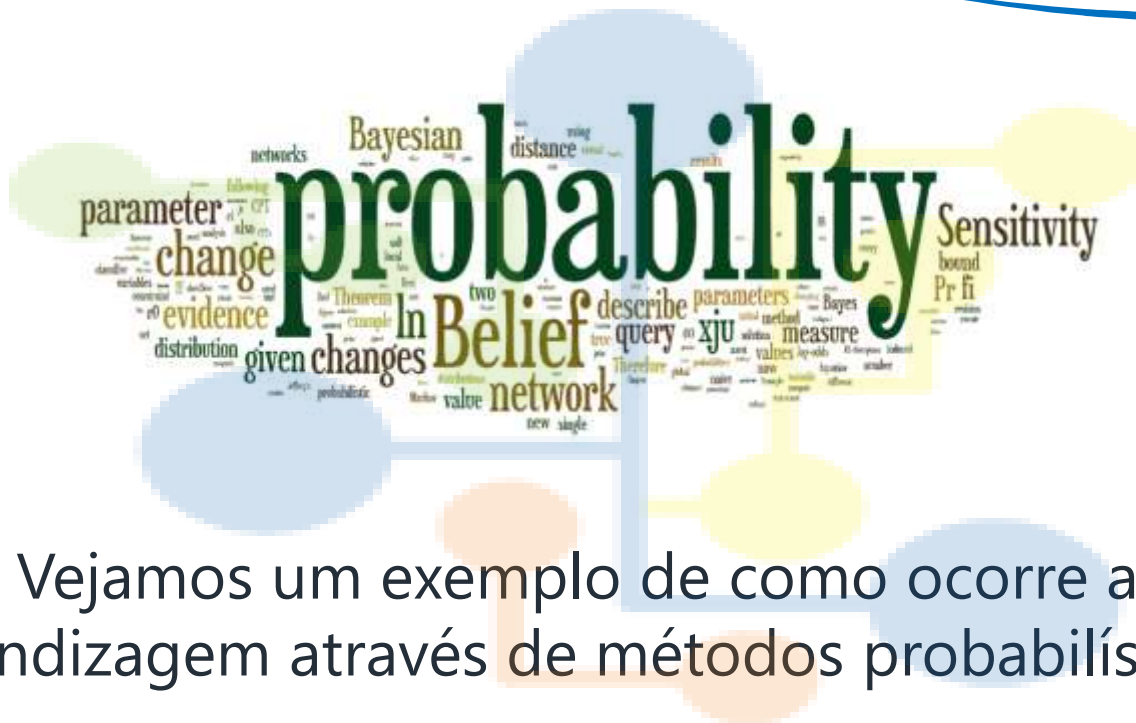
Conhecimento a priori pode ser combinado com os dados observados para determinar a probabilidade de uma hipótese.



Métodos Bayesianos podem acomodar hipóteses que fazem
predições probabilísticas.
(Ex: Este paciente tem uma chance de 93% de se recuperar)



Novas instâncias podem ser classificadas combinando a probabilidade de múltiplas hipóteses ponderadas pelas suas probabilidades.



Vejamos um exemplo de como ocorre a aprendizagem através de métodos probabilísticos.



$P(\text{Doença} = \text{presente}) = 0.08$

$P(\text{Doença} = \text{ausente}) = 0.92$

$P(\text{Teste} = \text{positivo} \mid \text{Doença} = \text{presente}) = 0.75$

$P(\text{Teste} = \text{negativo} \mid \text{Doença} = \text{ausente}) = 0.96$



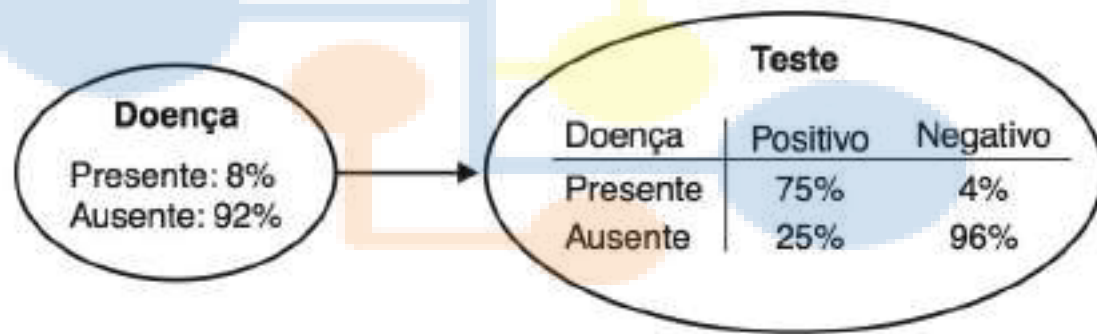


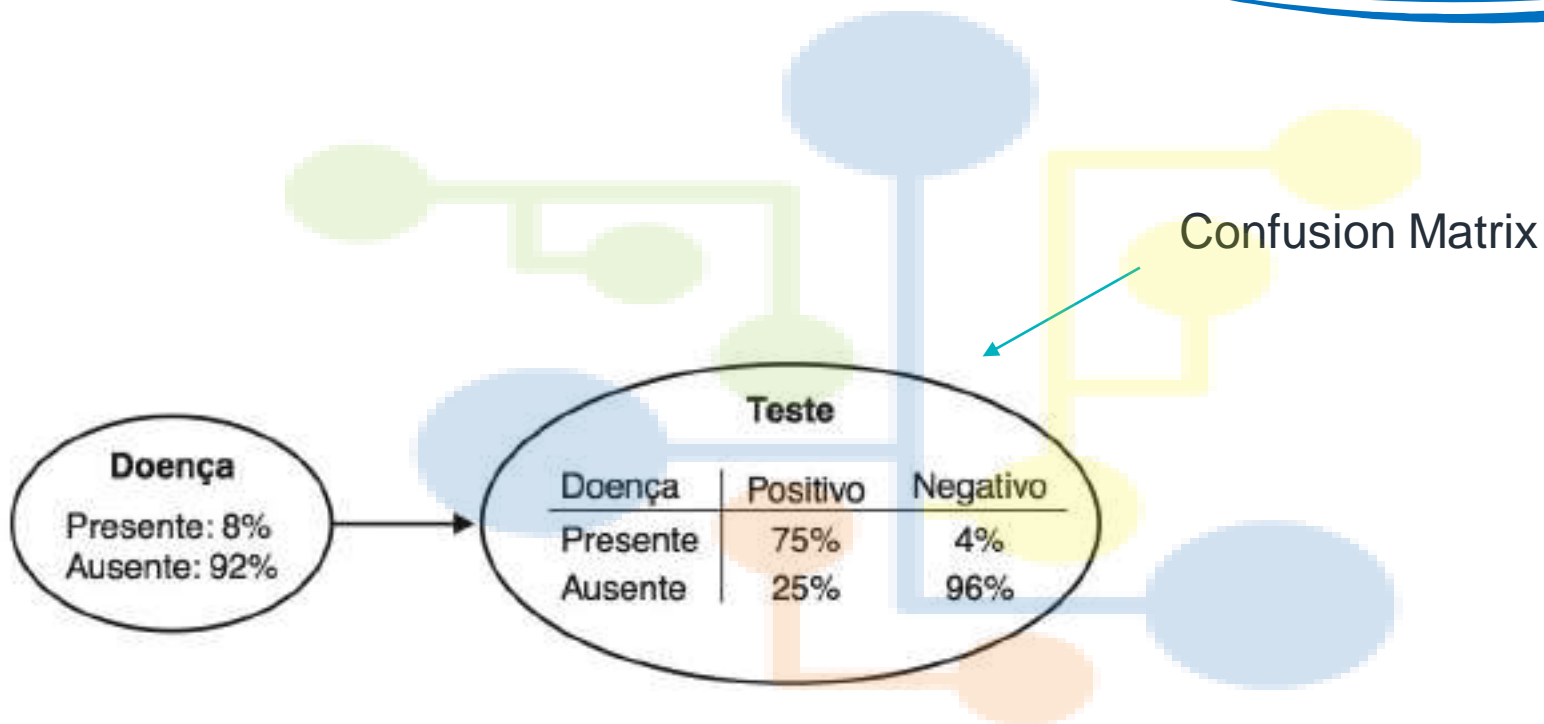
$$P(\text{Doença} = \text{presente}) = 0.08$$

$$P(\text{Doença} = \text{ausente}) = 0.92$$

$$P(\text{Teste} = \text{positivo} \mid \text{Doença} = \text{presente}) = 0.75$$

$$P(\text{Teste} = \text{negativo} \mid \text{Doença} = \text{ausente}) = 0.96$$







$$P(\text{Doença} = \text{presente}) = 0.08$$

$$P(\text{Doença} = \text{ausente}) = 0.92$$

$$P(\text{Teste} = \text{positivo} \mid \text{Doença} = \text{presente}) = 0.75$$

$$P(\text{Teste} = \text{negativo} \mid \text{Doença} = \text{ausente}) = 0.96$$

$$P(A) = P(A|B) \times P(B)$$

Qual é o poder preditivo do teste com respeito à doença? É possível calcular as probabilidades a priori para a variável **Teste**.





P(Teste = positivo) =

$$\begin{aligned} &= P(\text{Teste} = \text{positivo} \mid \text{Doença} = \text{presente}) \times P(\text{Doença} = \text{presente}) \\ &+ P(\text{Teste} = \text{positivo} \mid \text{Doença} = \text{ausente}) \times P(\text{Doença} = \text{ausente}) \\ &= 0.75 \times 0.08 \times 0.04 \times 0.92 = 0.0968 \end{aligned}$$

P(Teste = negativo) =

$$\begin{aligned} &= P(\text{Teste} = \text{negativo} \mid \text{Doença} = \text{presente}) \times P(\text{Doença} = \text{presente}) \\ &+ P(\text{Teste} = \text{negativo} \mid \text{Doença} = \text{ausente}) \times P(\text{Doença} = \text{ausente}) \\ &= 0.25 \times 0.08 \times 0.96 \times 0.92 = 0.9032 \end{aligned}$$





Naïve Bayes Classifier

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$



Thomas Bayes
1702 - 1761

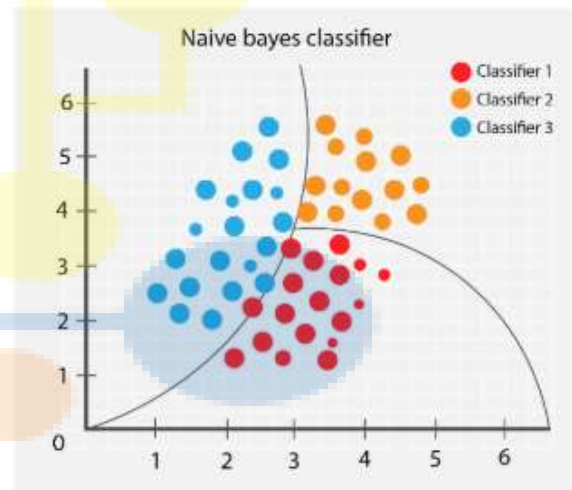
Naive Bayes



Naive Bayes

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$



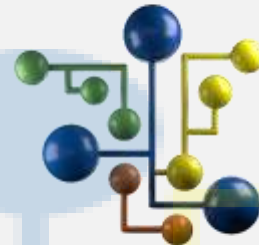


O classificador Naïve Bayes é baseado na suposição simplificadora de que os valores dos atributos são condicionalmente independentes dado o valor alvo.



Data Science
Academy

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



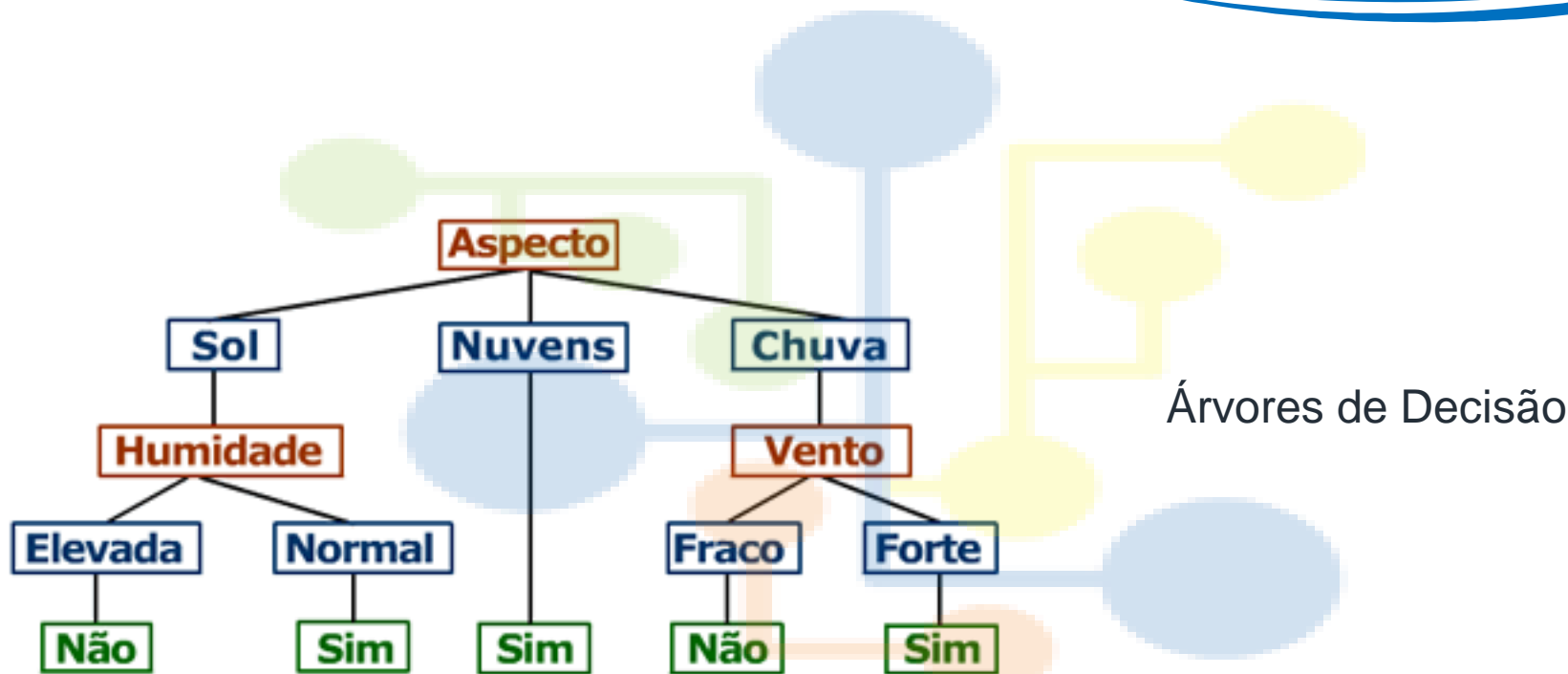
**Data Science
Academy**

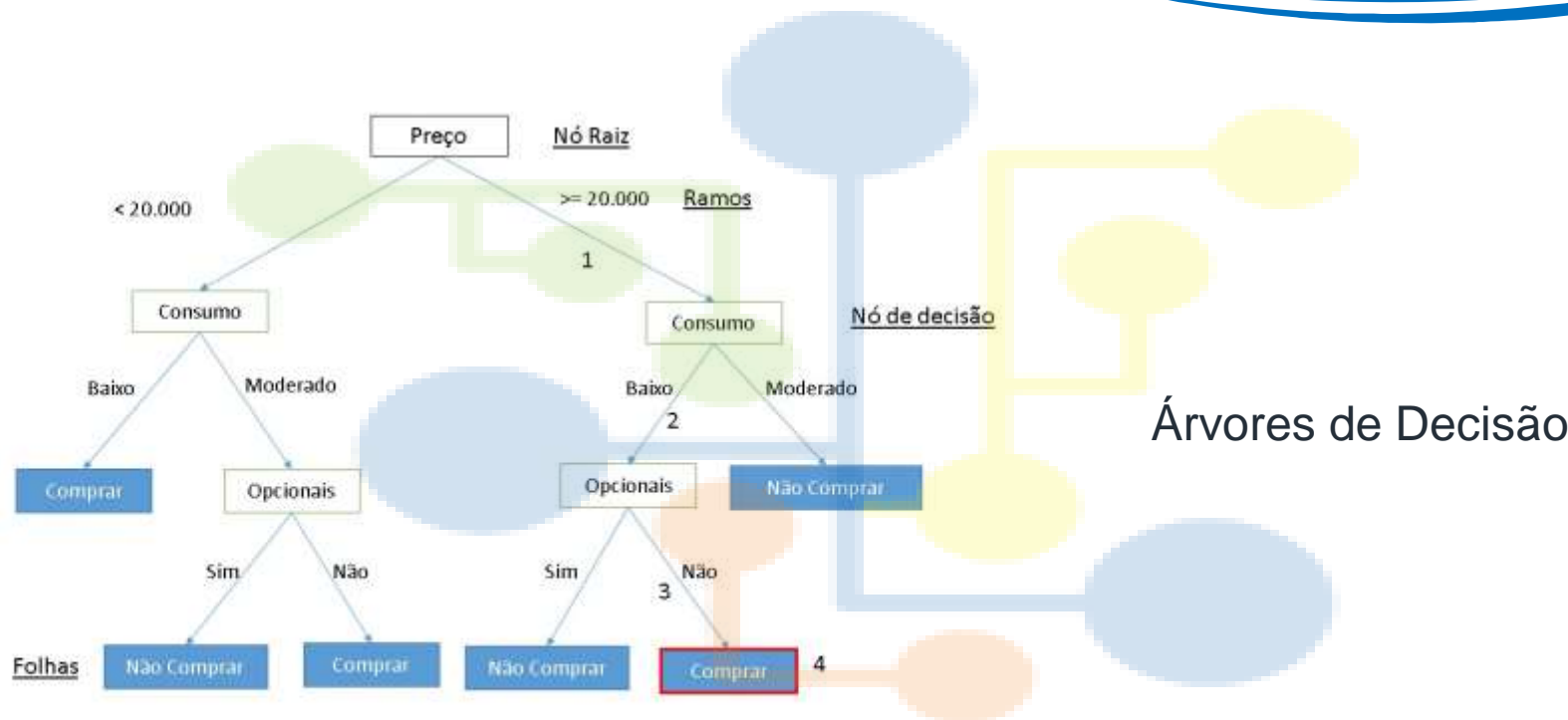
Aprendizagem Baseada em Procura

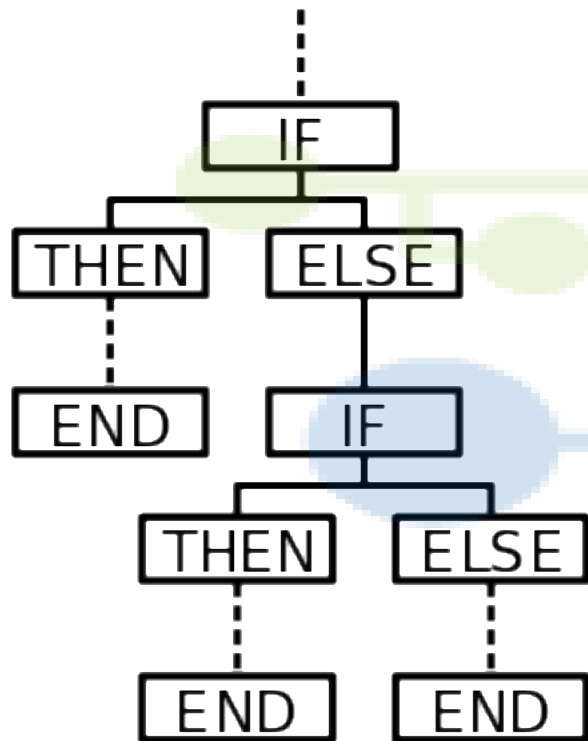


Árvores de Decisão

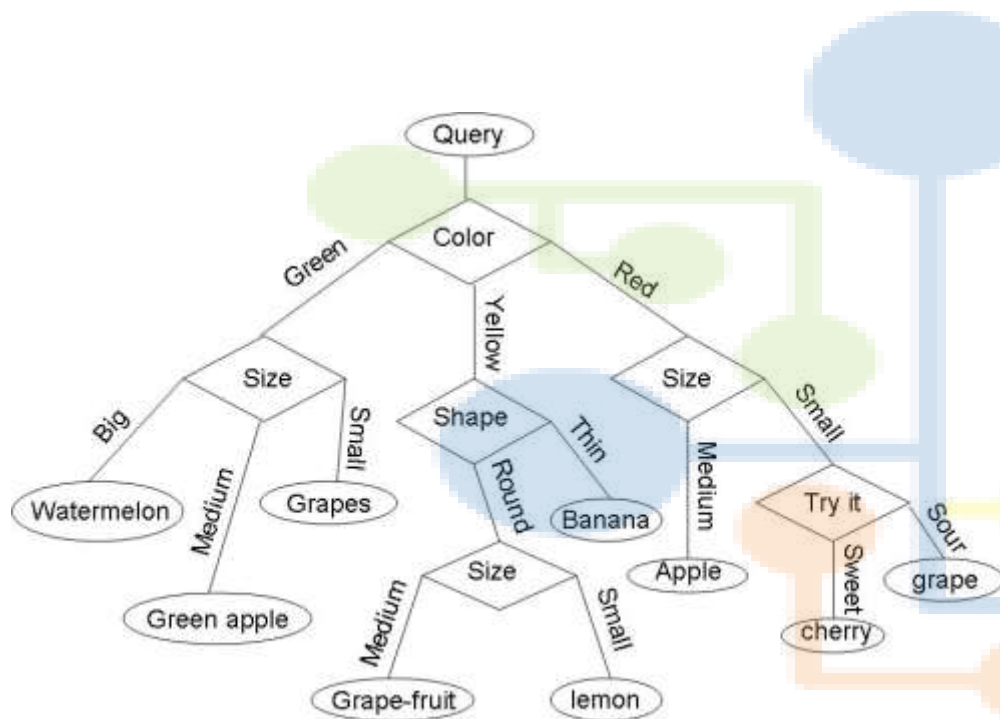








Árvores de decisão também podem ser representadas como conjuntos de regras SE-ENTÃO-SENÃO (IF-THEN-ELSE).

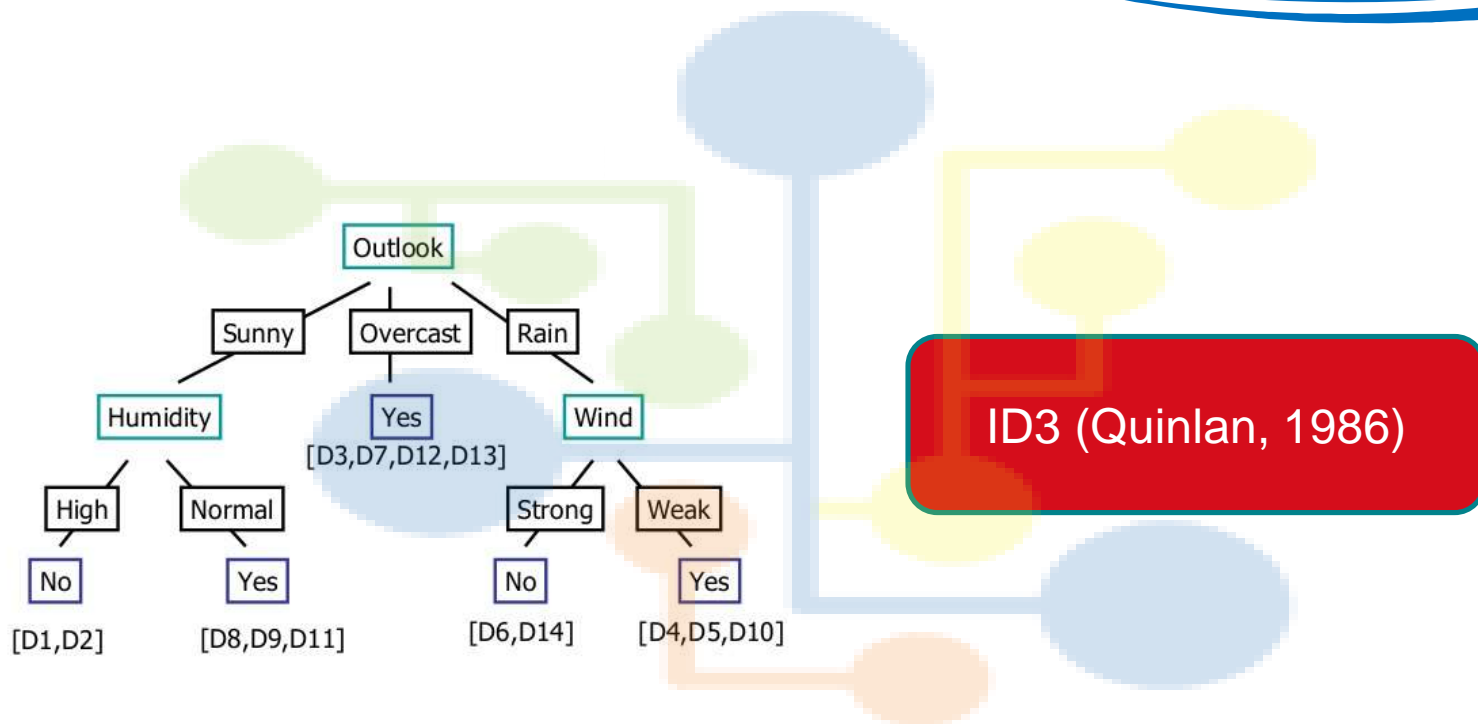


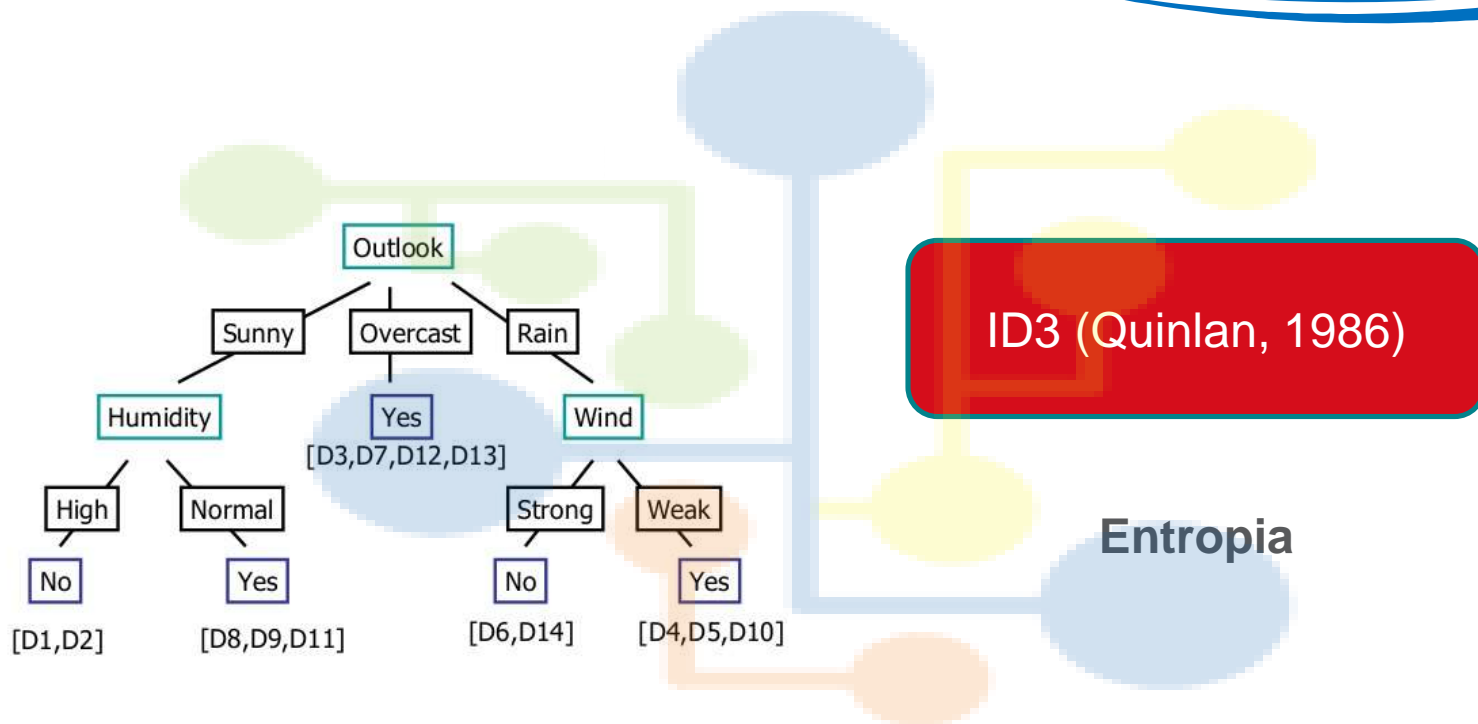
Árvores de decisão classificam instâncias ordenando as árvores acima (ou abaixo), a partir da raiz até alguma folha.

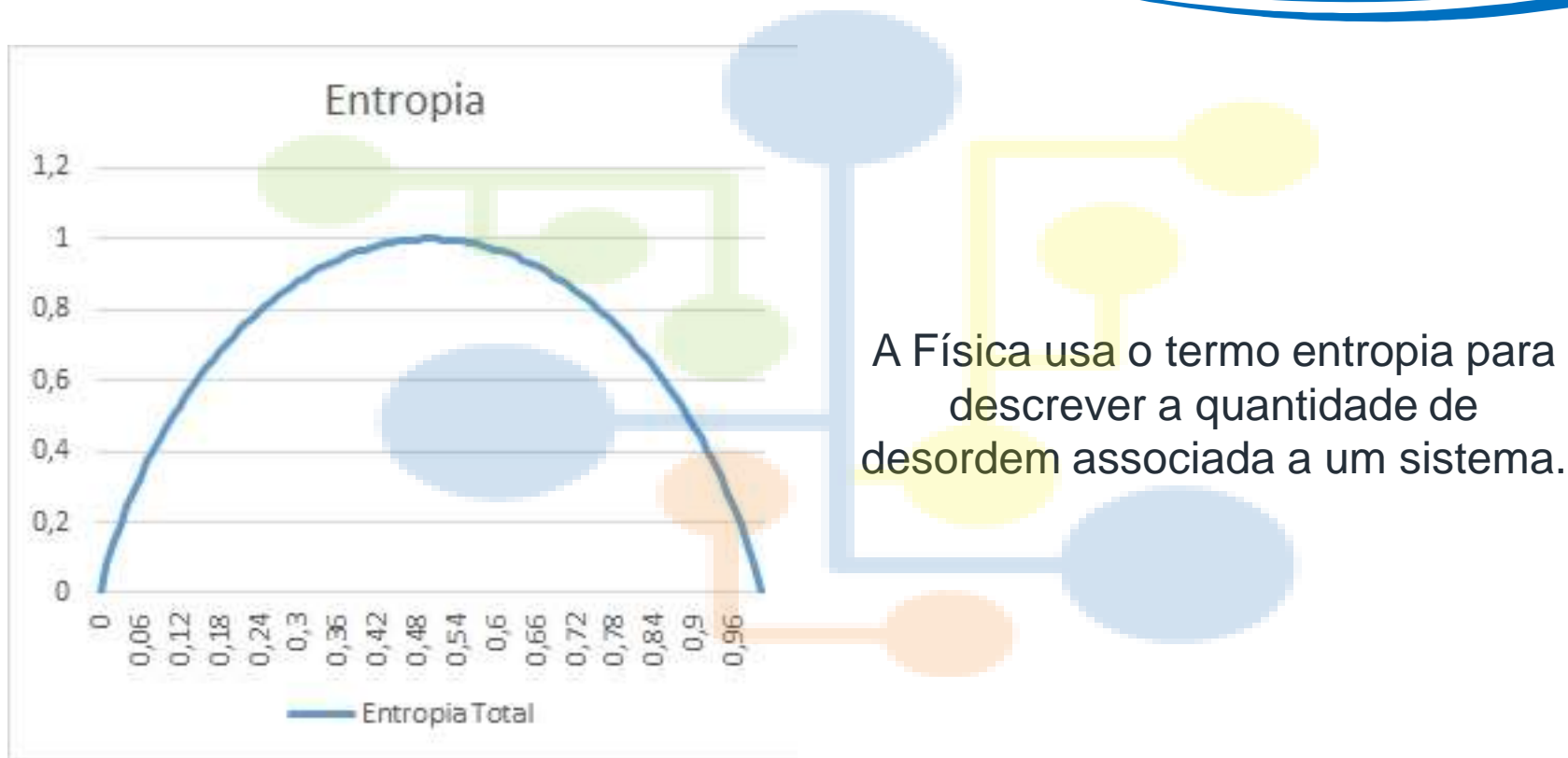


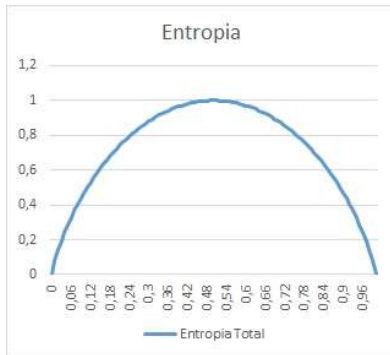
ID3 (Quinlan, 1986)

C4.5 (Quinlan, 1993)





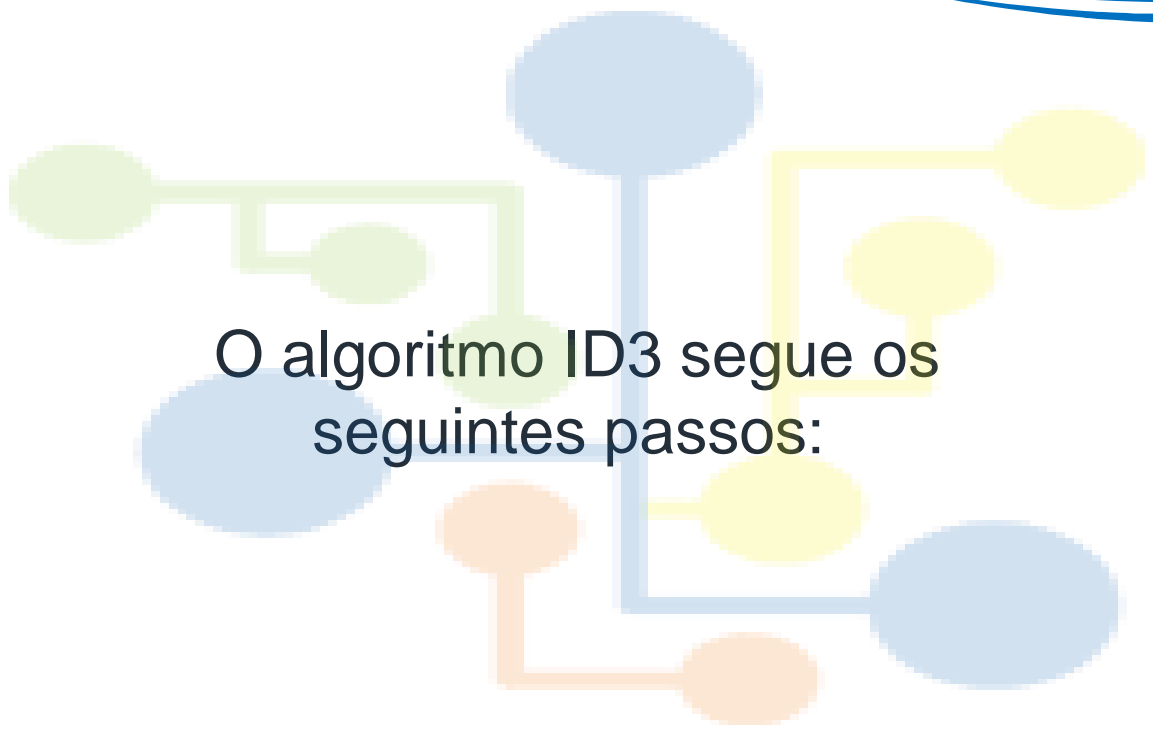




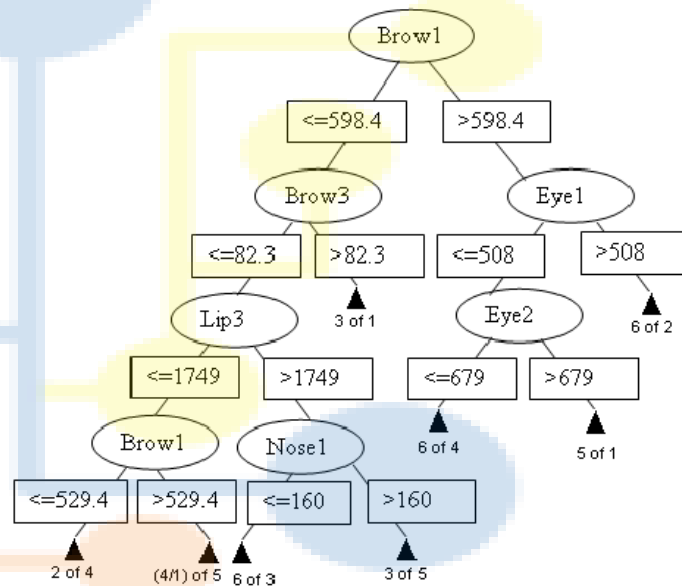
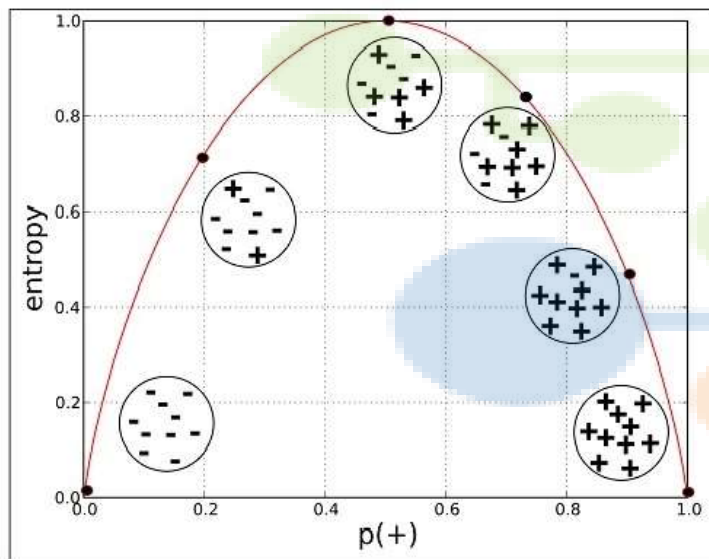
A incerteza ou impureza em um nó pode ser medida através da Entropia.

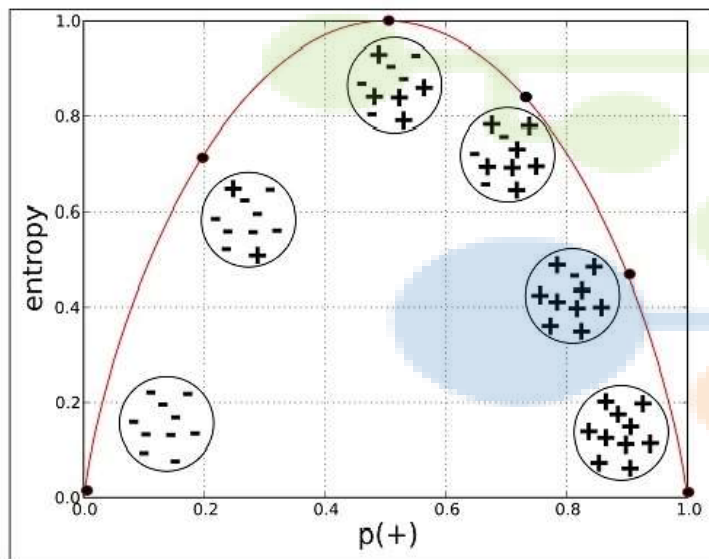
Se todos os exemplos são da mesma classe, então a entropia assume valor mínimo.

Se todas as classes têm o mesmo número de exemplos então a entropia assume o valor máximo.

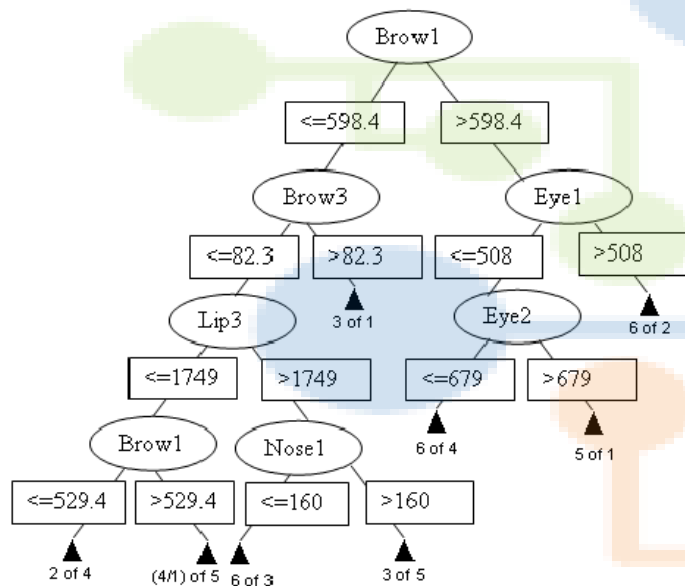
A decorative background diagram consisting of several interconnected nodes and lines. The nodes are colored in light blue, light green, light yellow, and light orange. The lines are colored in light blue, light green, and light yellow, forming a complex network structure.

O algoritmo ID3 segue os seguintes passos:





$$Entropia(S) = \sum p_i \log_2 p_i$$

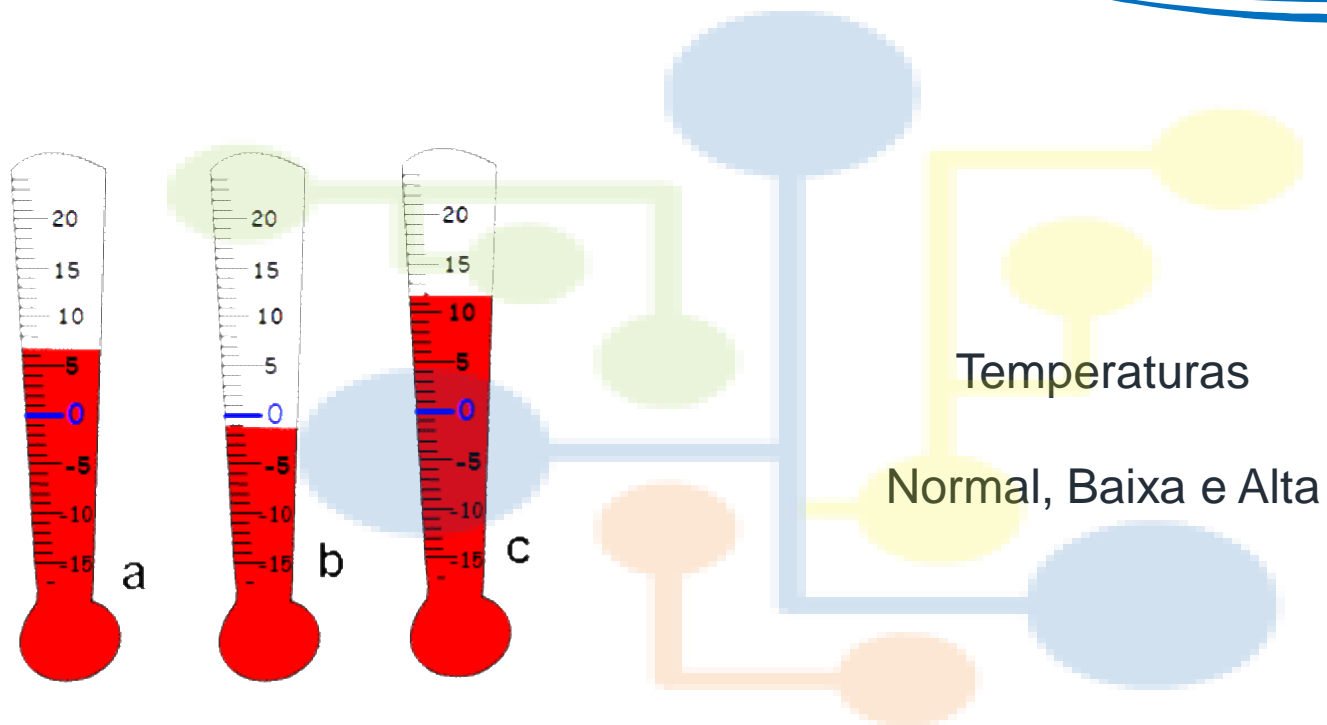


Ganho de Informação



Entropia mede o nível de certeza que temos sobre um evento.

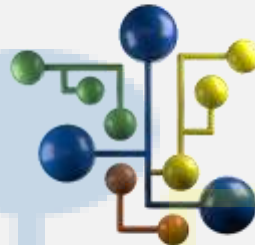
Ganho de Informação mede a efetividade de um atributo em classificar um conjunto de treinamento.





Data Science
Academy

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



**Data Science
Academy**

Aprendizagem Baseada em Otimização

A background diagram showing a network of interconnected nodes and lines. The nodes are colored in shades of blue, green, yellow, and orange. The lines are colored in shades of blue, green, yellow, and orange. The diagram is centered on the slide and serves as a background for the two main text boxes.

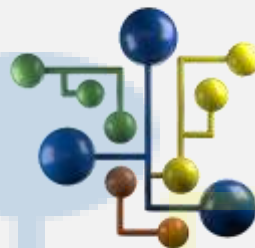
Redes Neurais Artificiais

SVM
(Support Vector Machines)



Data Science
Academy

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



**Data Science
Academy**

Aprendizagem com Redes Neurais Artificiais



A máquina mais incrível
do Planeta Terra



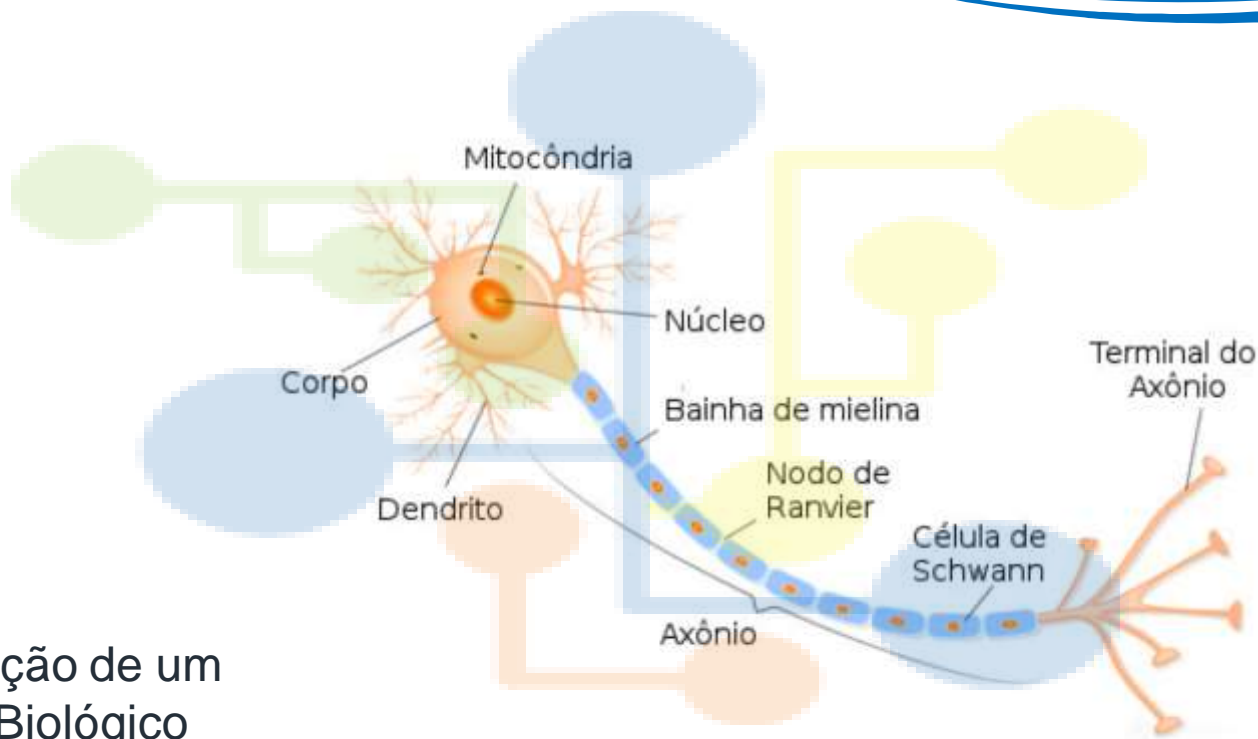
Data Science
Academy

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



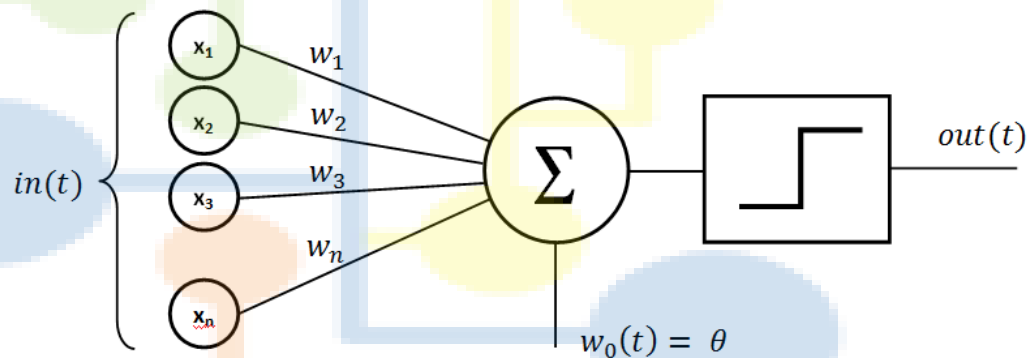


Representação de um Neurônio Biológico





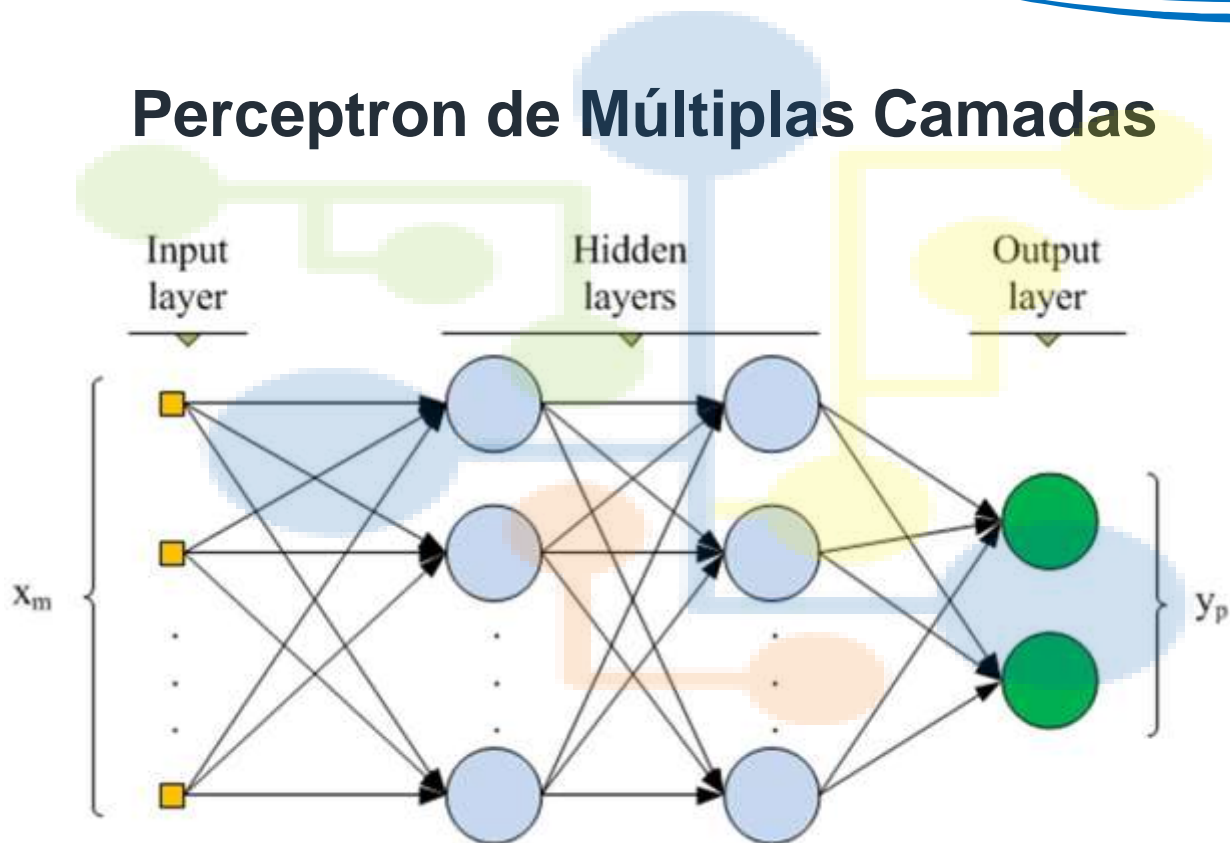
Perceptron

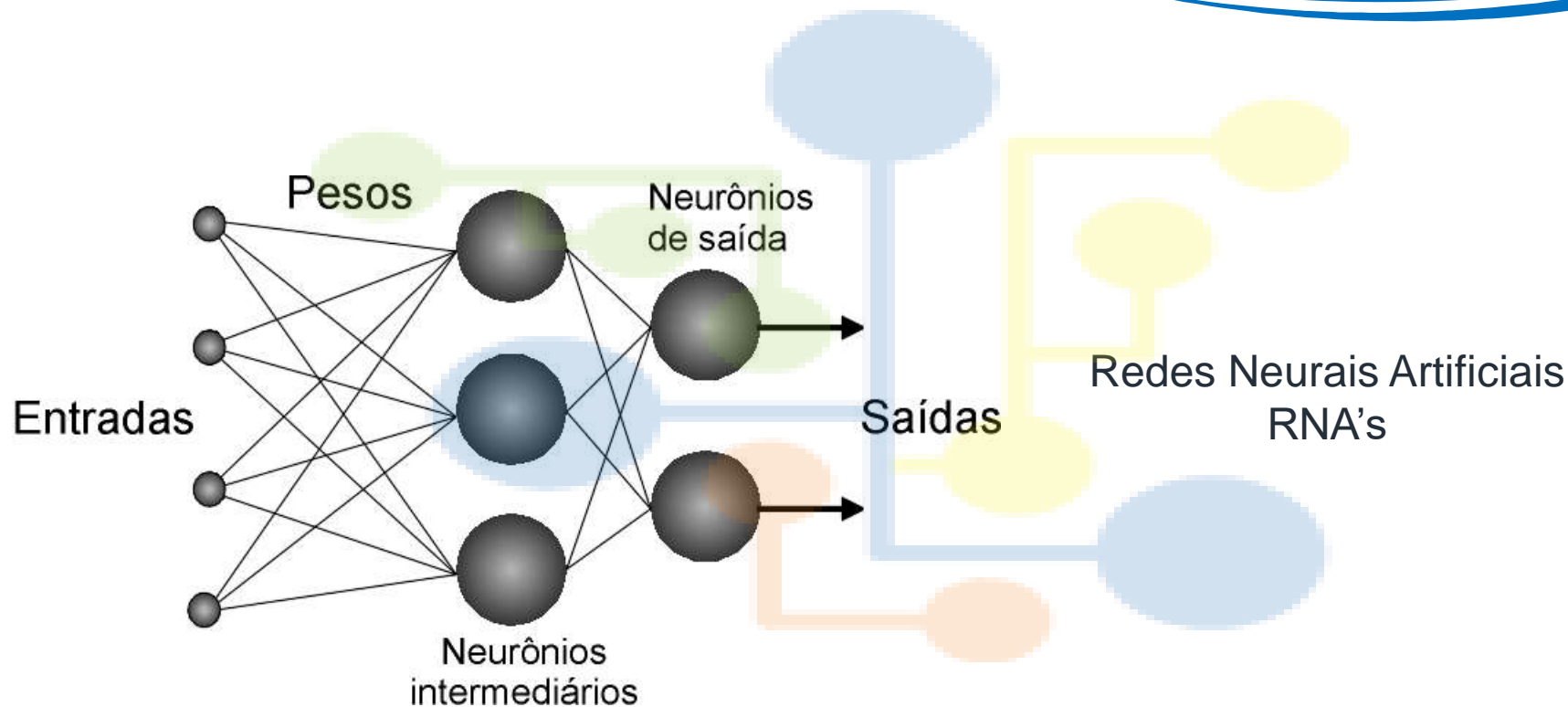


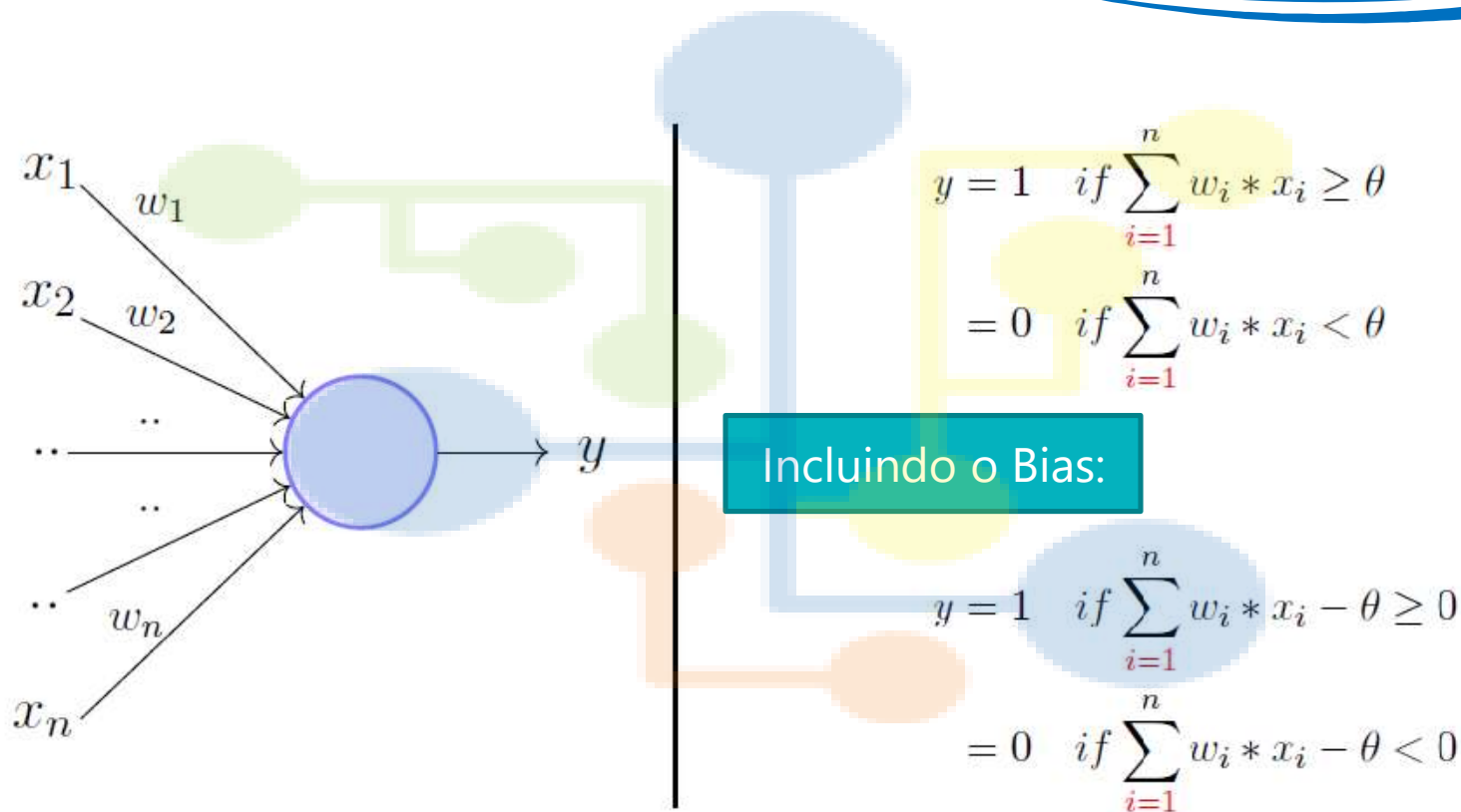
Representação de um
Neurônio Matemático

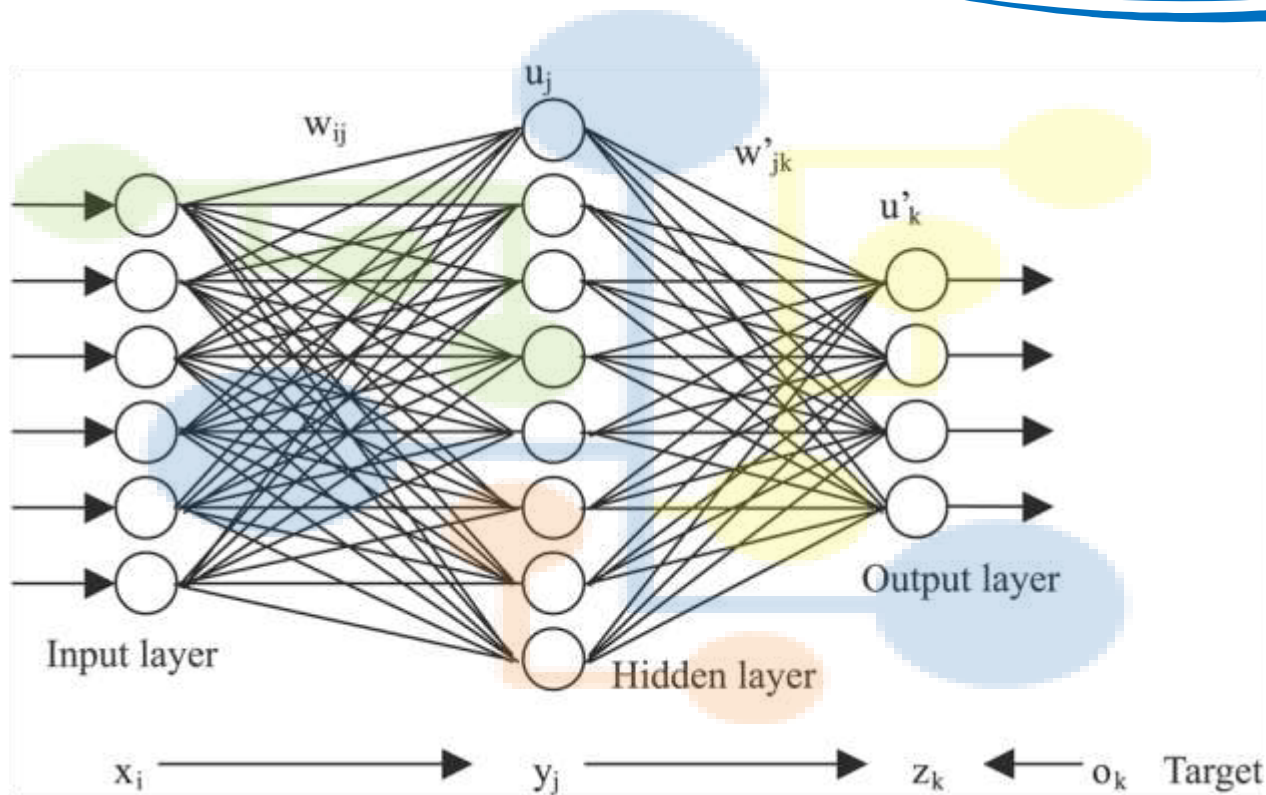


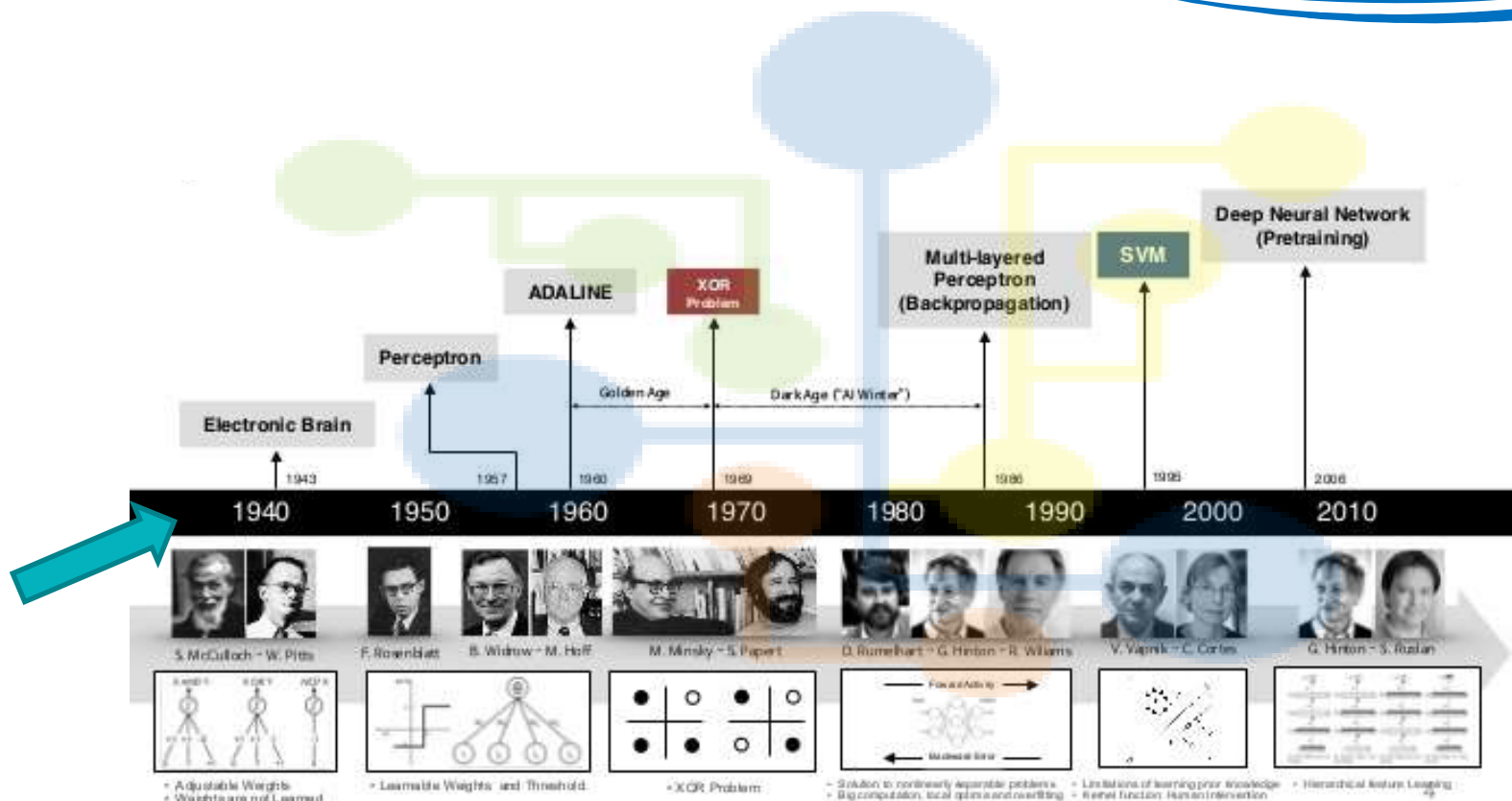
Perceptron de Múltiplas Camadas







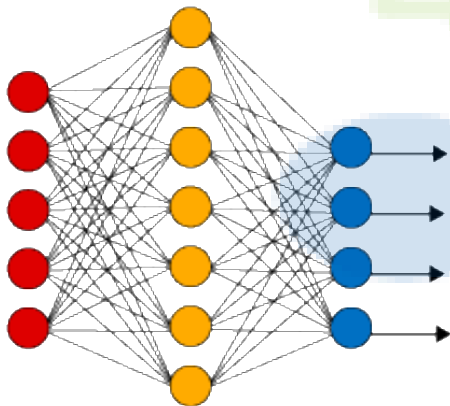




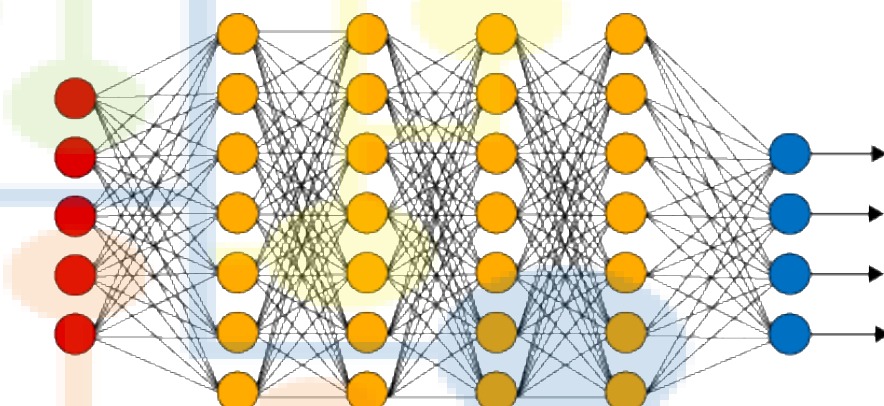




Simple Neural Network



Deep Learning Neural Network



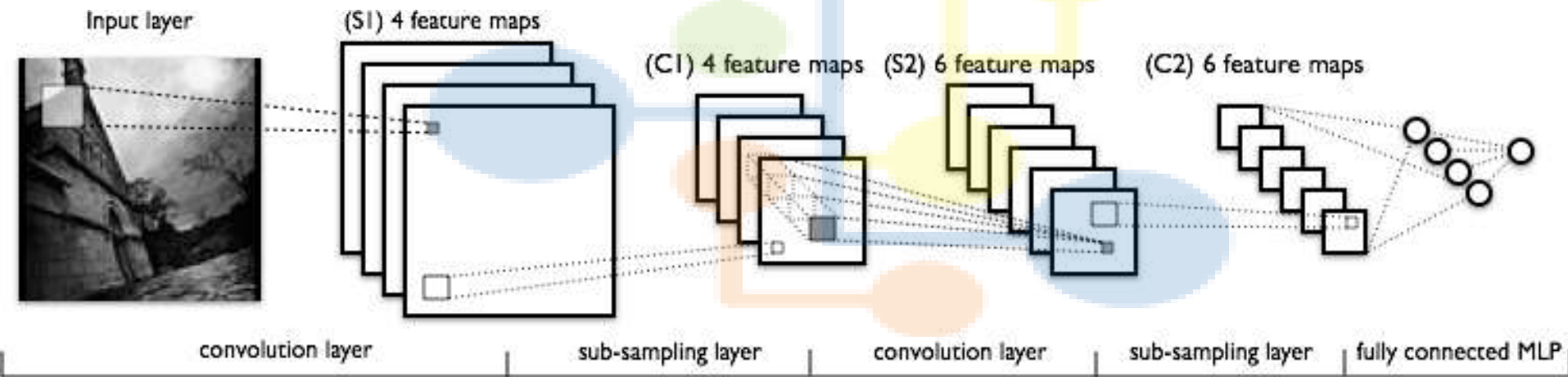
● Input Layer

● Hidden Layer

● Output Layer

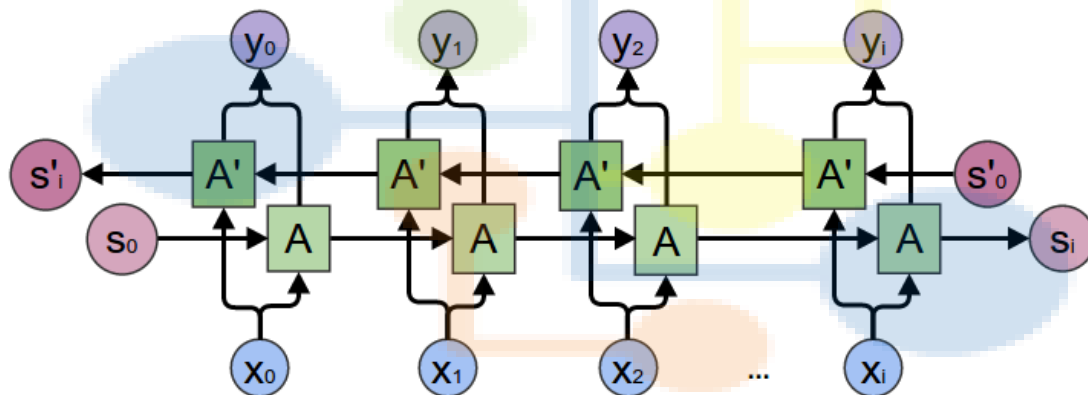


Redes Neurais Convulocionais (CNN – Convolutional Neural Networks)



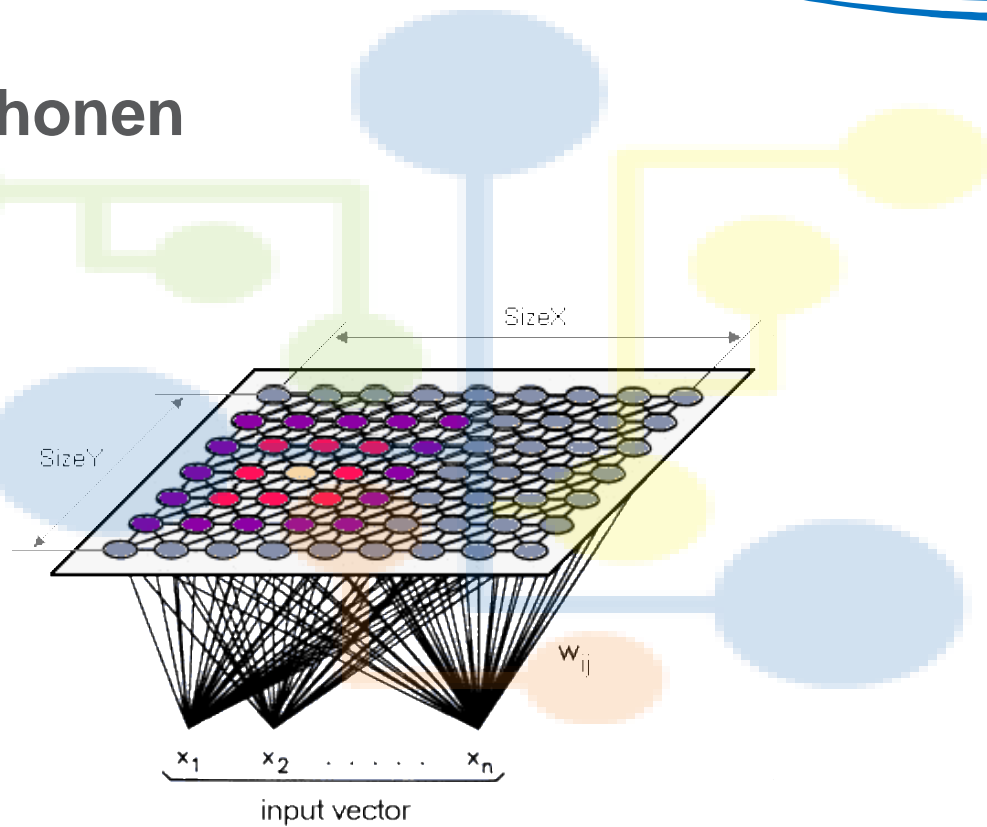


Redes Neurais Recorrentes (RNN – Recurrent Neural Networks)



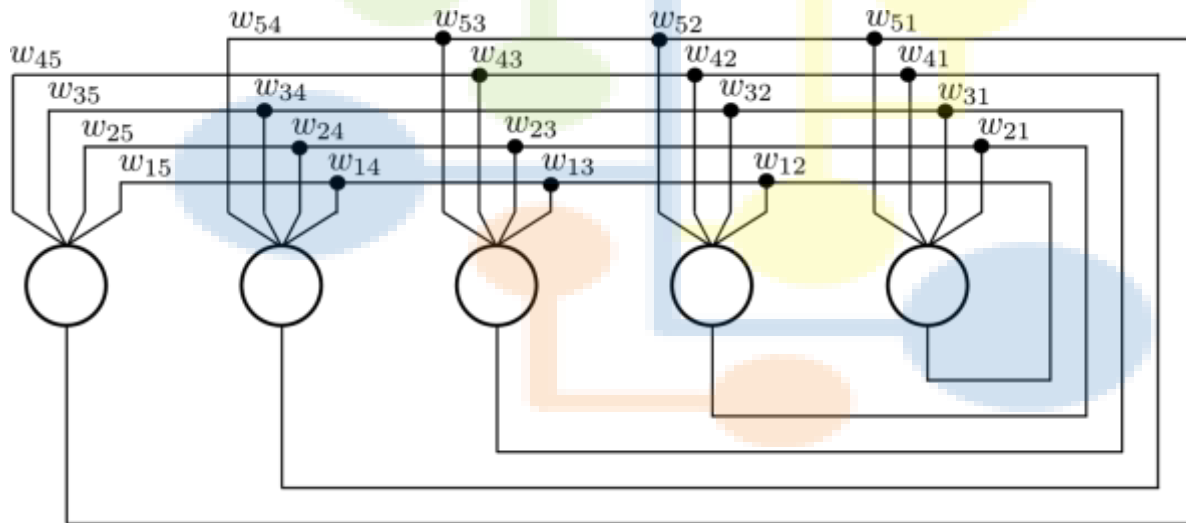


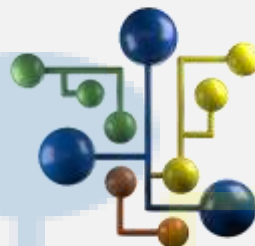
Rede de Kohonen





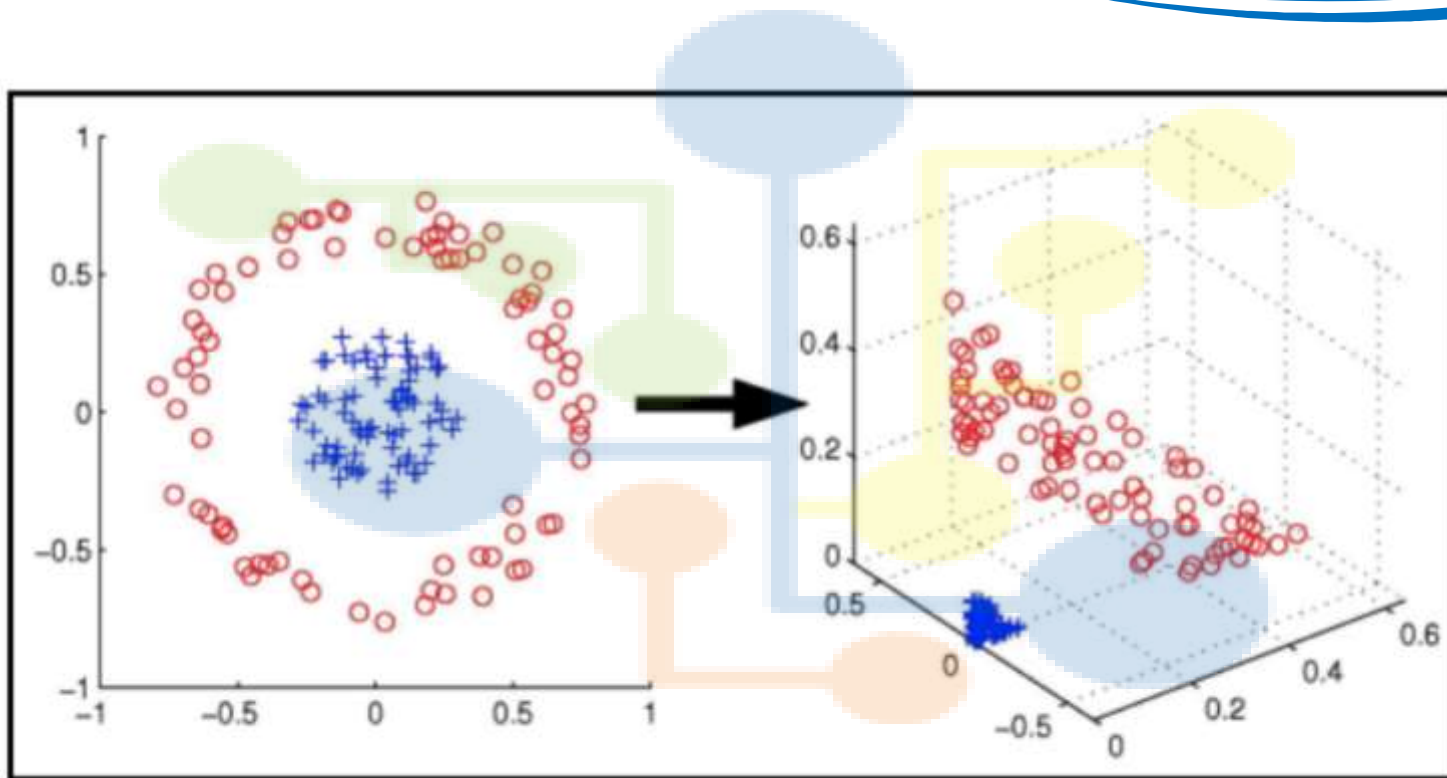
Rede de Hopfield





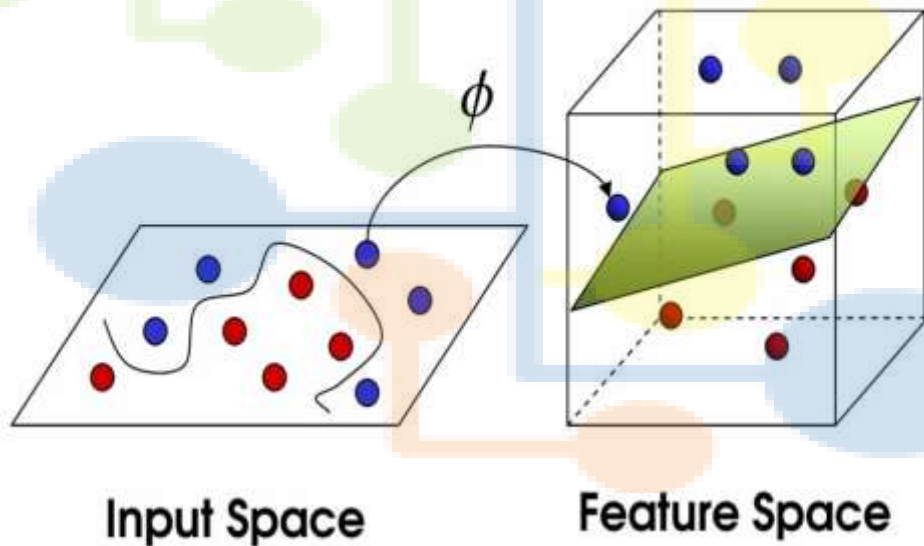
**Data Science
Academy**

Aprendizagem com SVMs



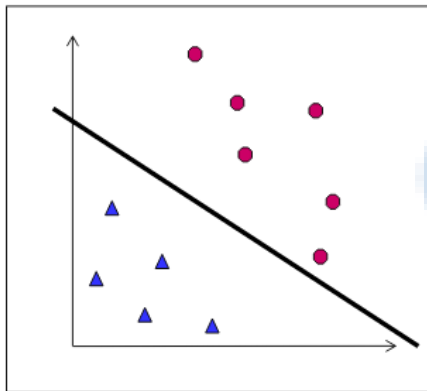


Principle of Support Vector Machines (SVM)

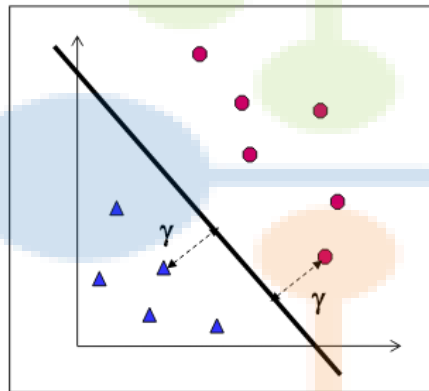




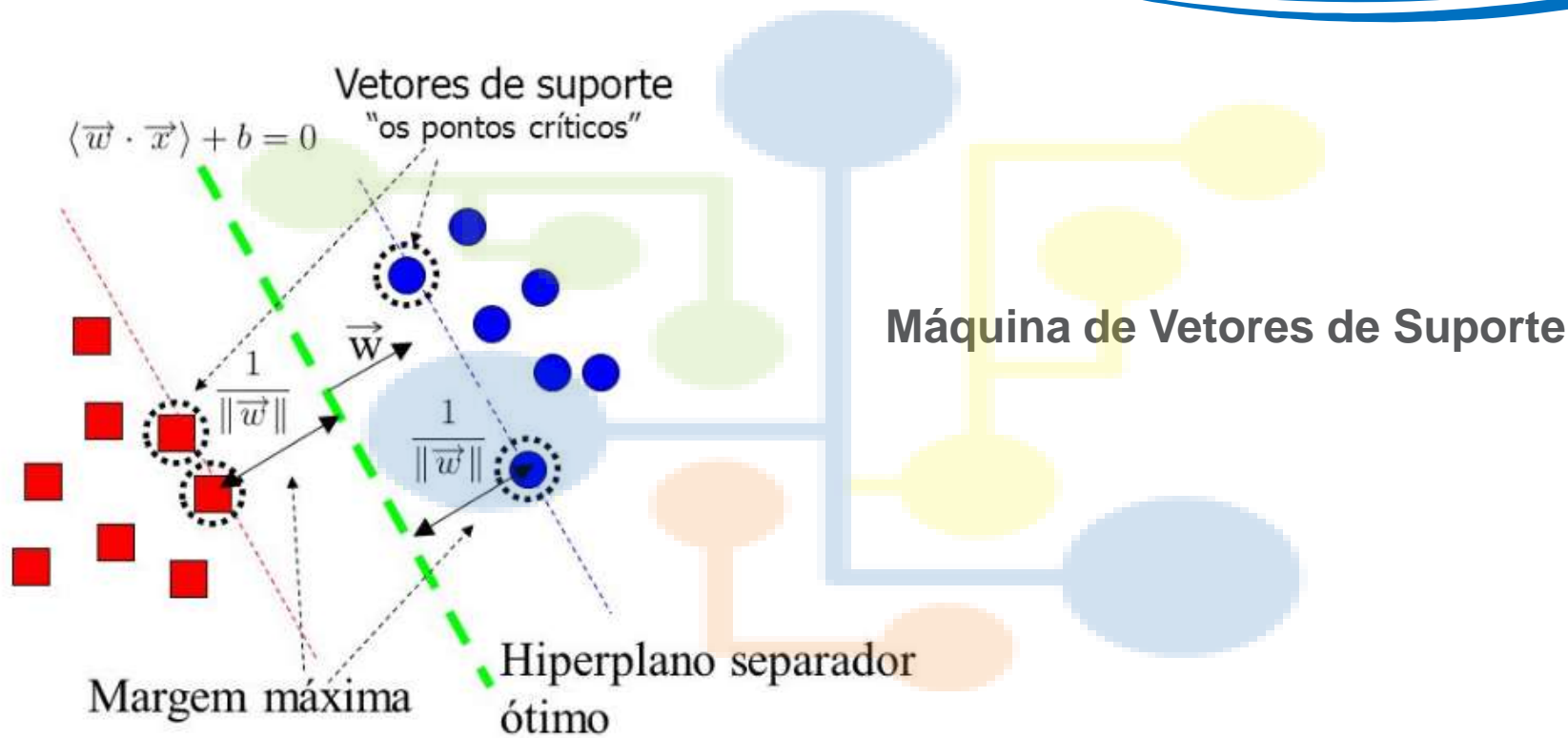
Rede Neural



SVMs



As máquinas de vetores de suporte são consideradas uma outra categoria das redes neurais.



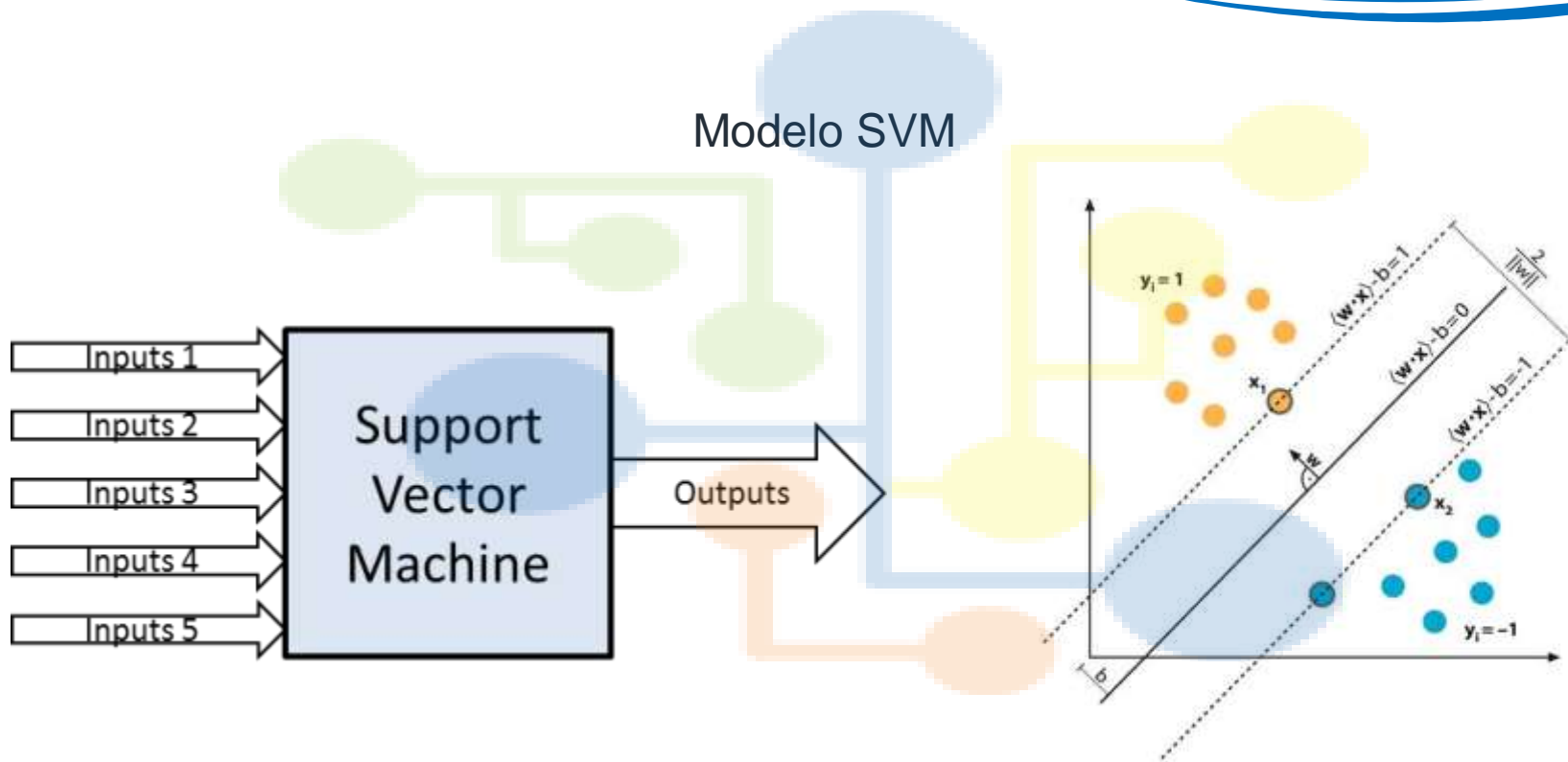


Vetores de suporte são simplesmente as coordenadas de observação individual.

Support Vector Machine é uma fronteira (hiperplano) que melhor segrega as duas classes.



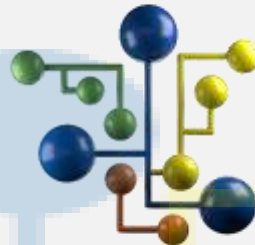
A SVM é uma técnica de aprendizado estatístico, baseada no princípio da Minimização do Risco Estrutural (SRM) e pode ser usada para resolver problemas de classificação ou de regressão.





Data Science
Academy

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



**Data Science
Academy**

Aprendizagem Não Supervisionada

Clusterização



Supervised Learning

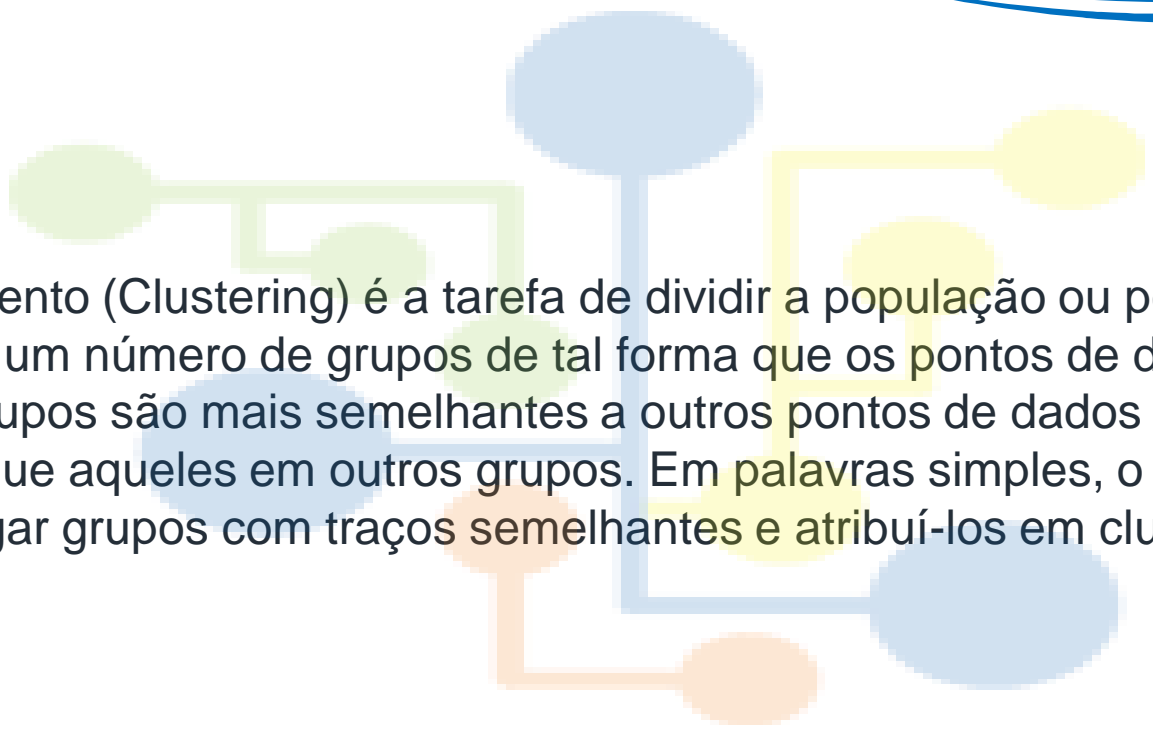


Unsupervised Learning

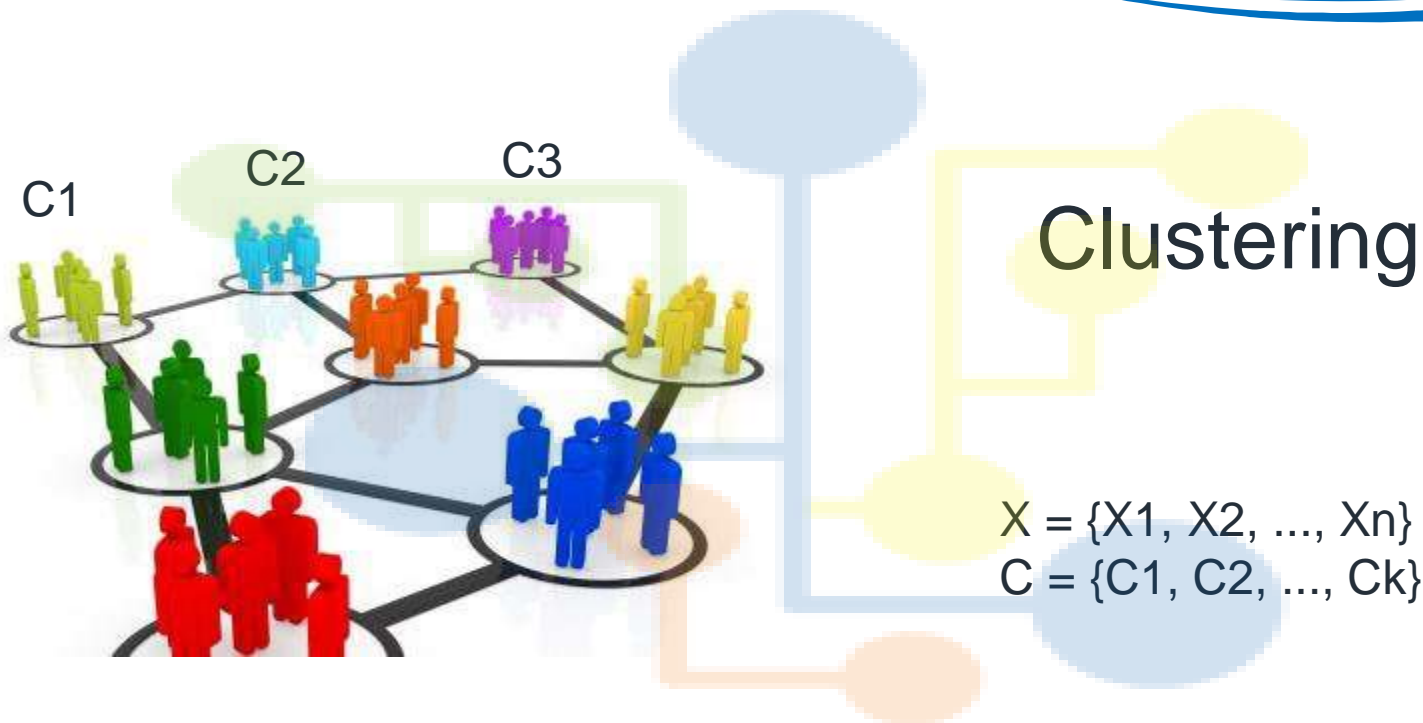


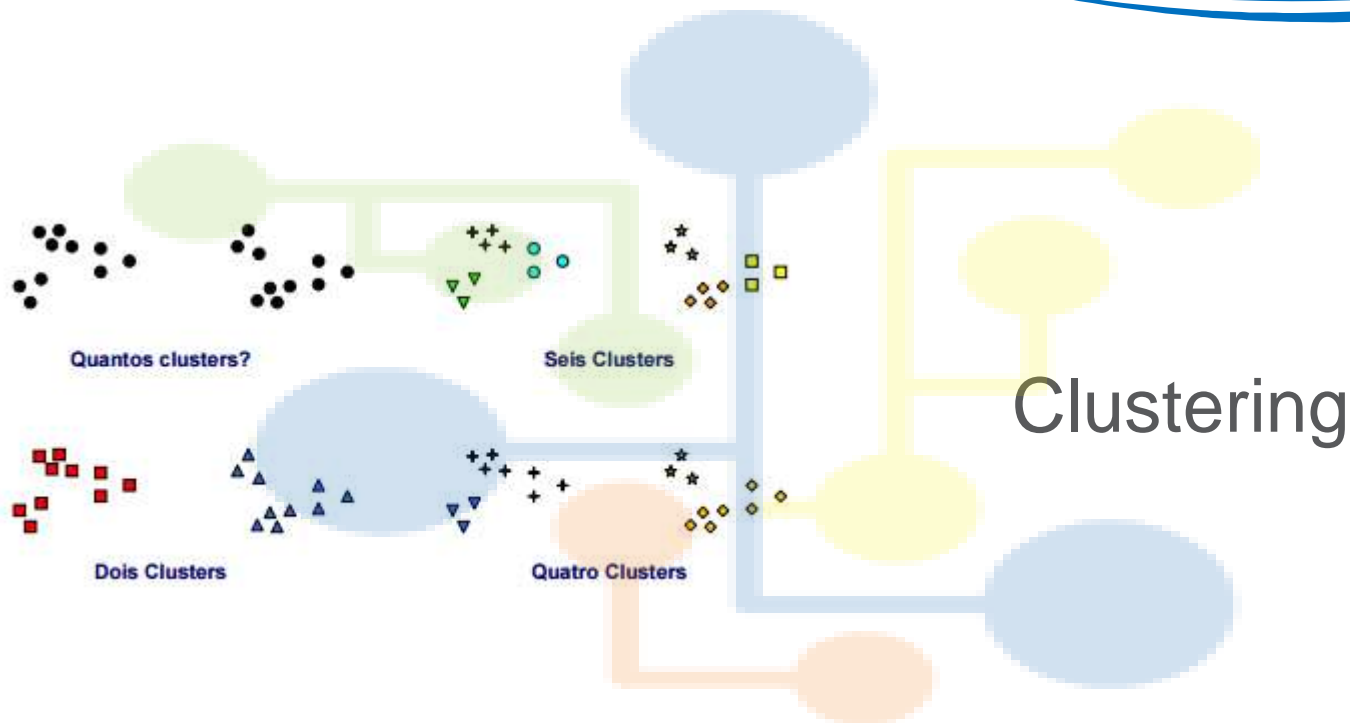


Agrupar todos os clientes de sua locadora de automóveis em 10 grupos com base em seus hábitos de compra e usar uma estratégia separada para os clientes em cada um desses 10 grupos. Isso é Clusterização.

A dendrogram illustrating the clustering process. It shows a hierarchical tree structure with nodes represented by colored circles (blue, green, yellow, orange) and lines connecting them. The tree starts from a single root at the bottom and branches out into several clusters. The clusters are color-coded: a large blue cluster, a green cluster, a yellow cluster, and an orange cluster. The lines connecting the nodes are also color-coded to match the clusters they represent.

Agrupamento (Clustering) é a tarefa de dividir a população ou pontos de dados em um número de grupos de tal forma que os pontos de dados nos mesmos grupos são mais semelhantes a outros pontos de dados no mesmo grupo do que aqueles em outros grupos. Em palavras simples, o objetivo é segregar grupos com traços semelhantes e atribuí-los em clusters.







Medidas de Distância

Distância euclidiana: considera a distância entre dois elementos X_i e X_j no espaço p -dimensional:

$$d(X_i, X_j) = \left[\sum_{l=1}^p (x_{il} - x_{jl})^2 \right]^{\frac{1}{2}}$$

Distância “city-block”: corresponde à soma das diferenças entre todos os p atributos de dois elementos X_i e X_j , não sendo indicada para os casos em que existe uma correlação entre tais atributos:

$$d(X_i, X_j) = \sum_{l=1}^p |x_{il} - x_{jl}|$$



Normalização \Rightarrow variáveis com mesmo peso.

- **Min-Max para um atributo f :**

$$s_{if} = \frac{x_{if} - \min_f}{\max_f - \min_f} \times (\text{novoMax} - \text{novoMin}) + \text{novoMin}$$

- **Z-score** $z_{if} = \frac{x_{if} - m_f}{\sigma_f}$

- **Desvio absoluto médio**

$$s_f = \frac{1}{n} (|x_{1f} - m_f| + |x_{2f} - m_f| + \dots + |x_{nf} - m_f|)$$

Clustering



Tipos de Clustering

Hard Clustering

Soft Clustering



Modelos de Conectividade

Modelos Centroides

Modelos de Distribuição

Modelos de Densidade

Tipos de Algoritmos de Clustering

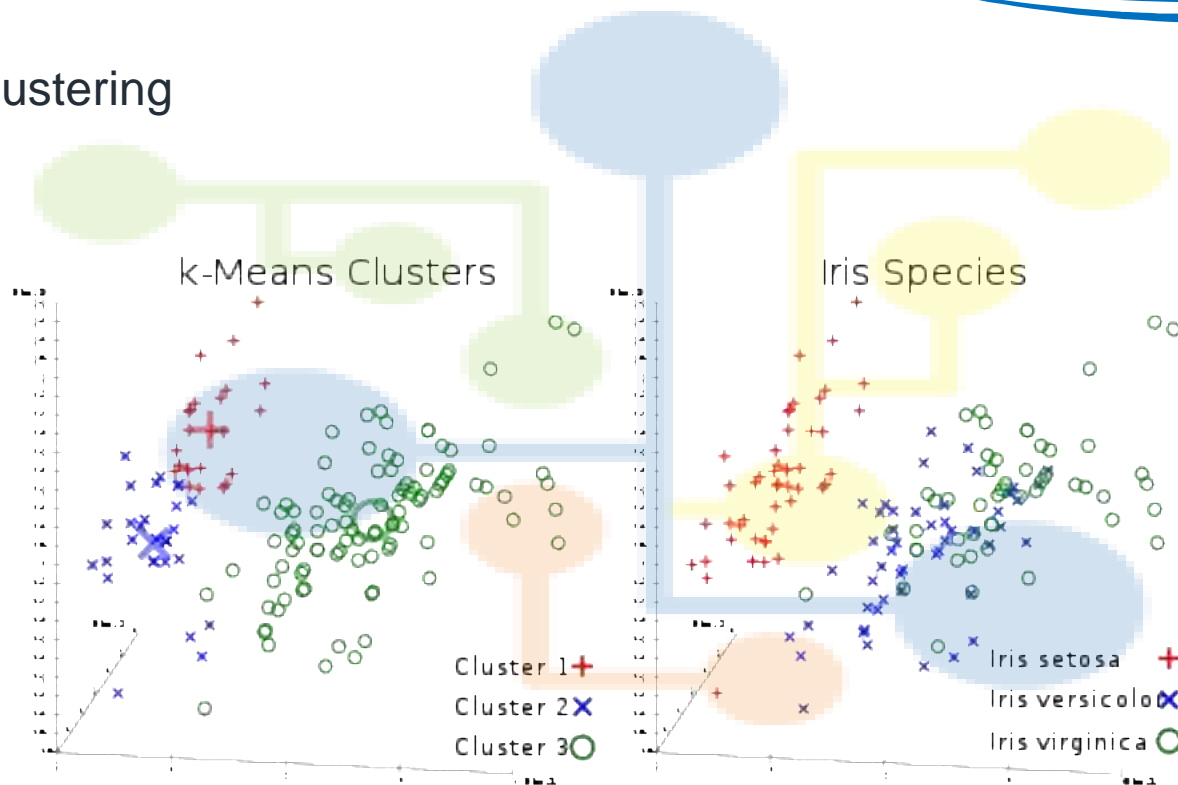


Métodos Utilizados para Clusterização

As heurísticas existentes para a solução de problemas de clusterização podem ser classificadas, de forma geral, em métodos hierárquicos e métodos de particionamento.

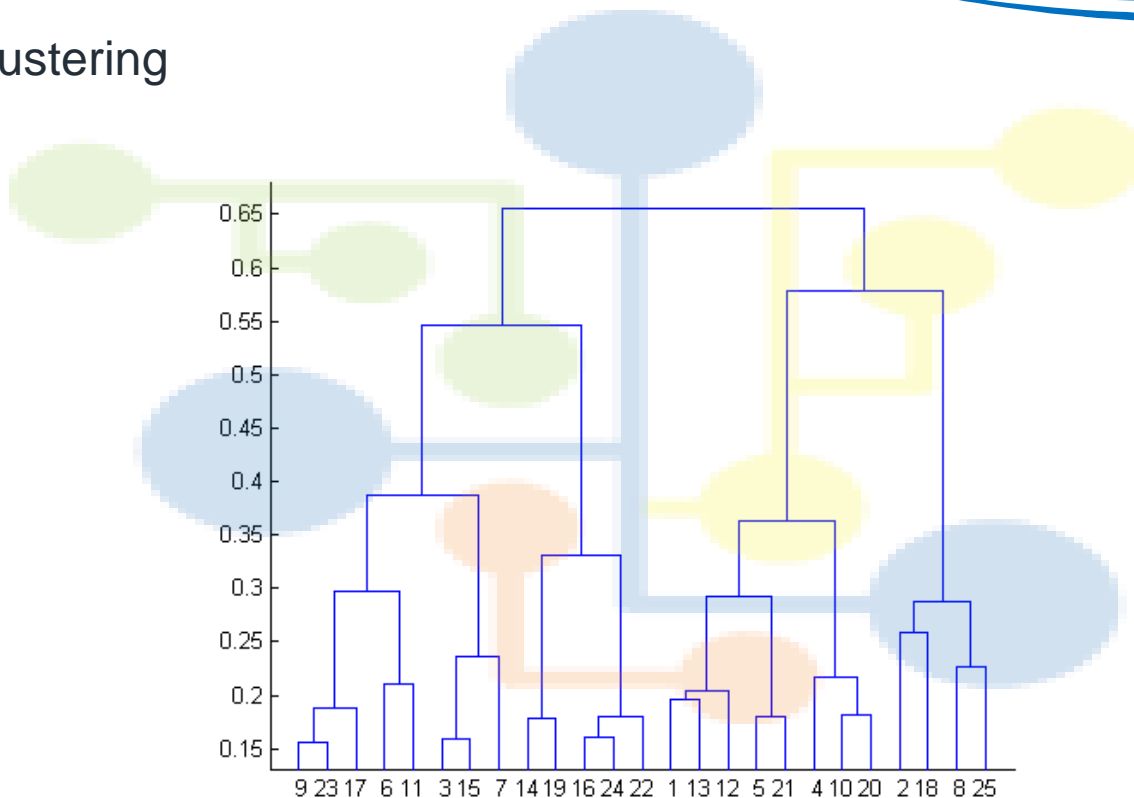


K-Means Clustering





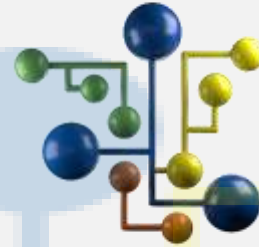
Hierarchical Clustering





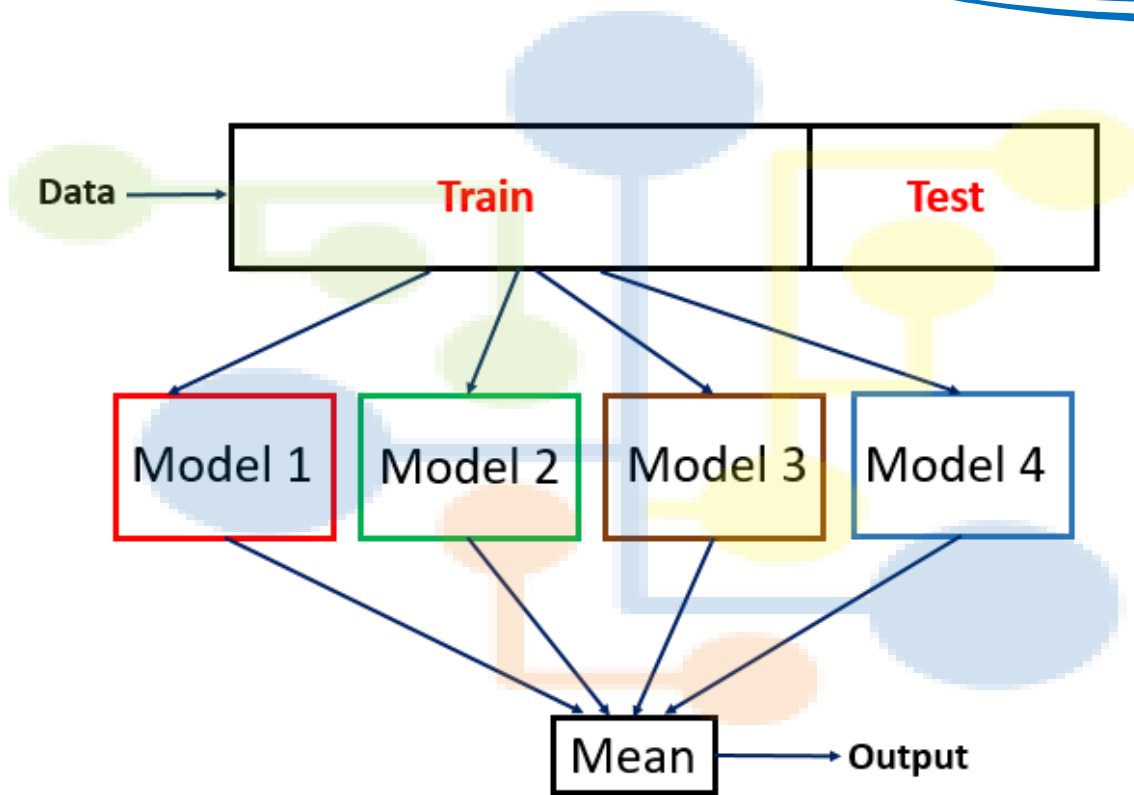
Data Science
Academy

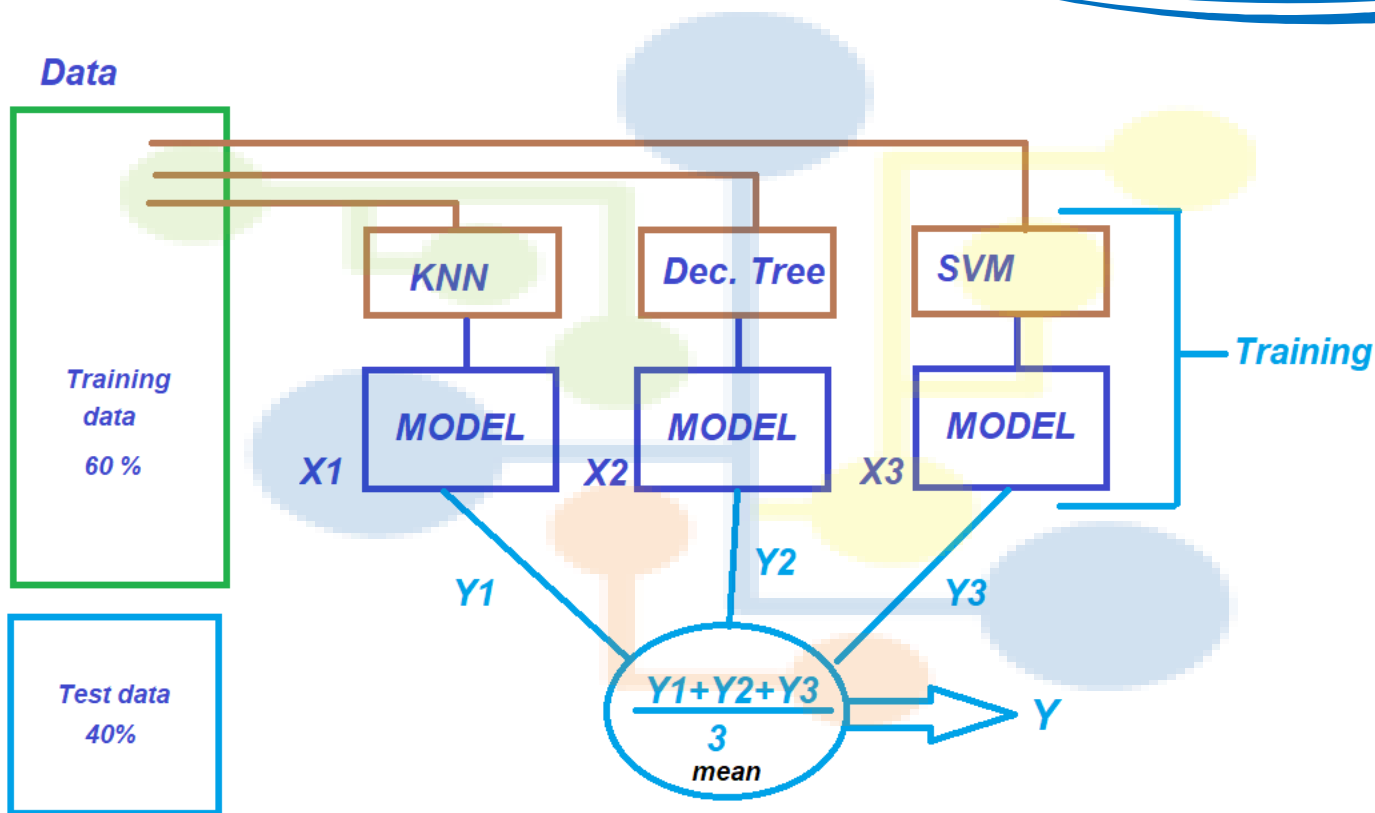
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

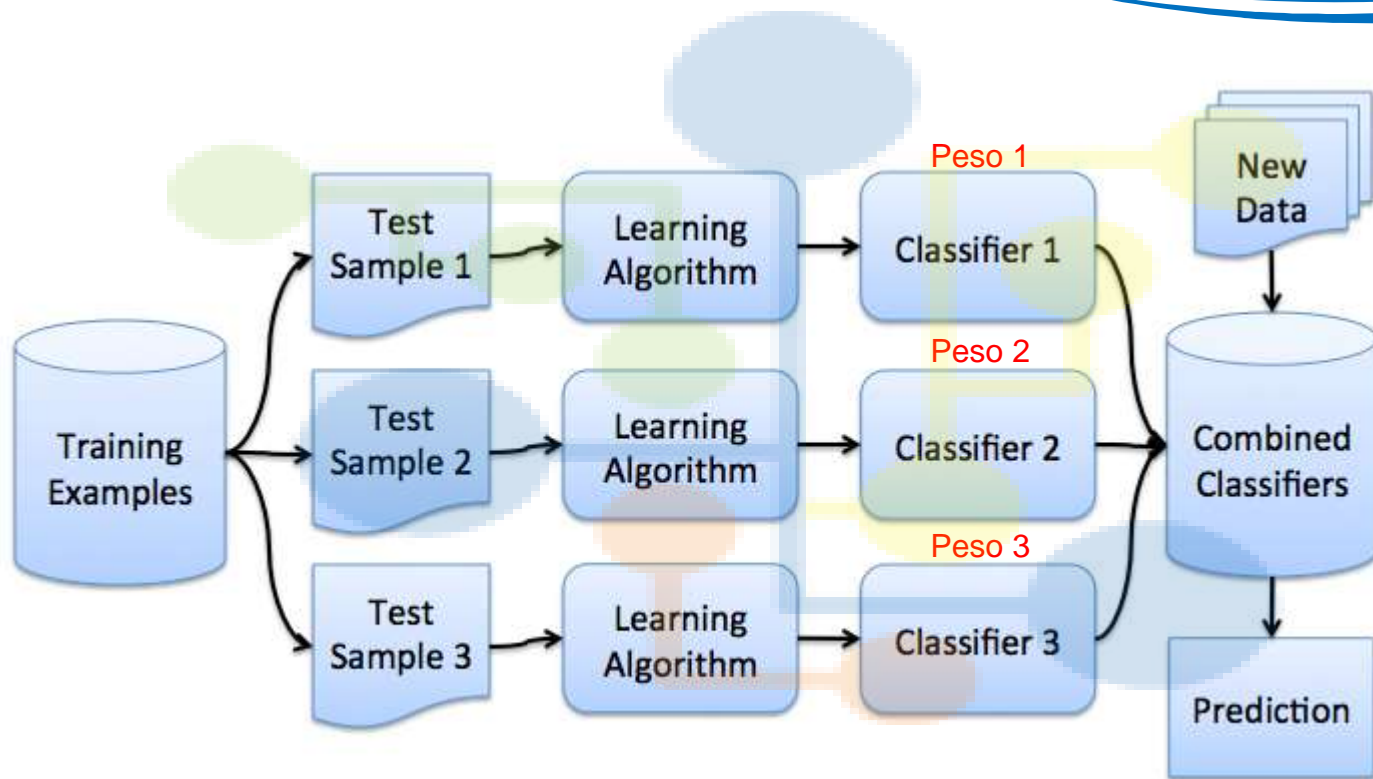


**Data Science
Academy**

Aprendizagem com Métodos Ensemble



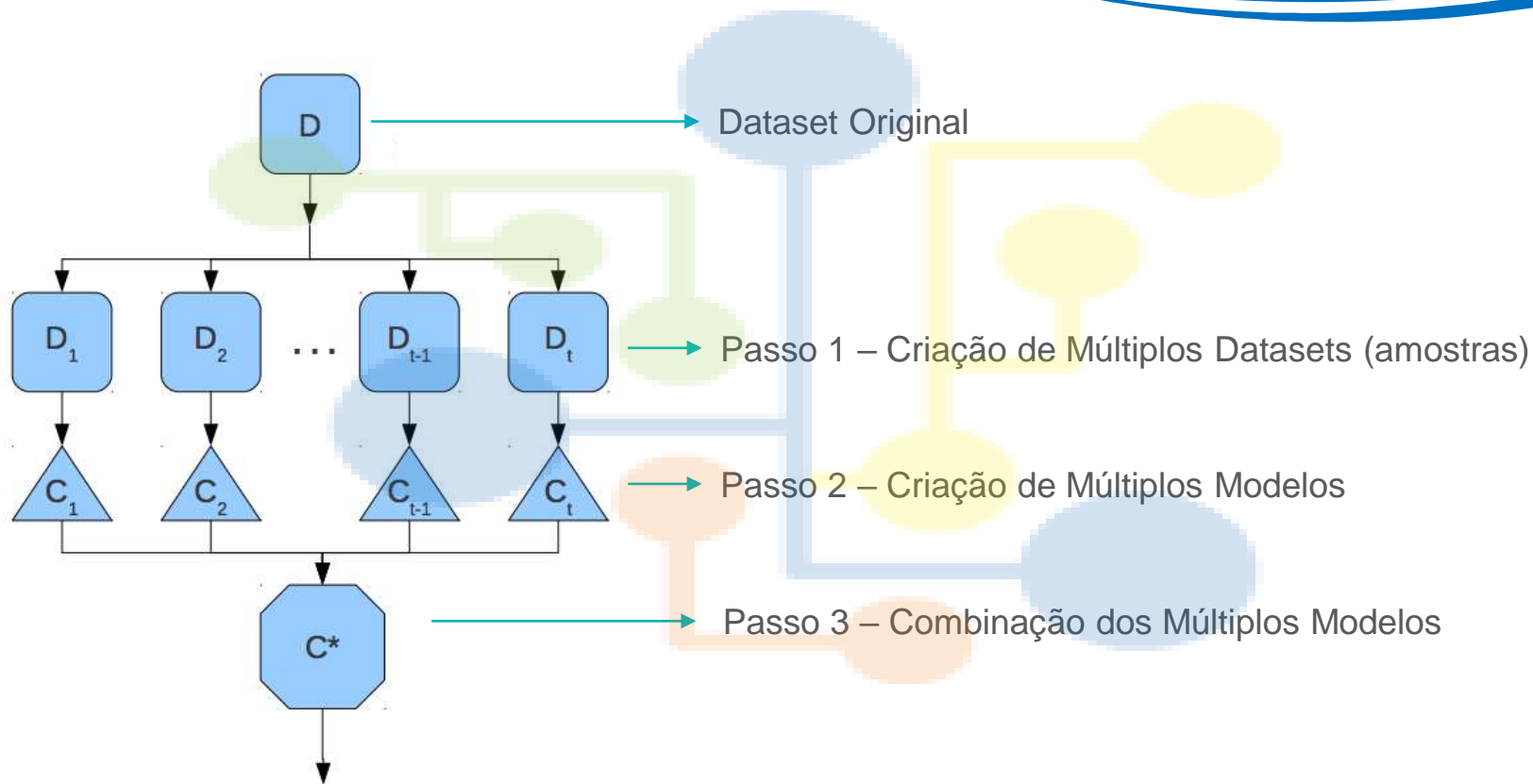






Isso é Método Ensemble

Unimos as saídas de diferentes modelos para encontrar a melhor resposta para o problema.





Temos duas abordagens principais com Métodos Ensemble:

Bootstrap Aggregation ou Bagging

- Bagged CART
- Random Forest

Boosting

- C5.0
- Stochastic Gradient Boosting
- AdaBoost



Data Science
Academy

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



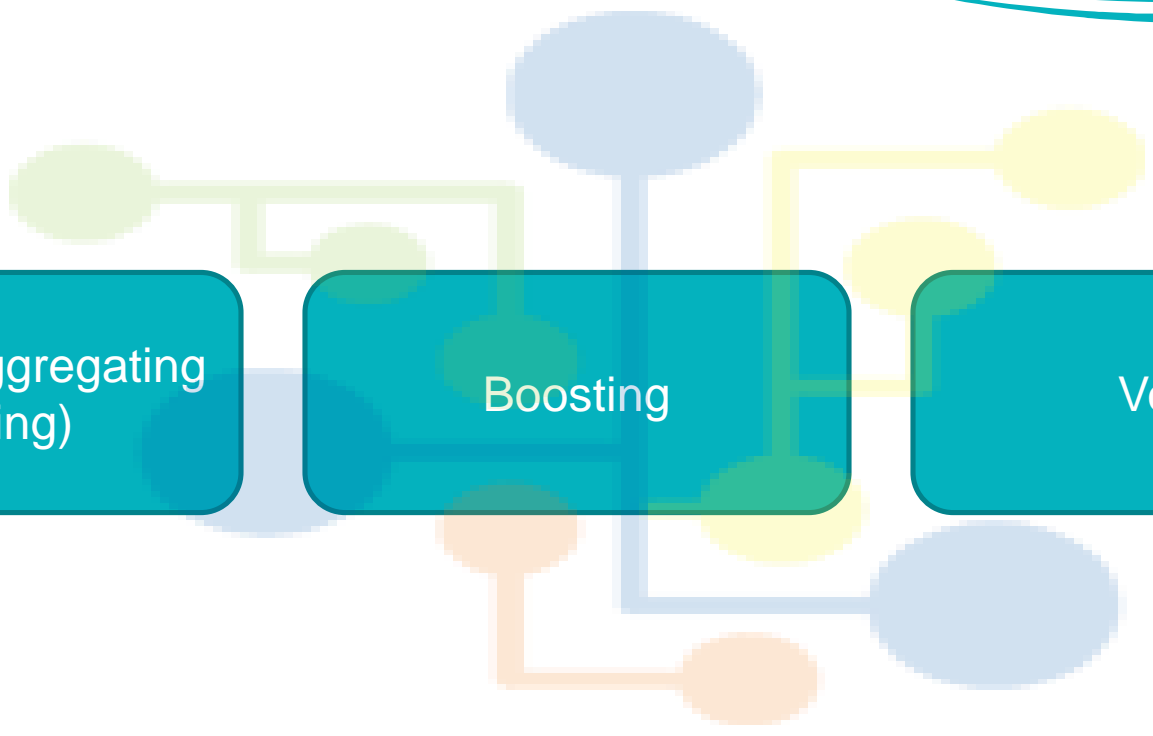
Data Science Academy

Métodos Ensemble – Parte II



Combinação de Preditores





Bootstrap Aggregating
(Bagging)

Boosting

Voting



Bootstrap Aggregating (Bagging)

Para construção de múltiplos modelos (normalmente do mesmo tipo) a partir de diferentes subsets no dataset de treino.





Boosting

Para construção de múltiplos modelos (normalmente do mesmo tipo), onde cada modelo aprende a corrigir os erros gerados pelo modelo anterior, dentro da sequência de modelos criados.

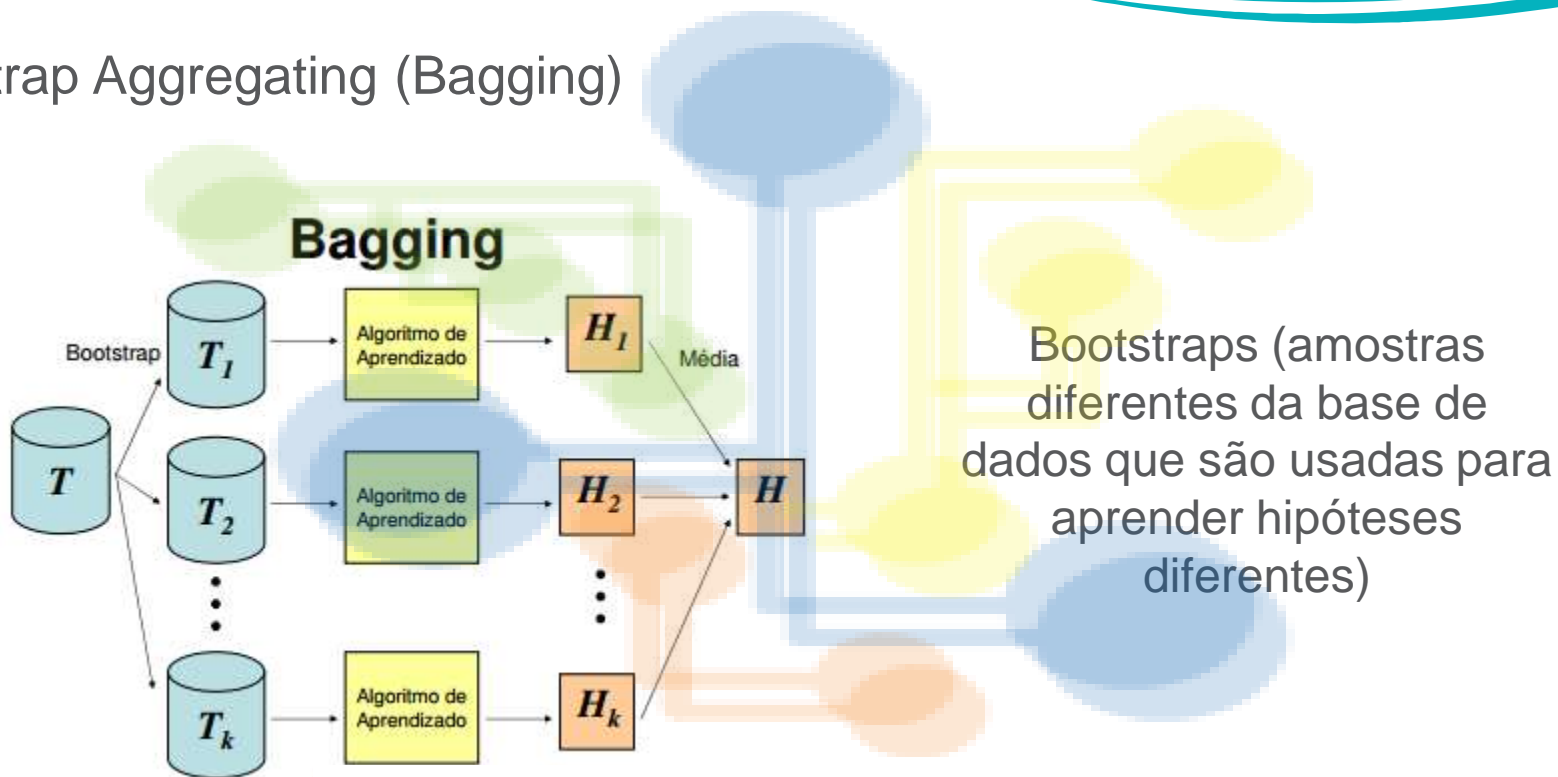


Voting

Para construção de múltiplos modelos (normalmente de tipos diferentes) e estatísticas simples (como a média) são usadas para combinar as previsões.



Bootstrap Aggregating (Bagging)





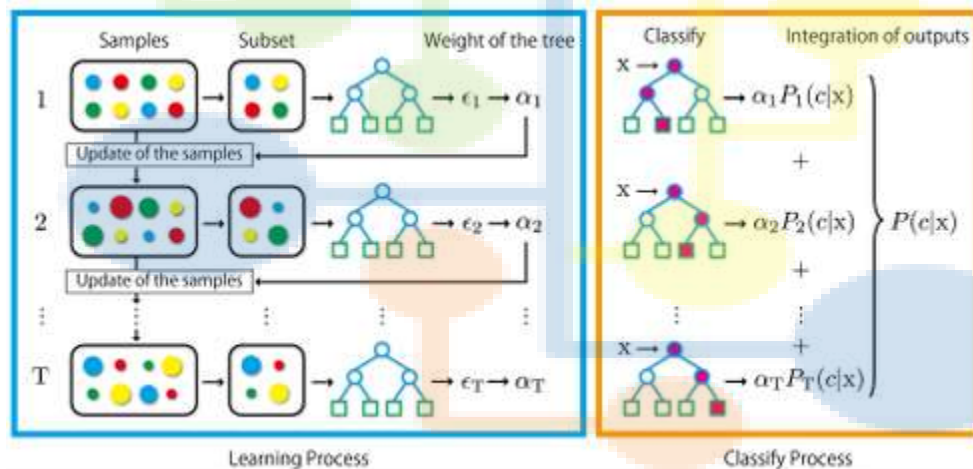
Boosting

- AdaBoost (**Adaptive Boosting**)
- Gradient Boosting
- XGBoost

O termo "Boosting" refere-se a uma família de algoritmos que converte modelos fracos em um modelo forte

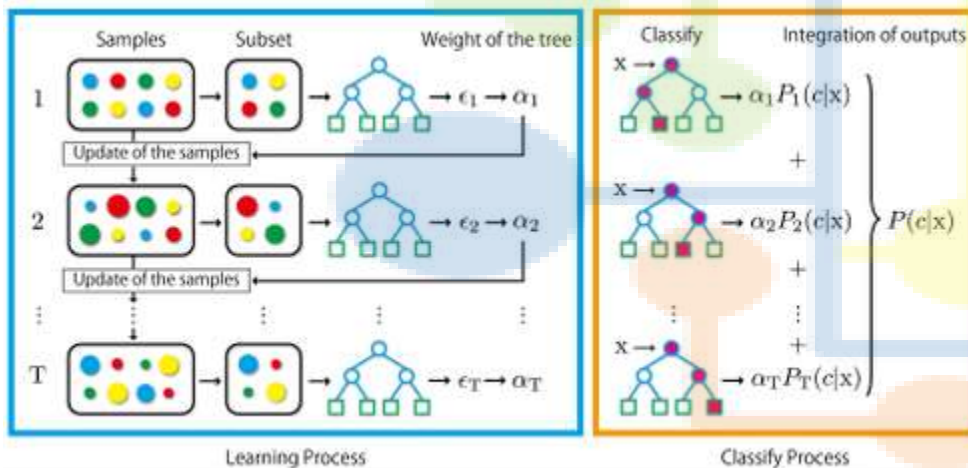


Boosting





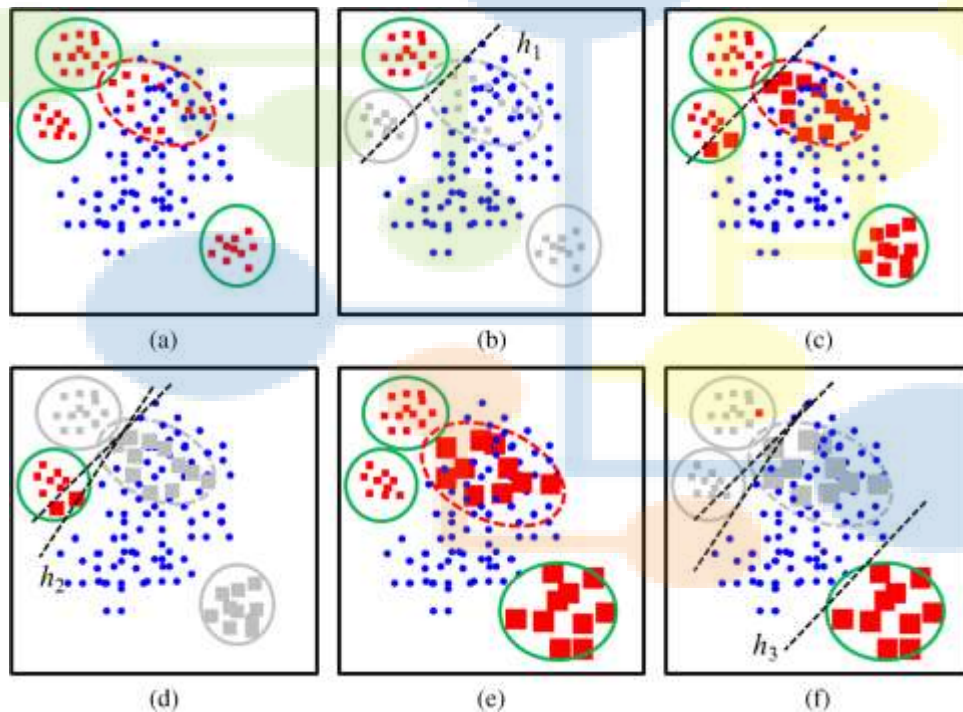
Boosting



- AdaBoost.M1
- AdaBoost.M2
- AdaBoost.R
- Adaboost.R2
- AdaBoost.RT
- Boosting Correlation Improvement (BCI)

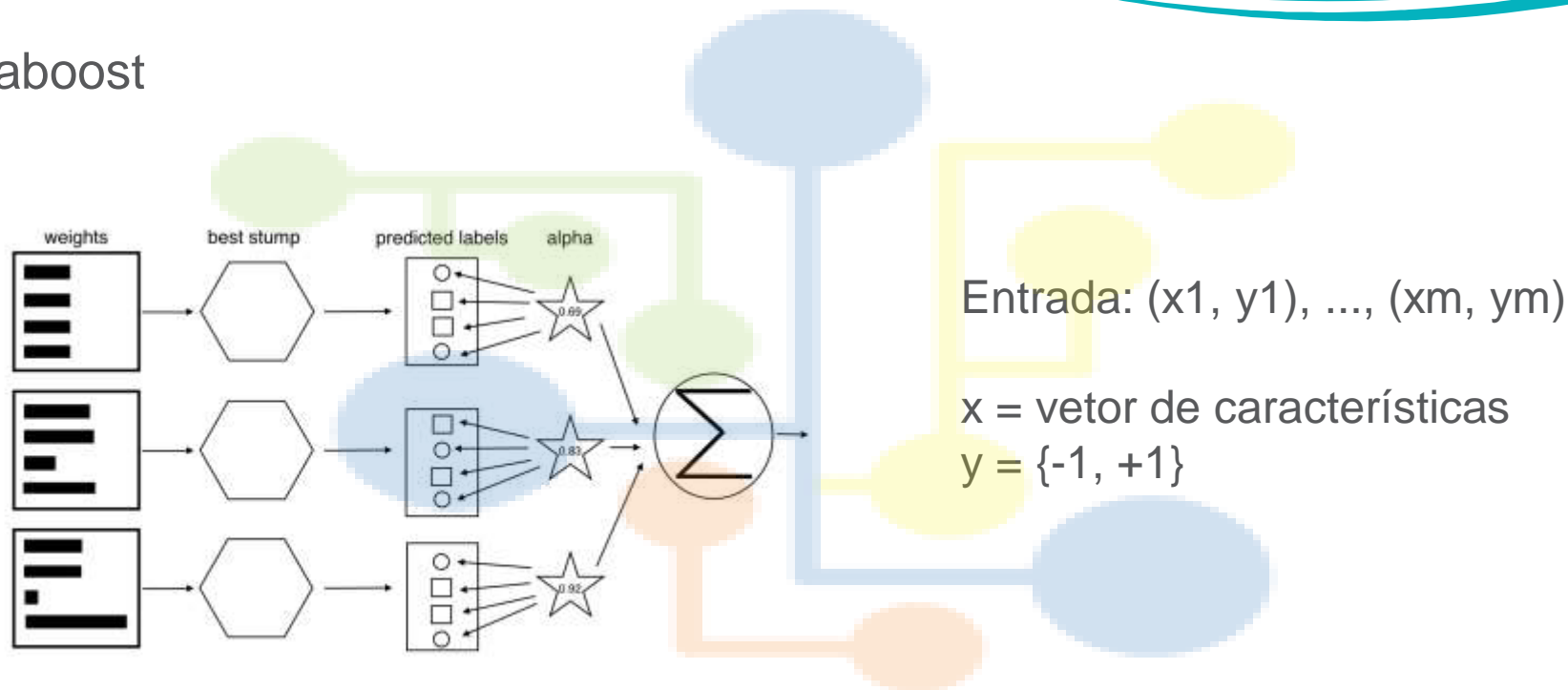


Adaboost



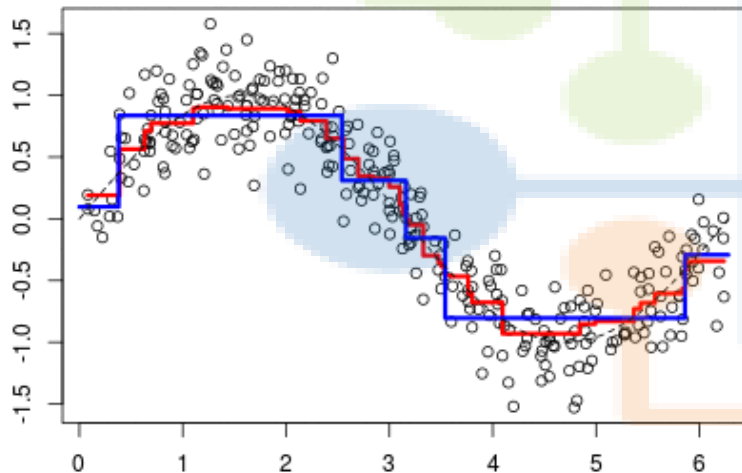


Adaboost





Boosting para Problemas de Regressão – AdaBoost.R



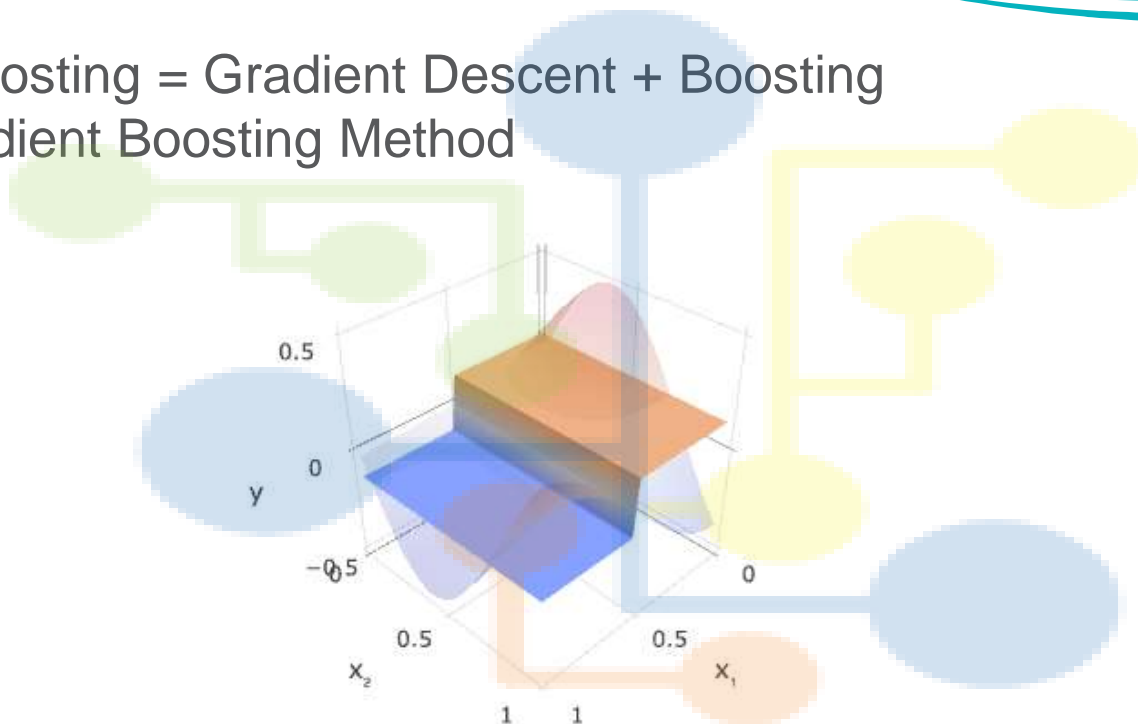
$f(x)$
 $g(x)$

$$f(x_i) = g(x_i), \forall x_i \in X$$



Gradient Boosting = Gradient Descent + Boosting

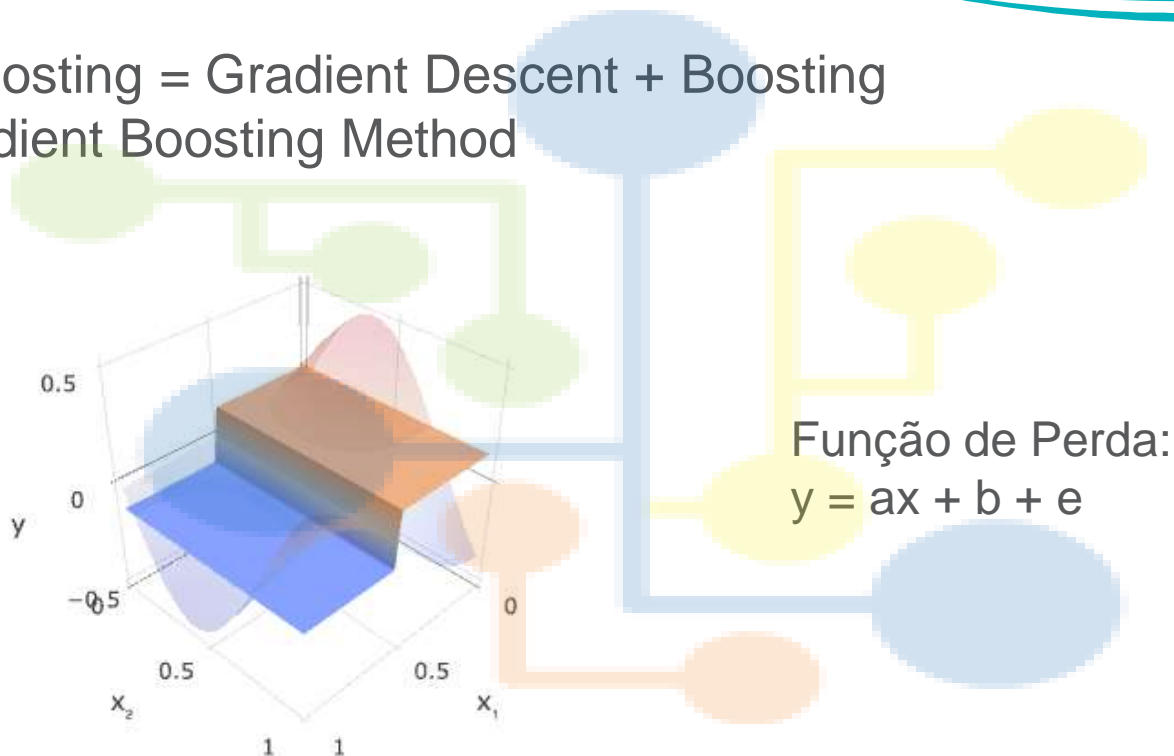
GBM = Gradient Boosting Method





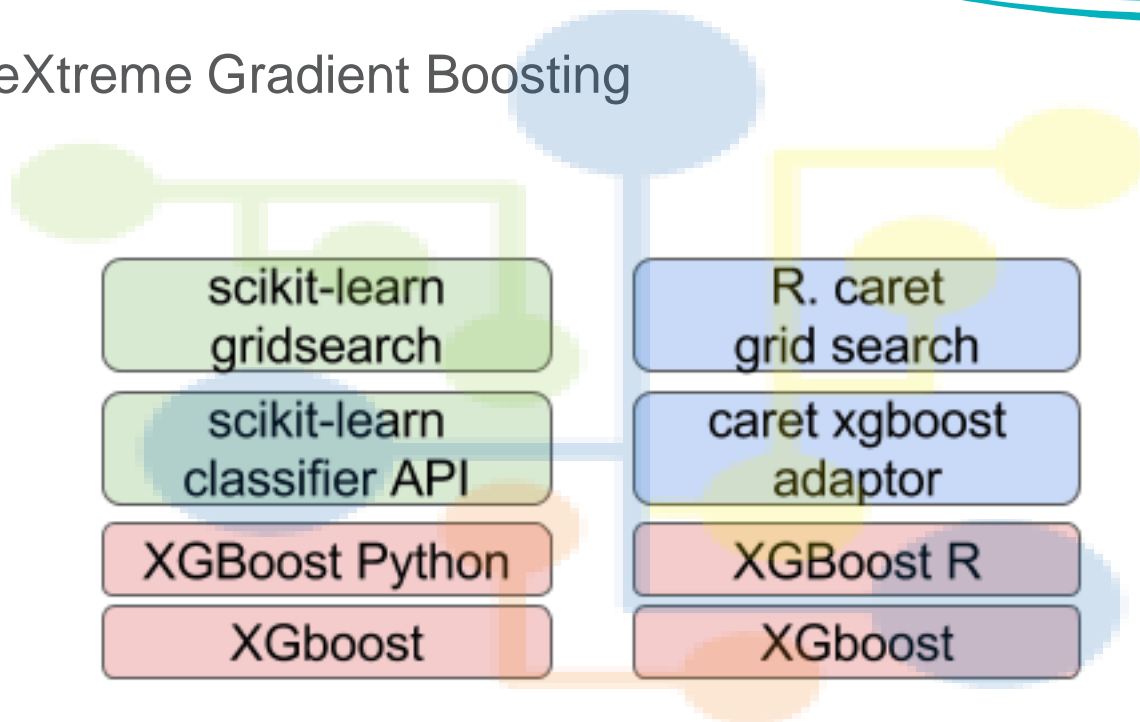
Gradient Boosting = Gradient Descent + Boosting

GBM = Gradient Boosting Method





XGBoost – eXtreme Gradient Boosting





A background diagram illustrating ensemble methods. It features a central vertical blue line with several horizontal branches. To the left, a green branch leads to a green circle. To the right, a yellow branch leads to a yellow circle. Below the central line, an orange branch leads to an orange circle. The diagram represents the combination of multiple models (represented by colored circles) to form a single ensemble model (represented by the central blue line).

Por que utilizar Métodos Ensemble?



Por que utilizar Métodos Ensemble?

Razões Estatísticas

Grandes volumes de dados

Pequenos volumes de dados



Por que utilizar Métodos Ensemble?

Razões Estatísticas

Dividir e Conquistar

Grandes volumes de dados

Seleção de modelo

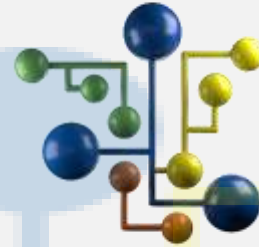
Pequenos volumes de dados

Diversidade



Data Science
Academy

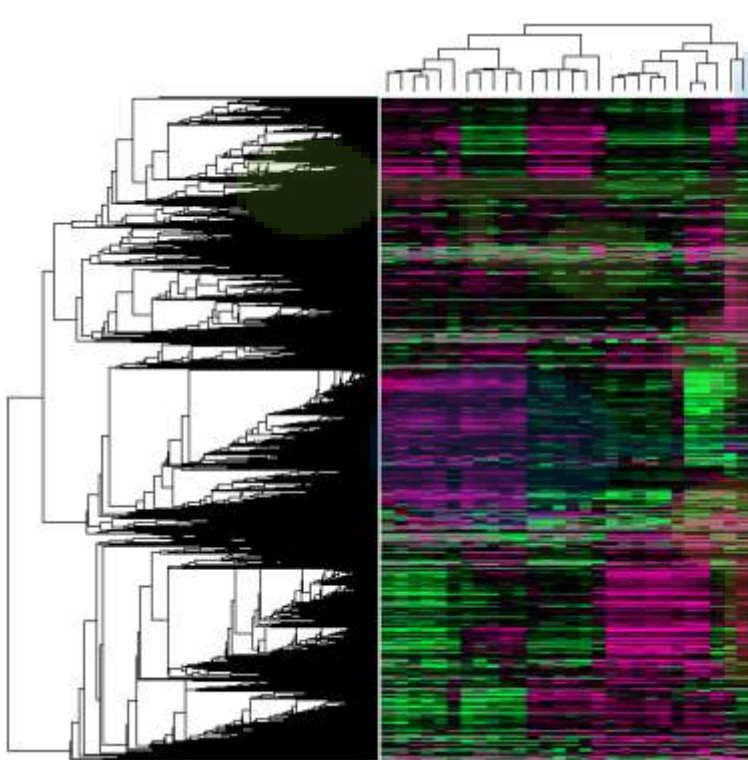
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Redução de Dimensionalidade

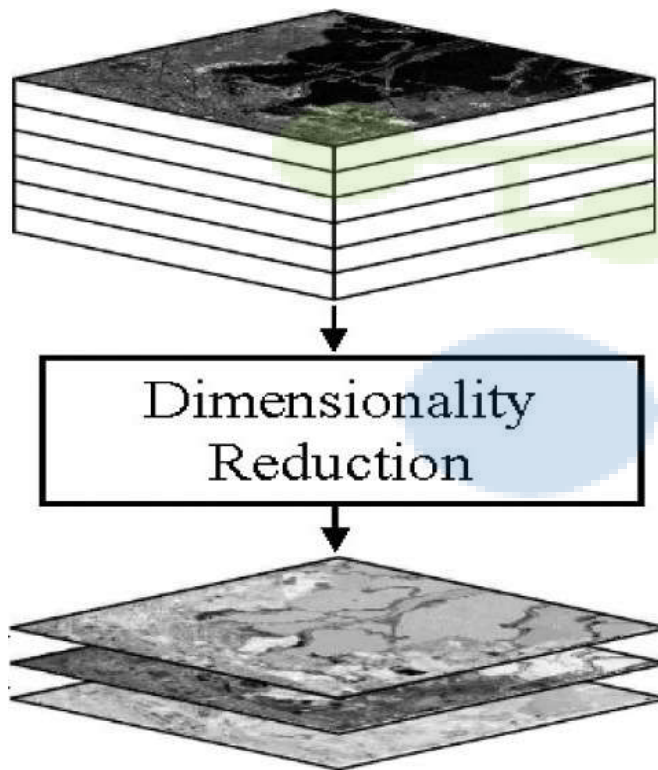




E esse aumento no volume de dados não é apenas em volume e de forma vertical, mas ocorre também na horizontal, com o aumento do número de dimensões ou atributos das bases de dados.



O termo ***dimensionalidade*** é atribuído ao número de características de uma representação de padrões, ou seja, a dimensão do espaço de características.



Extração de Atributos

Seleção de Atributos



Extração de Atributos

Principal Component Analysis, Multidimensional Scaling e o FastMap.

Seleção de Atributos

Algoritmos de aprendizado de máquina (Random Forest), cálculo de dimensão fractal e wrapper.



7 Técnicas para Redução da Dimensionalidade

Missing Values Ratio

Low Variance Filter

High Correlation Filter

**Random Forests /
Ensemble Trees**

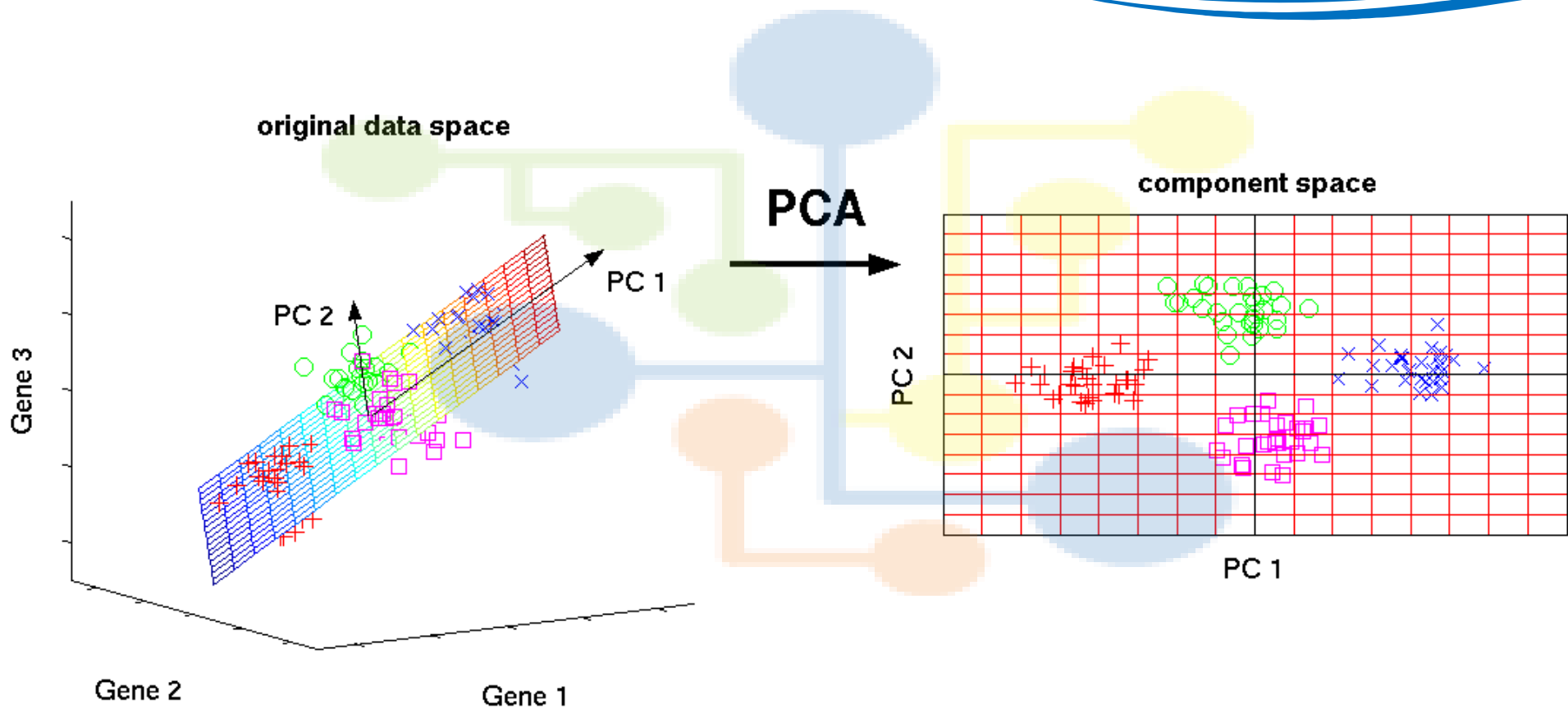
**Forward Feature
Construction**

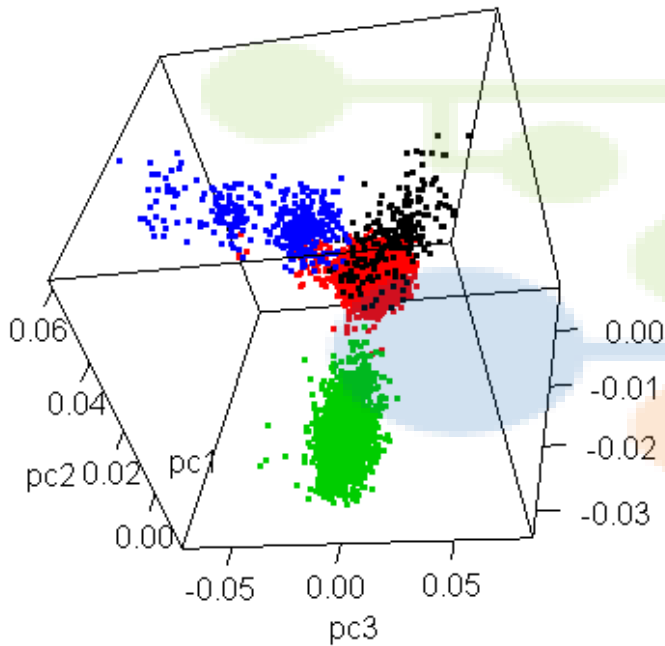
**Backward Feature
Elimination**

**Principal Component
Analysis (PCA)**

A decorative background diagram consisting of a central vertical blue line. From this line, several horizontal and vertical lines branch out to connect to colored circles. There are four blue circles, three green circles, three yellow circles, and two orange circles, all connected by lines of their respective colors. The circles are of varying sizes and are distributed across the slide, creating a network-like structure.

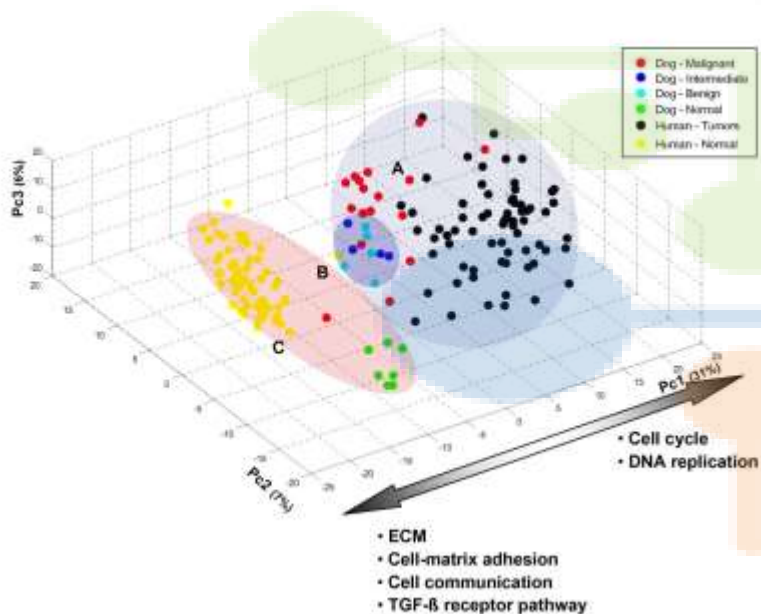
Principal Component Analysis (PCA)



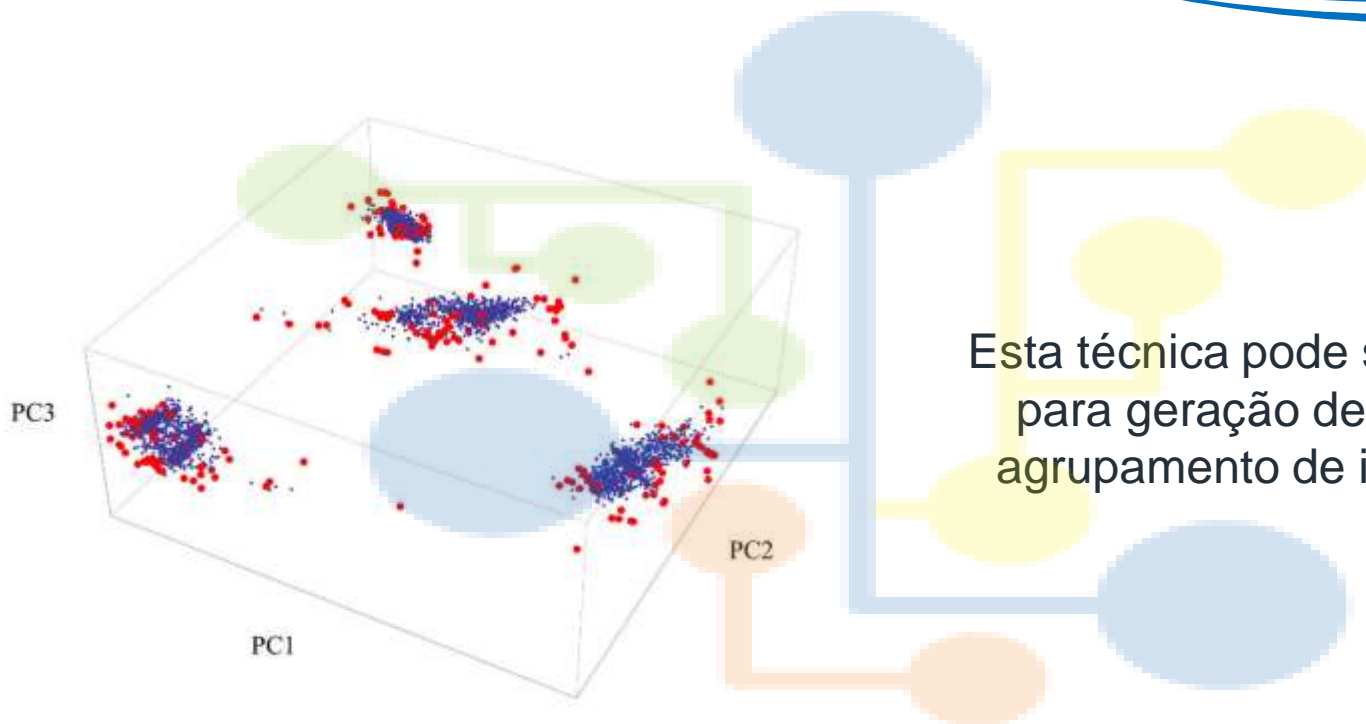


Cada componente resultante é uma combinação linear de n atributos.

Cada componente principal é uma combinação de atributos presentes no dataset



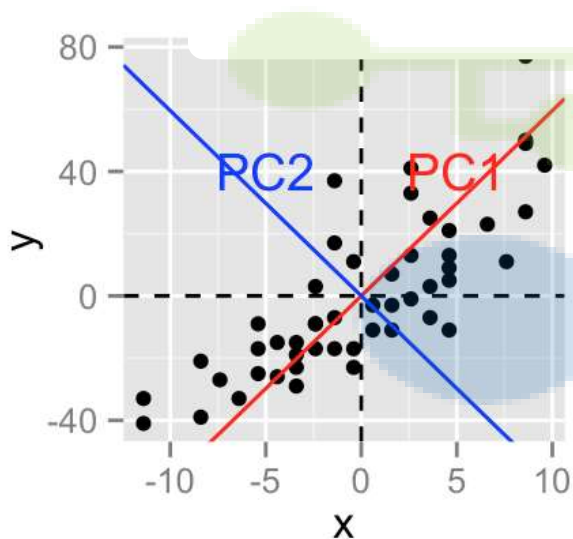
O PCA precisa ser alimentado com dados normalizados. Utilizar o PCA em dados não normalizados pode gerar resultados inesperados.



Esta técnica pode ser utilizada
para geração de índices e
agrupamento de indivíduos.



A análise de componentes principais é associada à ideia de redução de massa de dados, com menor perda possível da informação.



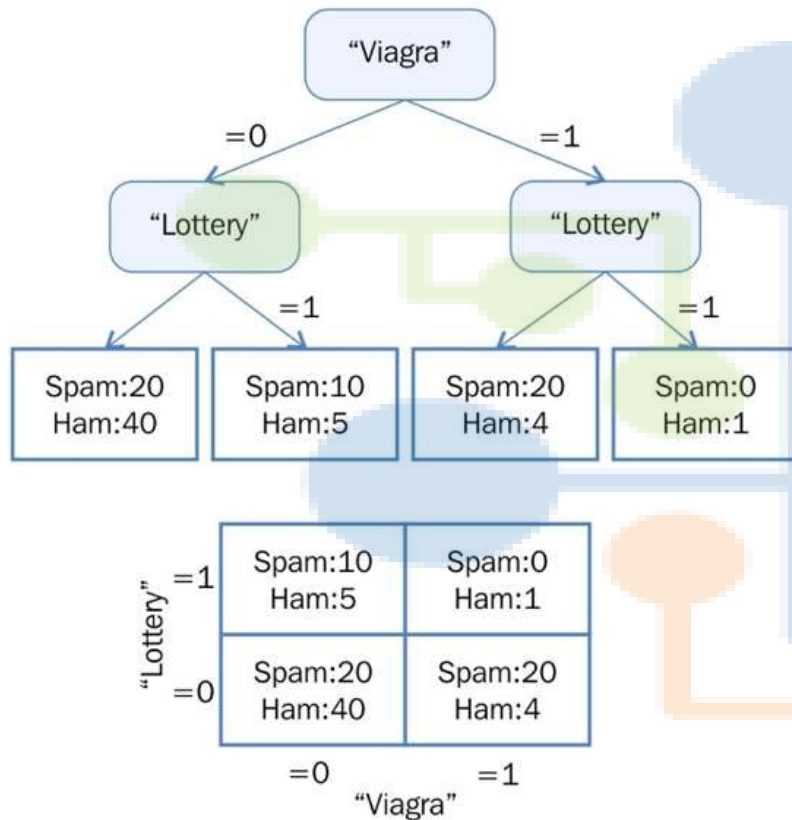
Em termos gerais a PCA busca reduzir o número de dimensões de um dataset, projetando os dados em um novo plano.

An abstract background diagram consisting of several interconnected nodes and lines in light blue, green, yellow, and orange. Three teal boxes with white text are overlaid on this diagram.

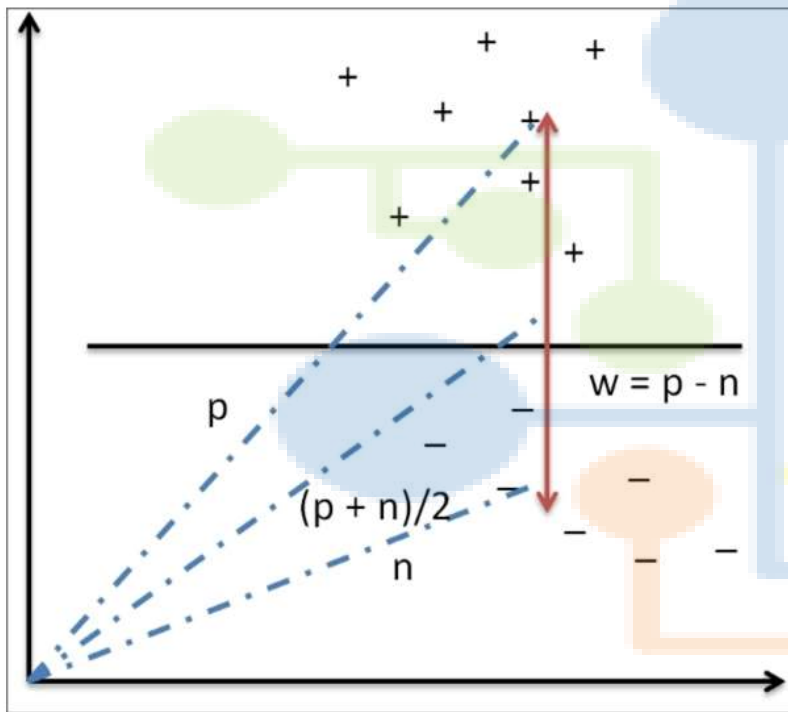
Lógicos

Geométricos

Probabilísticos



Lógicos

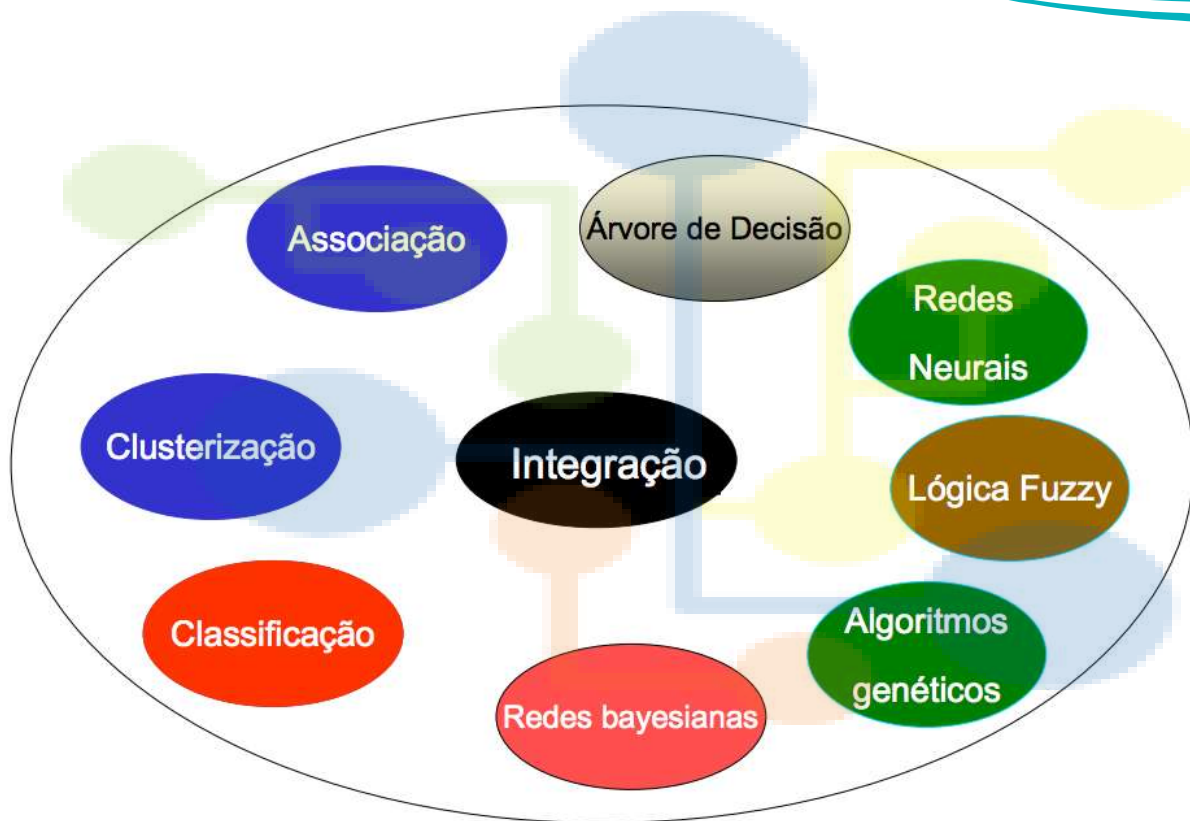


Geométricos



Viagra	Lottery	$P(Y = \text{Spam} (\text{Viagra}, \text{lottery}))$	$P(Y = \text{ham} (\text{Viagra}, \text{lottery}))$
0	0	0.31	0.69
0	1	0.65	0.35
1	0	0.80	0.20
1	1	0.40	0.60

Probabilísticos





Data Science
Academy

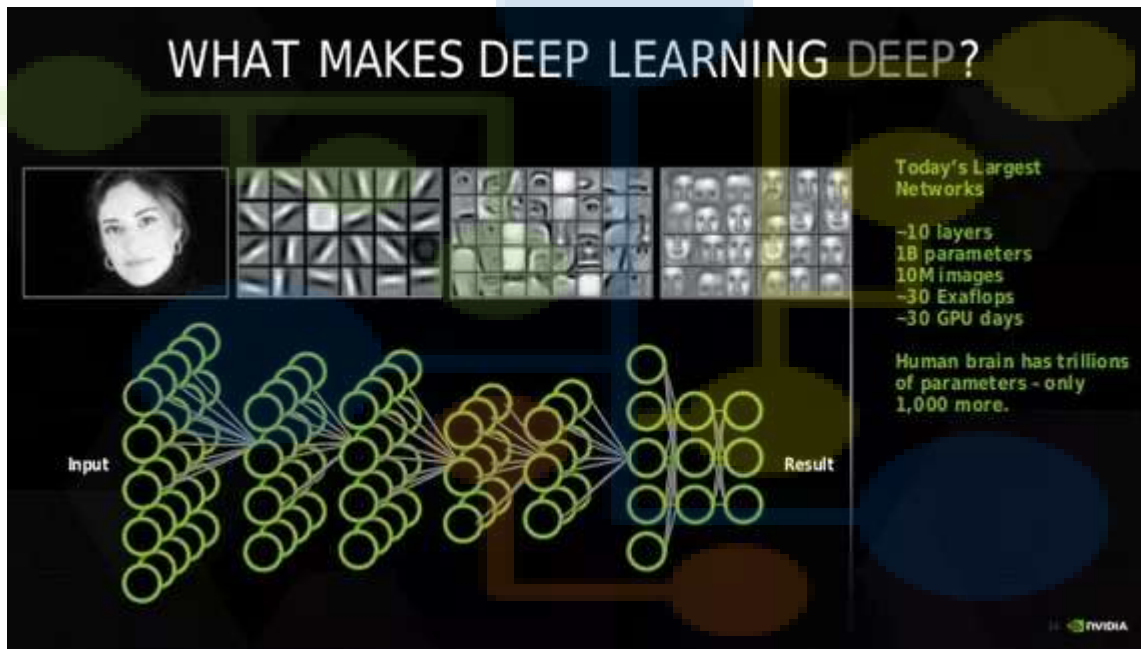
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

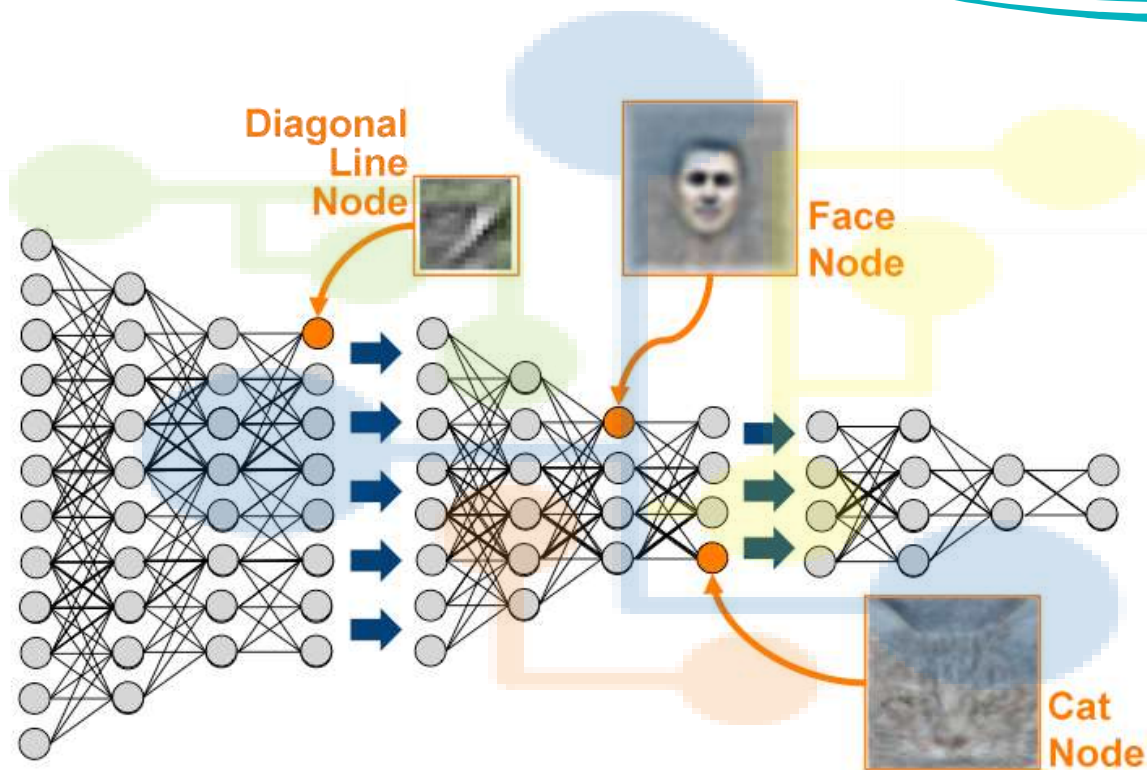


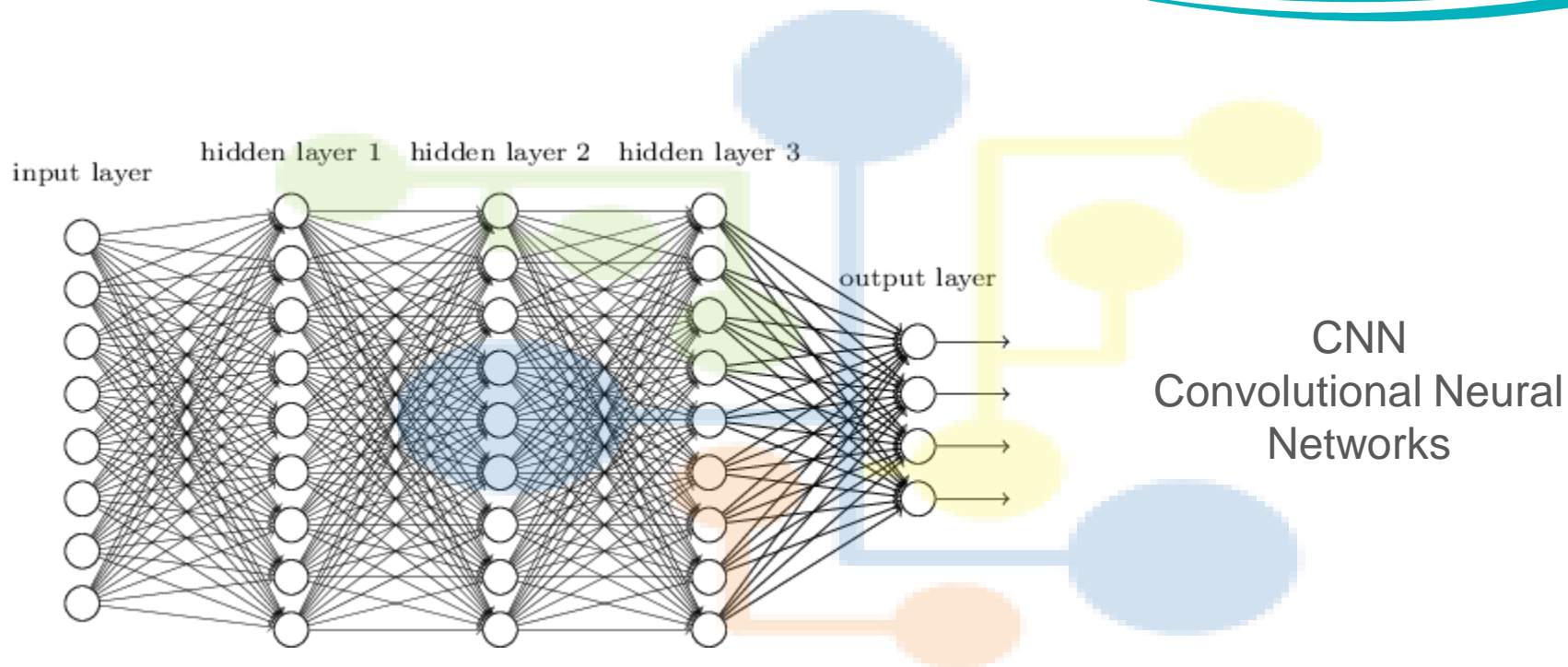
Data Science Academy

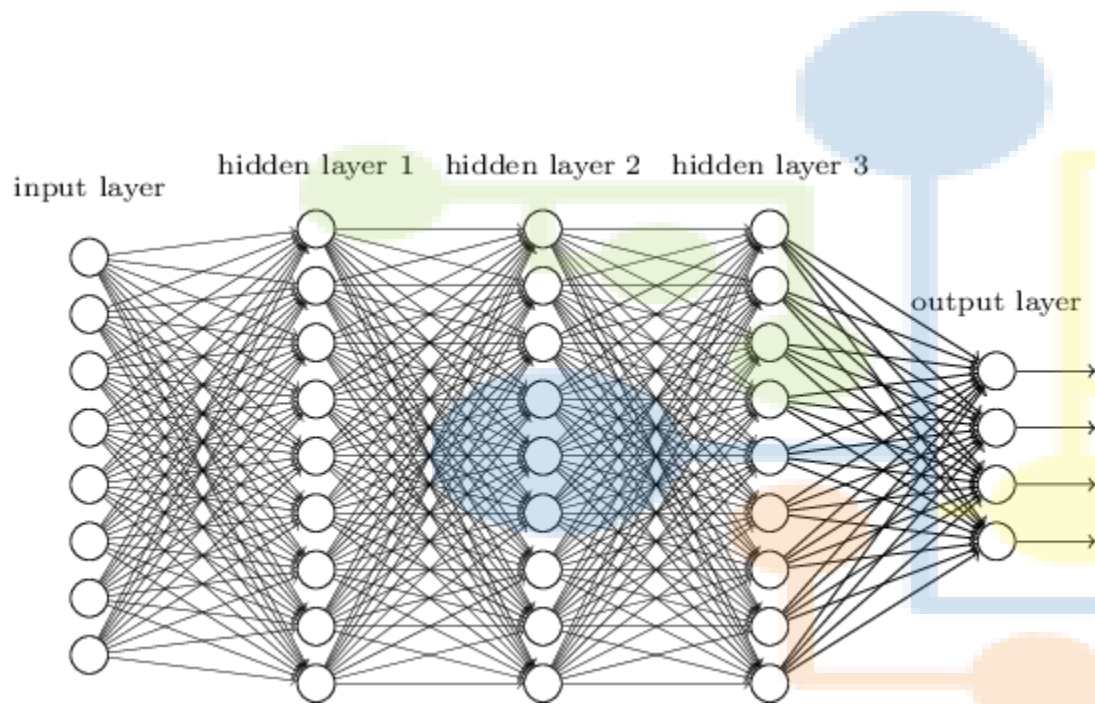
Deep Learning







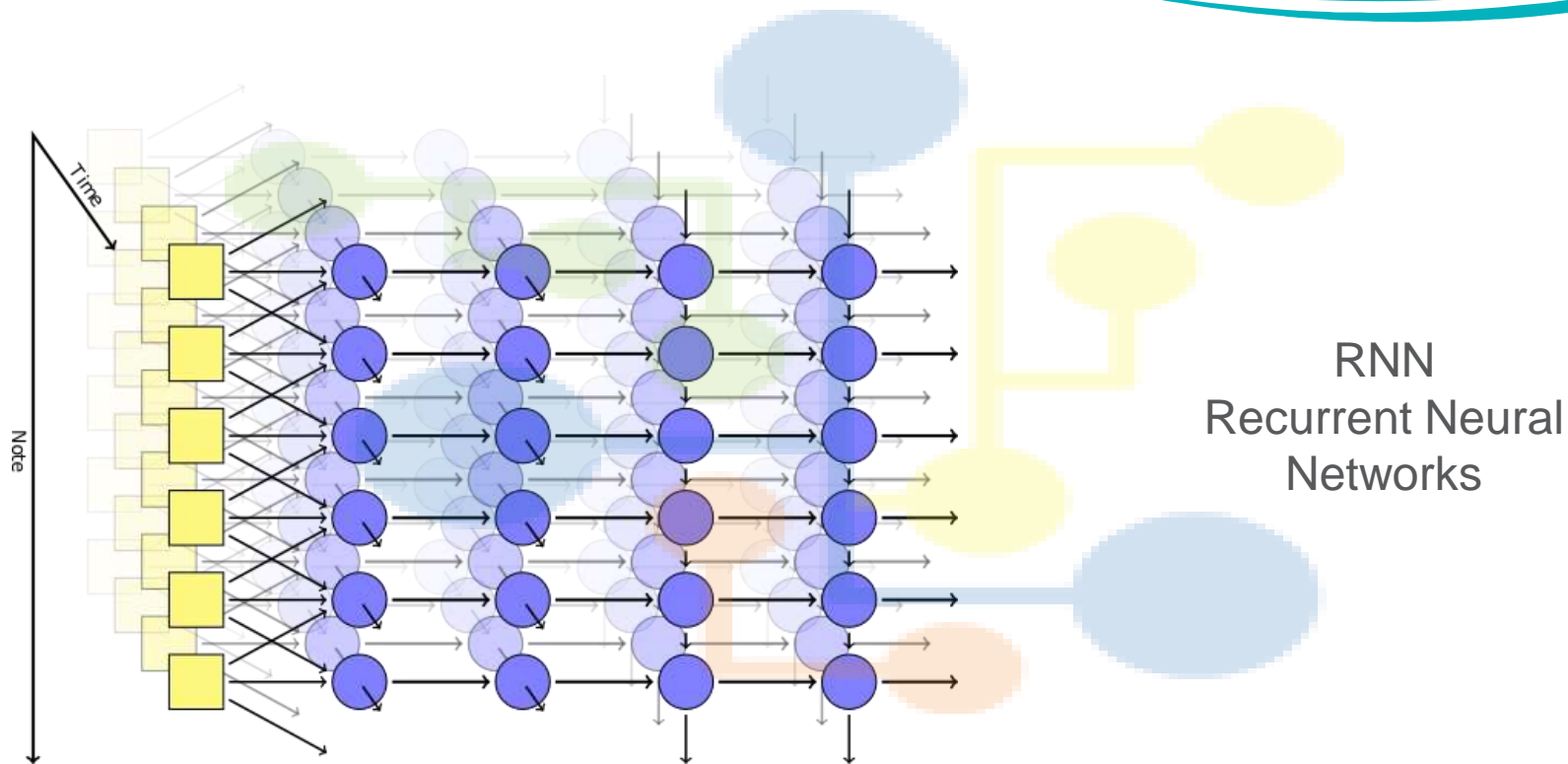


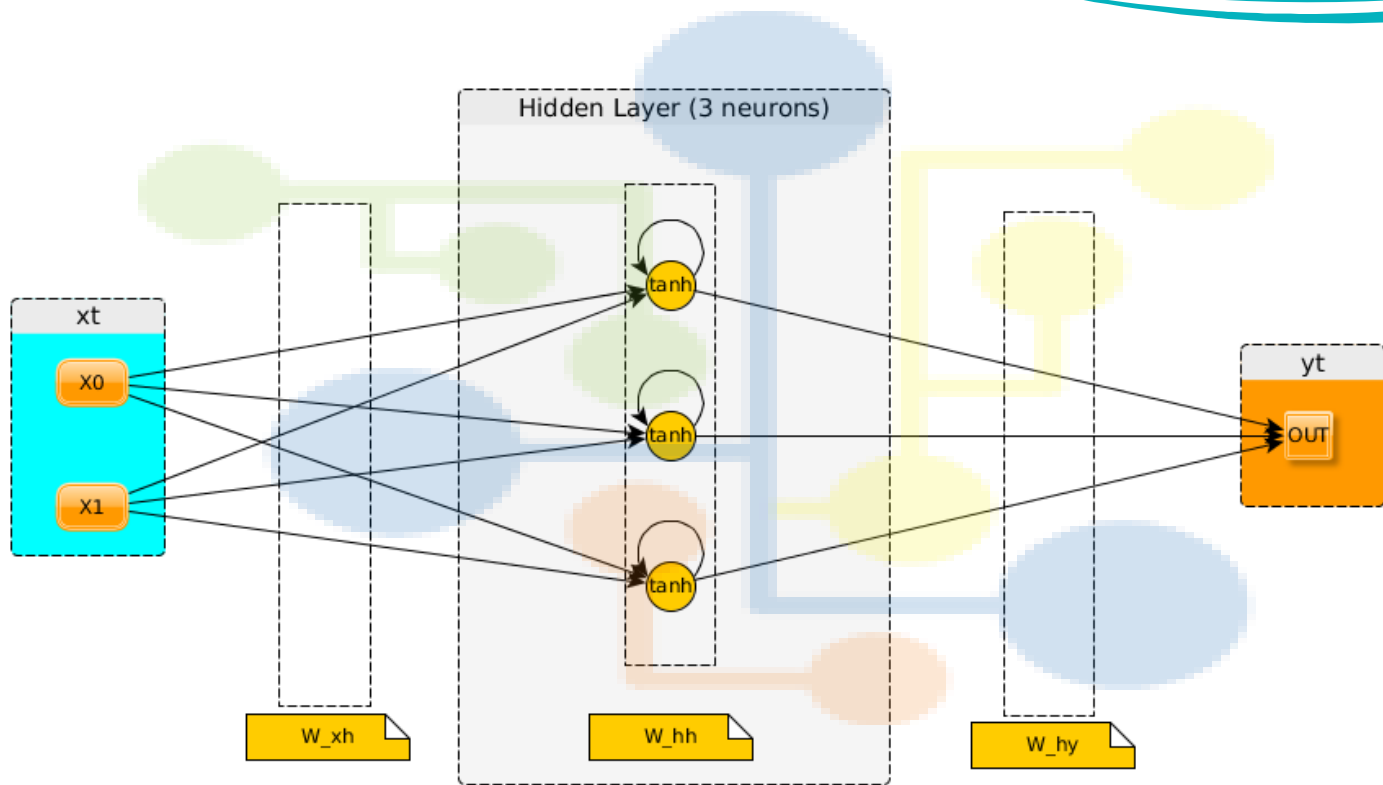


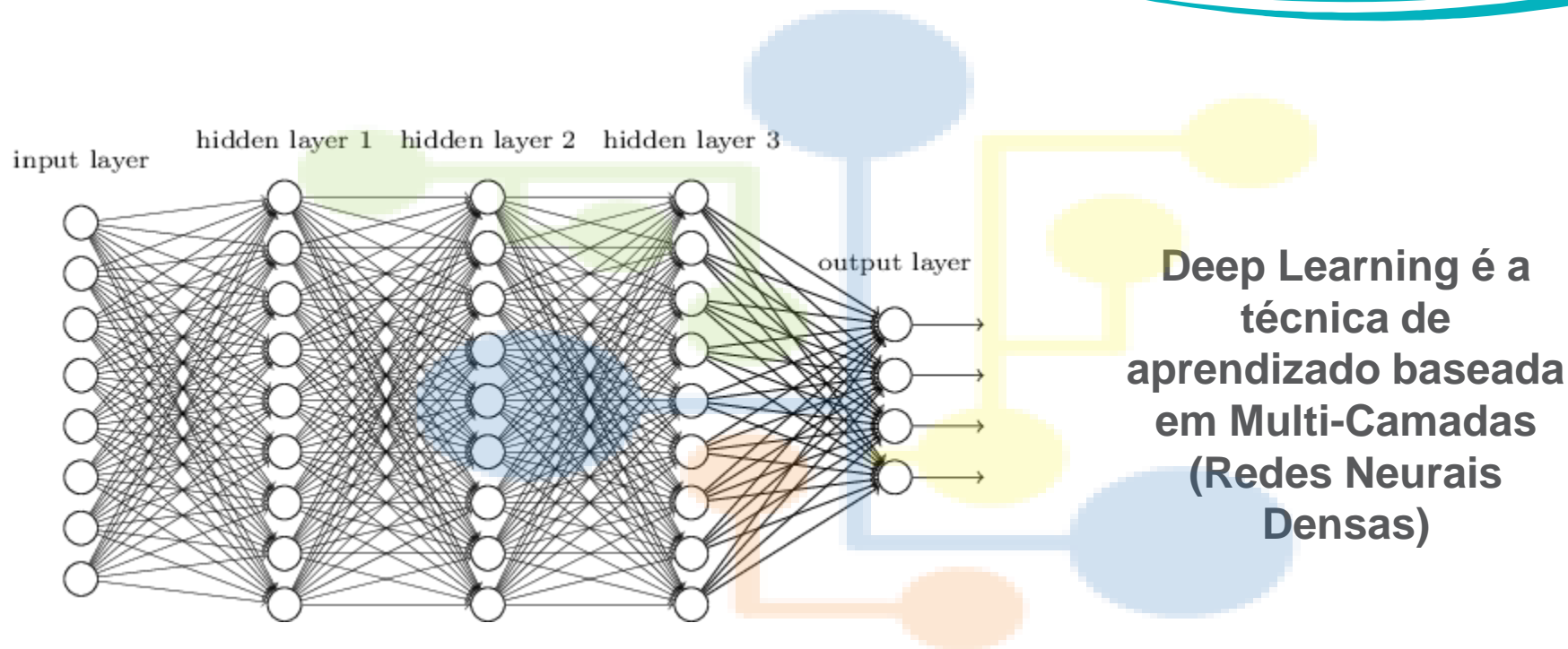
O nome "Rede Neural Convolucional" indica que a rede aplica uma operação matemática denominada convolução.

Convolução é um tipo especializado de operação linear.

Redes convolucionais são redes neurais simples que utilizam convolução no lugar de multiplicação geral de matrizes em pelo menos uma das camadas.



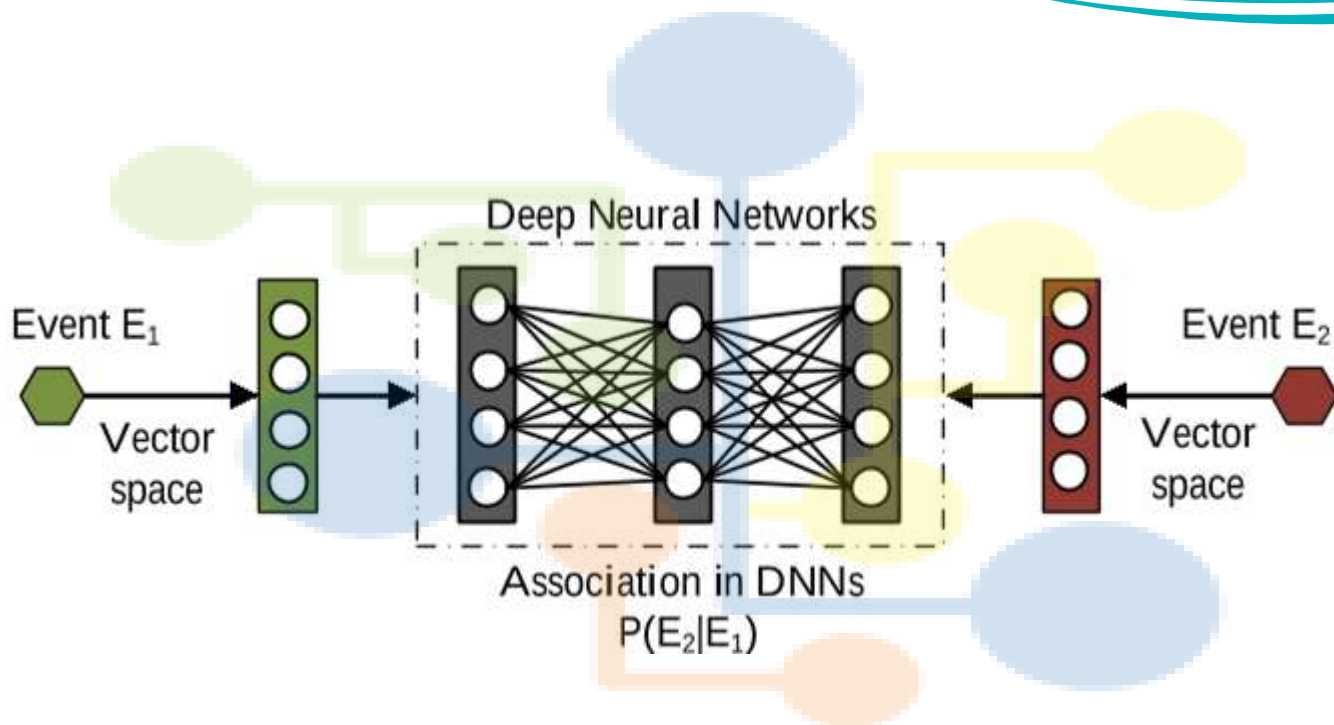


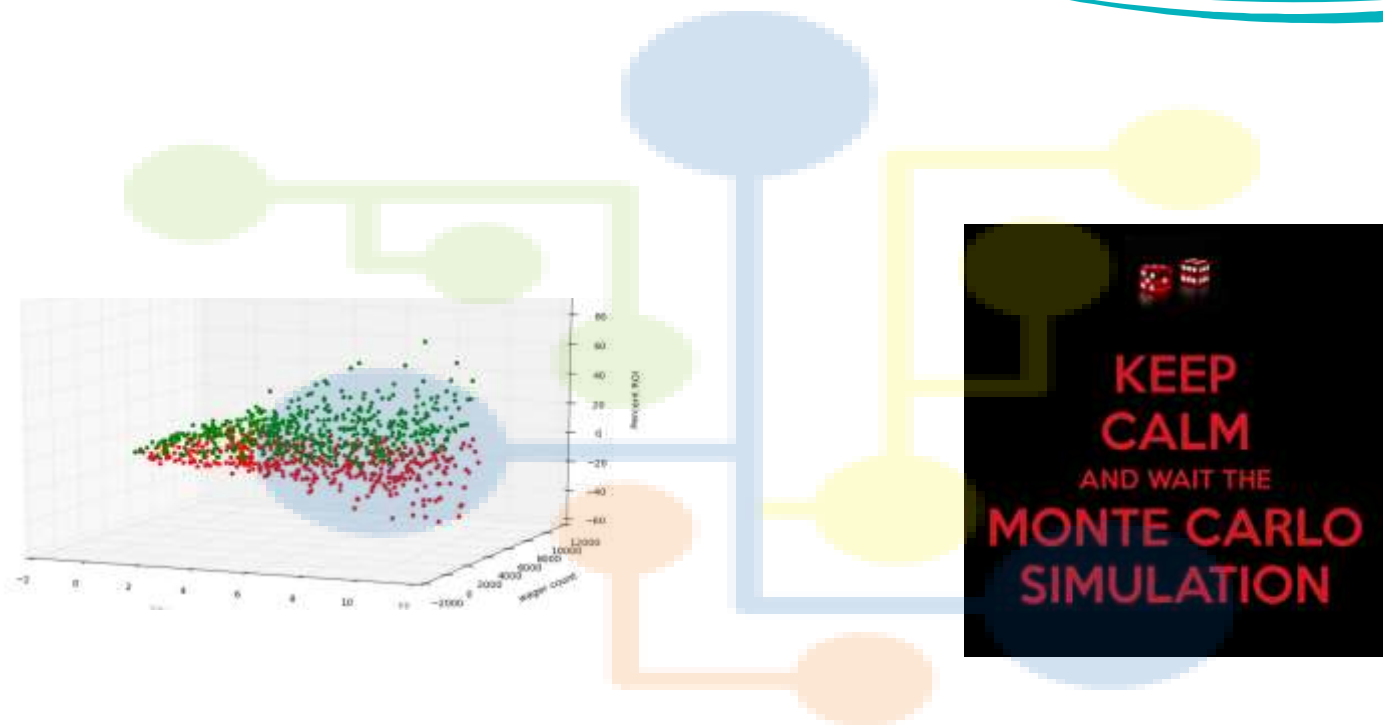




Reconhecimento de Voz e Processamento de Linguagem Natural

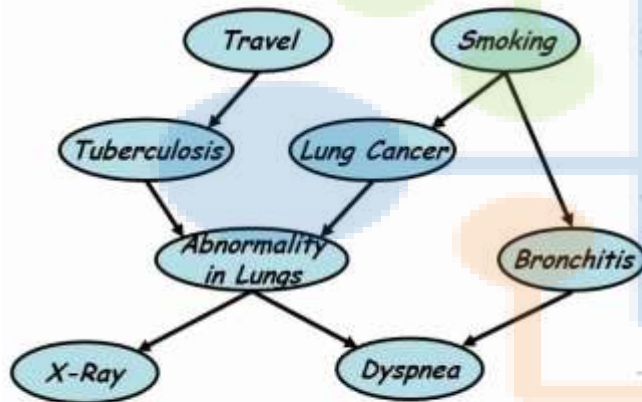








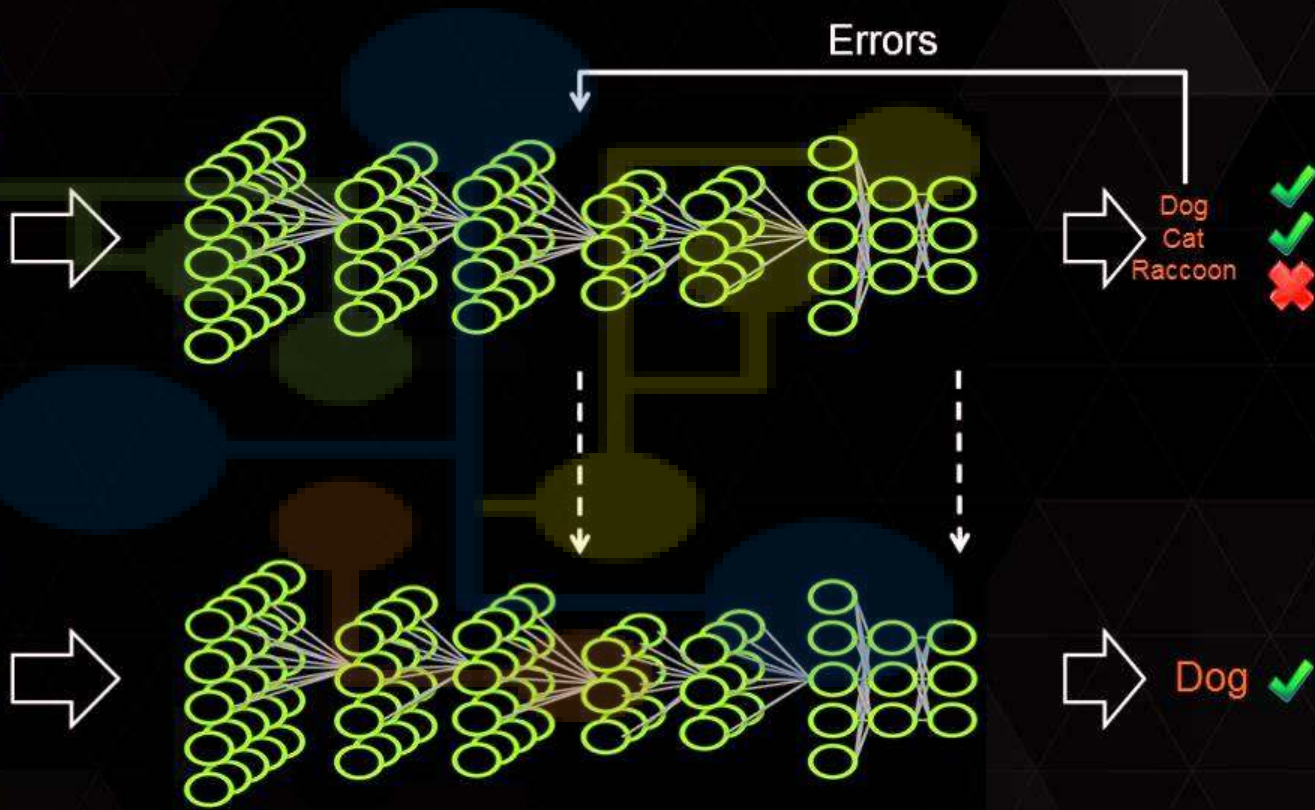
Muitos modelos probabilísticos são difíceis de treinar devido a dificuldade de realizar inferência



Data Science Academy

DEEP LEARNING APPROACH

Train:



Deploy:





Deep Learning está revolucionando a indústria aeroespacial





Data Science
Academy

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



Data Science Academy

Simulação e Otimização





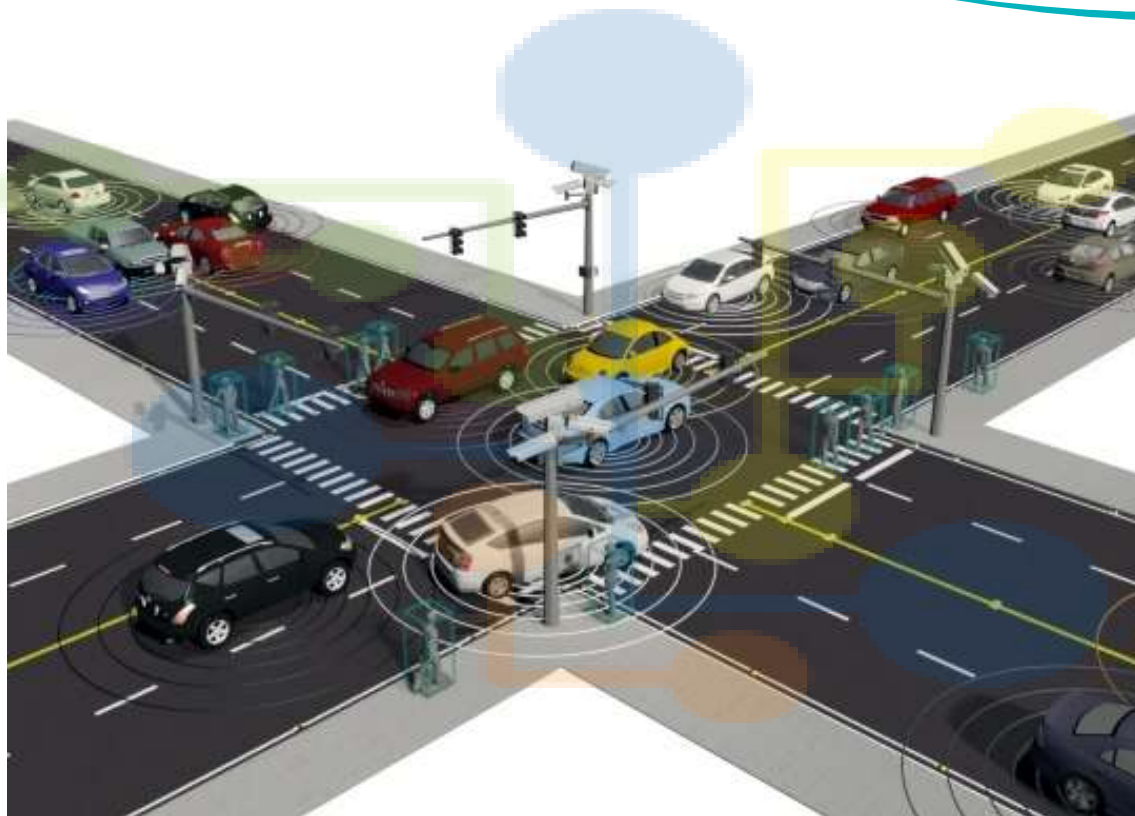
Podemos entender a simulação como um processo amplo que engloba não apenas a construção do modelo, mas todo o método experimental que se segue

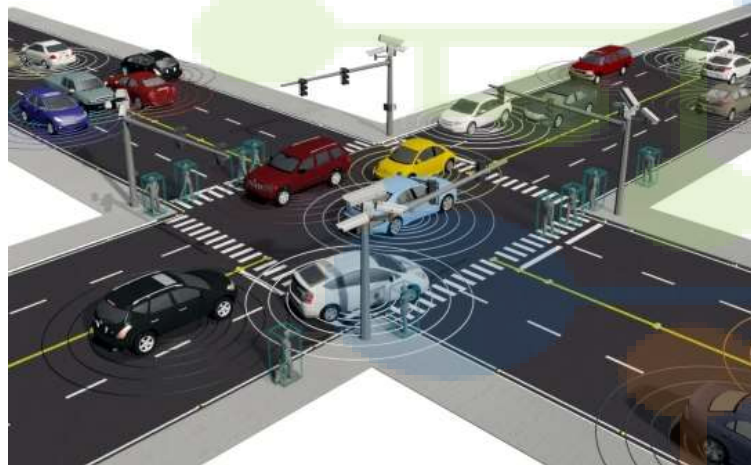


Simulação implica na modelagem de um processo ou sistema, de tal forma que o modelo imite as respostas do sistema real em uma sucessão de eventos que ocorrem ao longo do tempo

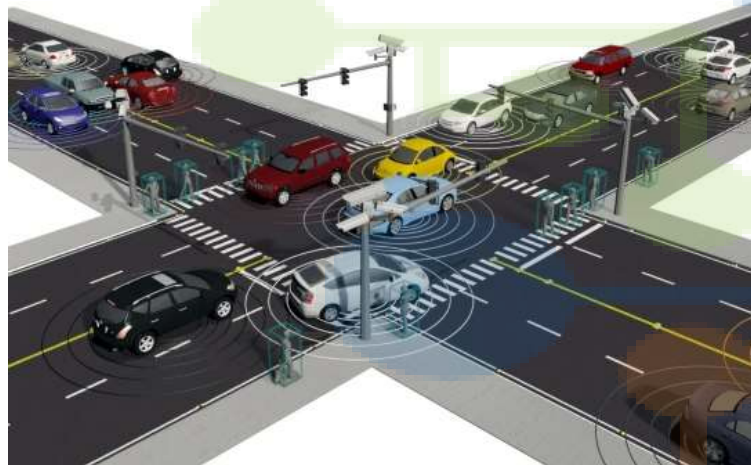


- Um dispositivo para compreensão de um problema;
- Um meio de comunicação para descrever a operação de um sistema;
- Uma ferramenta de análise para determinar elementos críticos e estimar medidas de desempenho;
- Uma ferramenta de projeto para avaliar problemas e propor soluções;
- Um sistema de planejamento de operações para trabalhos, tarefas e recursos;
- Um mecanismo de controle;
- Uma ferramenta de treinamento;
- Uma parte do sistema para fornecer informações on-line, projeções de situações e suporte à decisão.

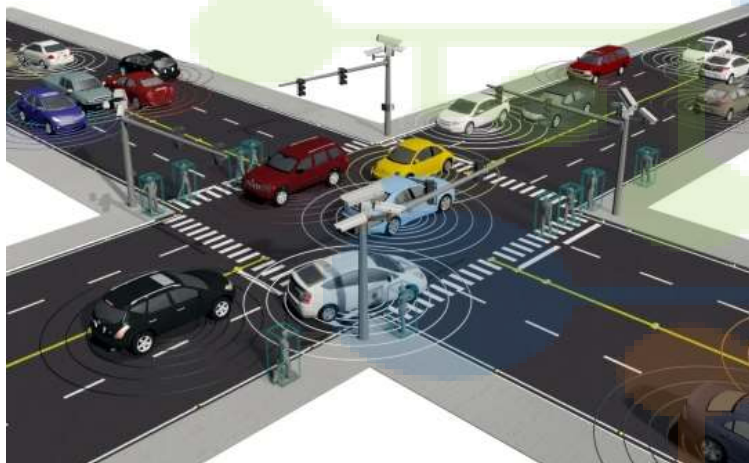




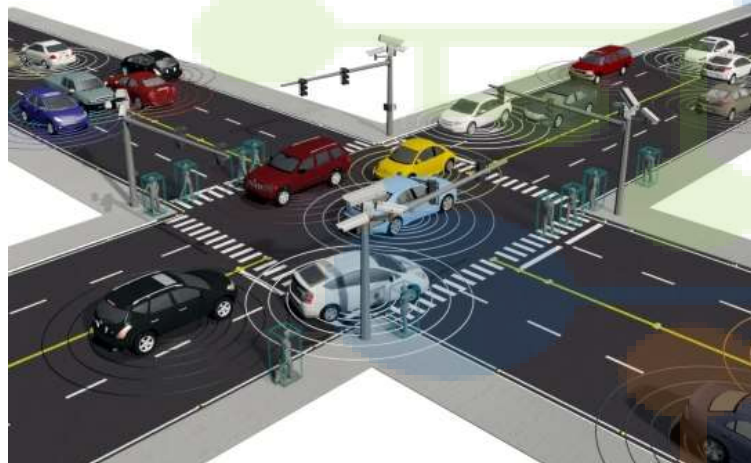
Variáveis



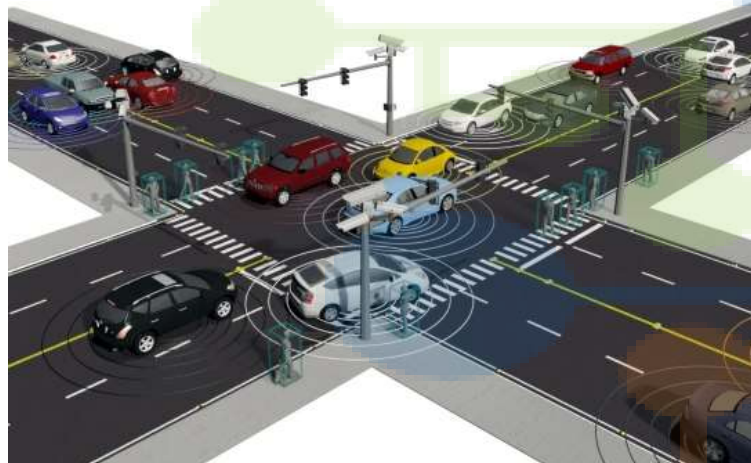
Variáveis de Estado



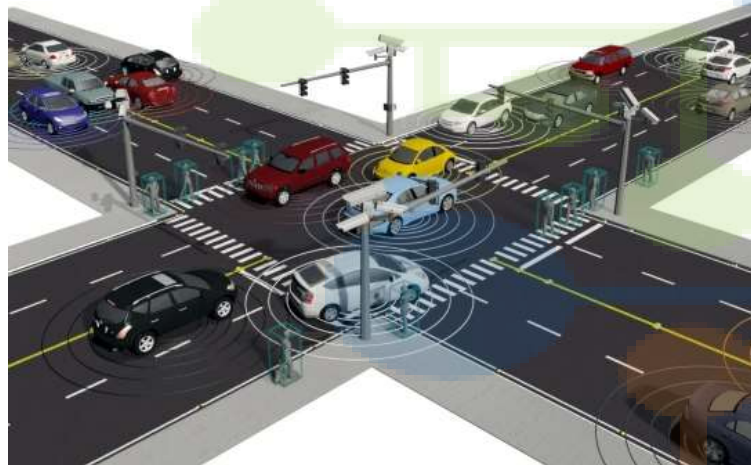
Entidade



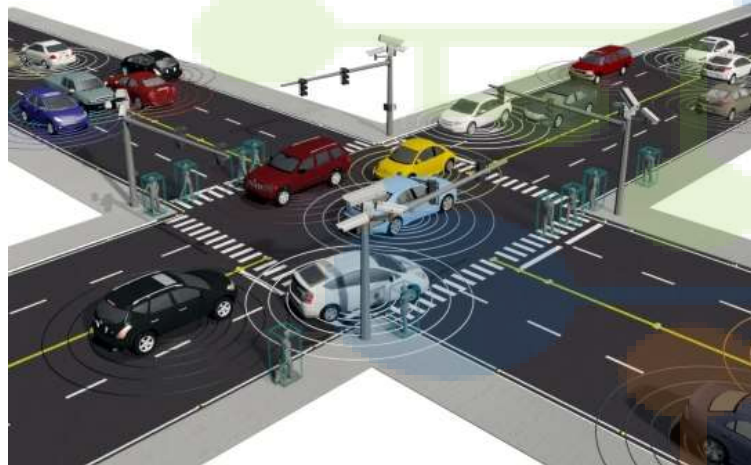
Atributo



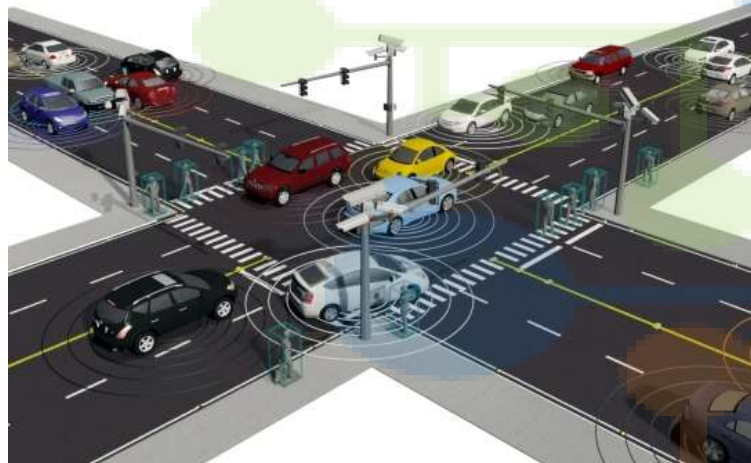
Recurso



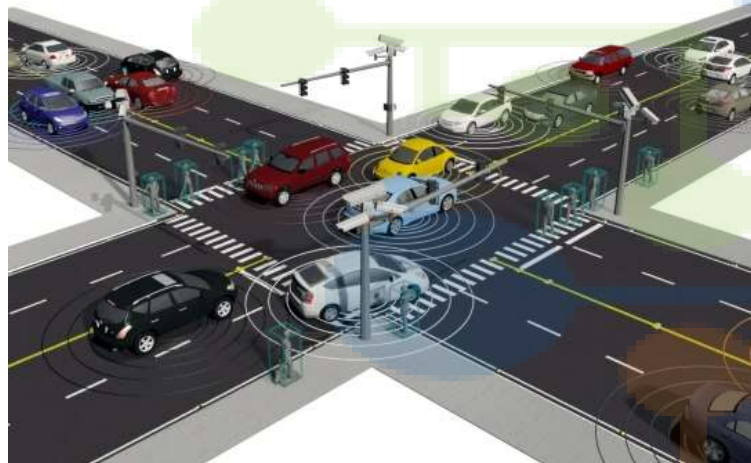
Processos



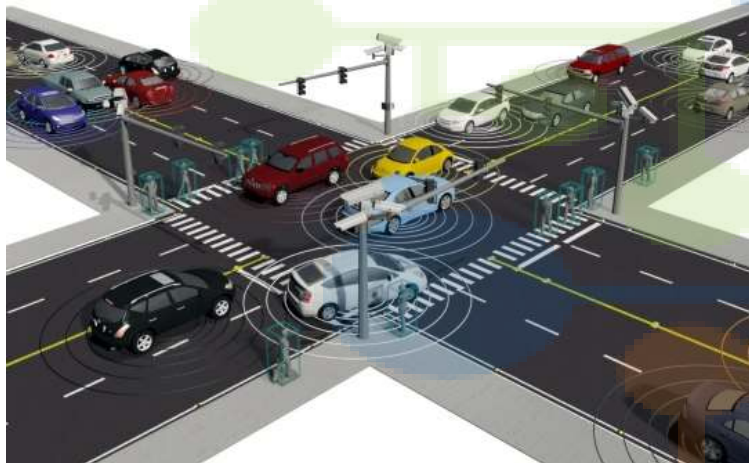
Tempo de
Simulação



Filas



Eventos



A simulação é a técnica
empregada durante a fase de
construção do modelo preditivo



Pode ser necessário a simulação
de uma grande variedade de
alternativas e a criação de
modelos preditivos pode gerar
bons resultados



Machine Learning nos ajuda a simular um determinado evento através de dados e com isso, prever o comportamento futuro.

An abstract background diagram consisting of several colored circles (blue, green, yellow, orange) connected by lines of the same color, forming a network-like structure. The text is centered over this diagram.

Modelos Determinísticos
X
Modelos Estocásticos



Modelos Determinísticos



Modelos Estocásticos



Quando uma variável de entrada de um sistema é aleatória, a variável de saída também será aleatória, no entanto, o sistema pode ter comportamento determinístico ou ser representado por um modelo determinístico





Modelo Computacional é um programa de computador cujas variáveis apresentam o mesmo comportamento dinâmico e estocástico do sistema real que representa

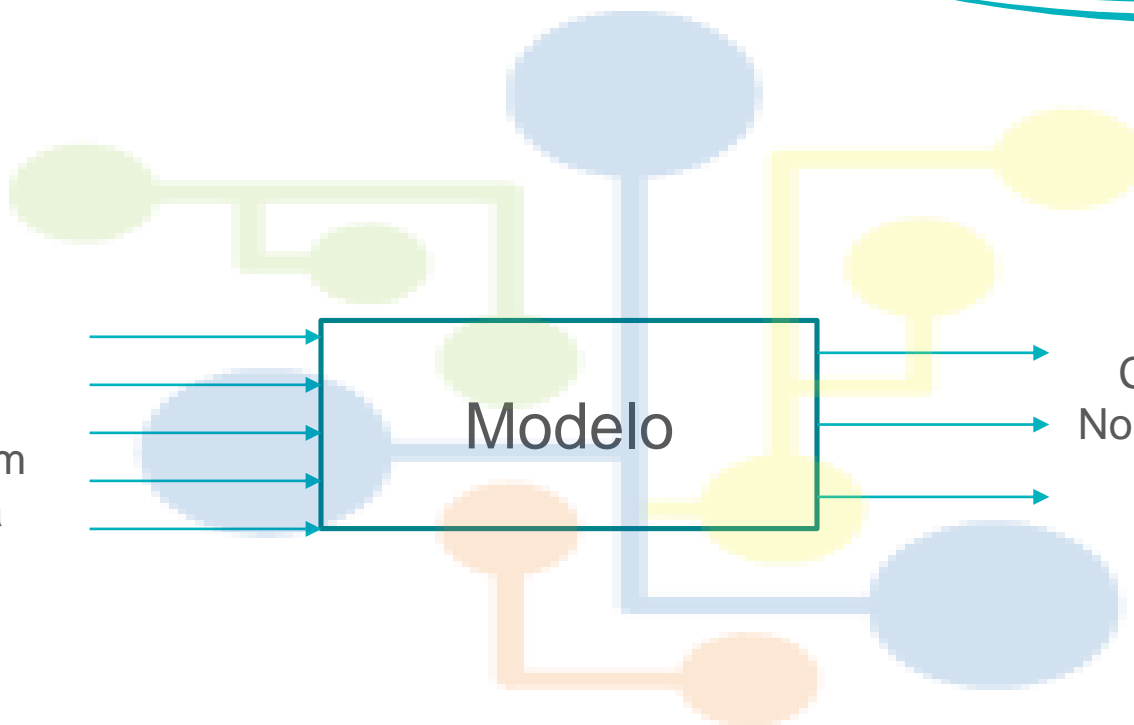




Velocidades
individuais
observadas em
uma avenida

Modelo

Comportamentos
Normais ou Anormais
e Tendências





Determinístico

Resultado do modelo é pré-determinado em função dos dados de entrada

Estocástico

Resultado do modelo não depende somente dos dados de entrada, mas também de outros fatores, normalmente aleatórios. Isso requer um modelo probabilístico.



Determinístico

Exemplo: Se uma pessoa tem mais de 16 anos, ela pode tirar carteira de motorista. Se tiver menos de 16, não pode.

Estocástico

Exemplo: Modelo para prever a reação de pessoas em um shopping, a uma situação de emergência. Um modelo probabilístico tenta descrever o comportamento "aleatório" das entidades



Modelos determinísticos e estocásticos podem ser combinados para resolver problemas que requerem muitas alternativas diferentes, tais como:

- Busca na Web e Extração de Informação
- Desenvolvimento de Novos Medicamentos
- Prever o Comportamento do Mercado Financeiro
- Compreender o Comportamento de Clientes
- Criação de Robôs

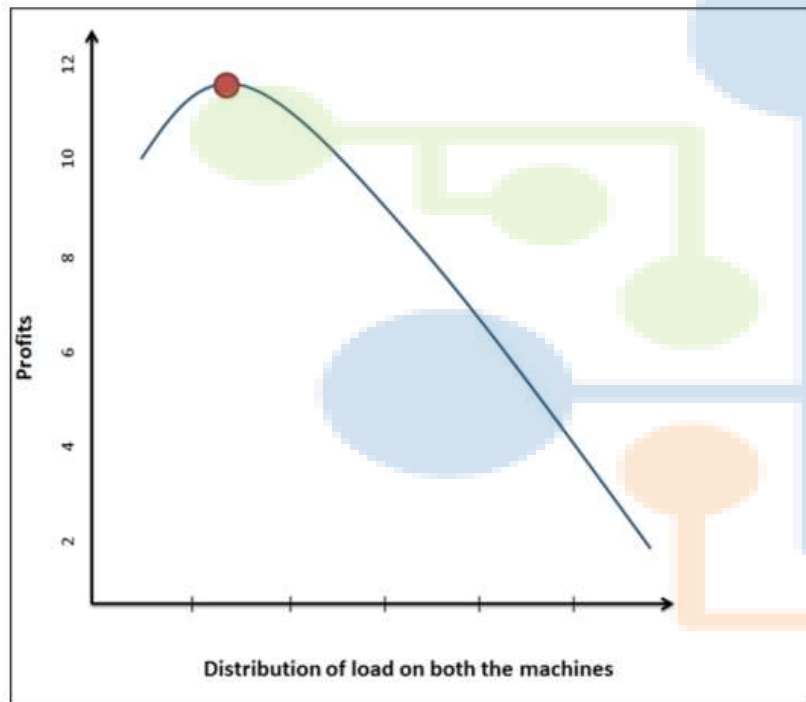




Otimização

An abstract network diagram with a central vertical blue line. To the left, a green line branches out to three green circles. To the right, a yellow line branches out to three yellow circles. At the bottom, an orange line branches out to two orange circles. A large blue circle is at the top, and another large blue circle is at the bottom right. The word 'Otimização' is centered over the diagram.





Otimização

An abstract background graphic consisting of several large, semi-transparent circles in blue, green, yellow, and orange, connected by thin lines of the same colors, creating a network-like structure.

Aprendizado = Representação + Avaliação + Otimização



Aprendizado =

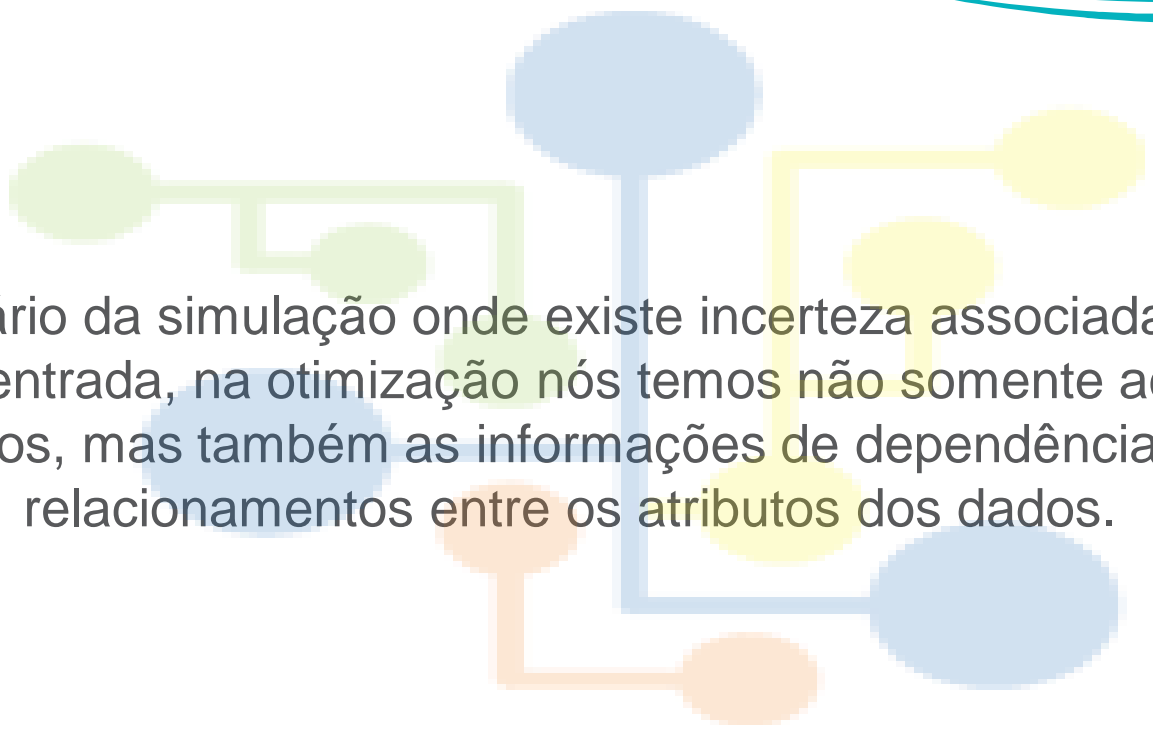
Representação

+

Avaliação

+

Otimização

An abstract background diagram consisting of several colored circles (blue, green, yellow, orange) connected by lines of the same color, forming a network-like structure. The circles are of varying sizes and are distributed across the slide, with some lines connecting them in a way that suggests a flow or relationship.

Ao contrário da simulação onde existe incerteza associada com os dados de entrada, na otimização nós temos não somente acesso aos dados, mas também as informações de dependências e relacionamentos entre os atributos dos dados.



Generalização → Principal Objetivo na Construção do Modelo Preditivo



Seleção de Modelo





Definir Espaço de
Parâmetro

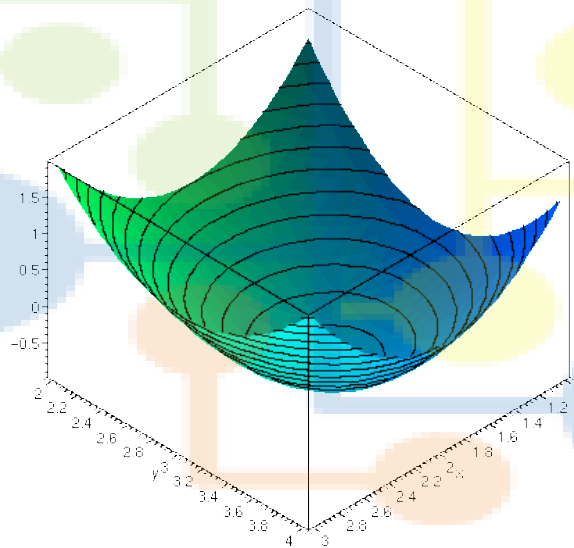
Definir
Configurações de
Validação Cruzada

Definir Métrica

Treinar, Avaliar e
Comparar

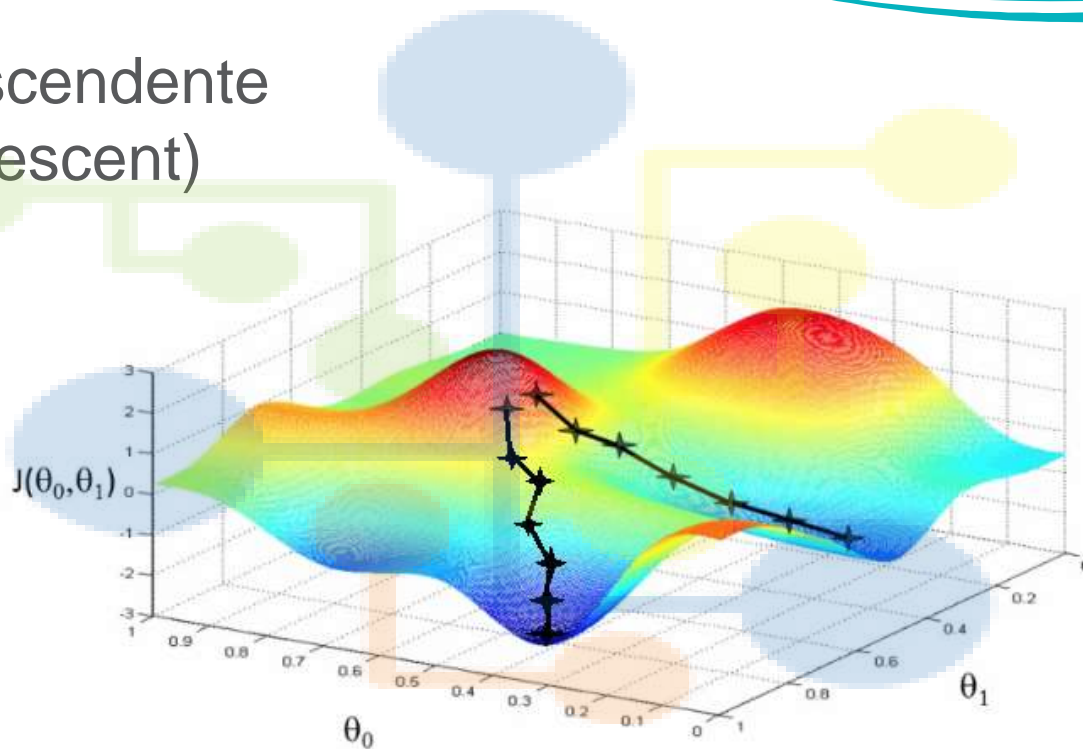


Gradiente Descendente (Gradient Descent)



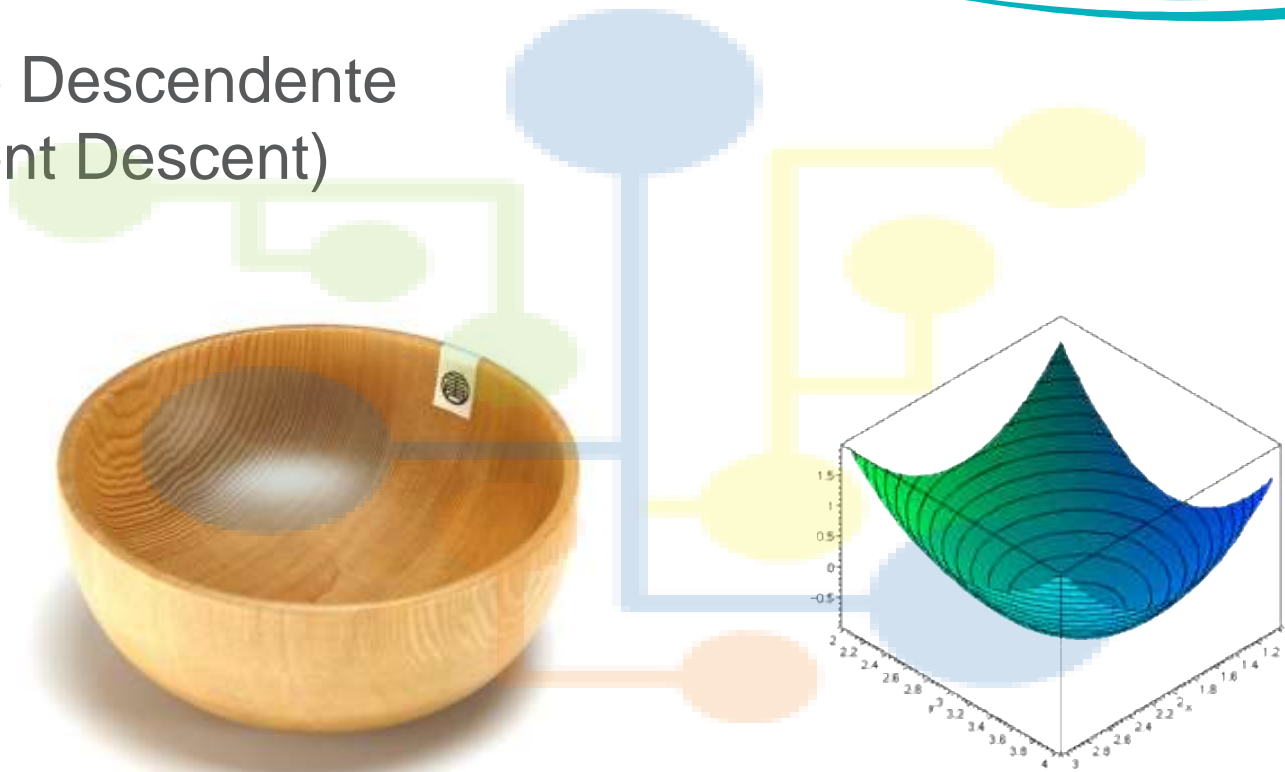


Gradiente Descendente (Gradient Descent)





Gradiente Descendente (Gradient Descent)





Gradiente Descendente (Gradient Descent)



coefficient = 0.0

cost = $f(\text{coefficient})$

delta = derivative(cost)

coefficient = coefficient - (alpha * delta)



Gradiente Descendente (Gradient Descent)

Batch Gradient Descent

Stochastic Gradient Descent



Data Science
Academy

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

Obrigado