



Data Science
Academy

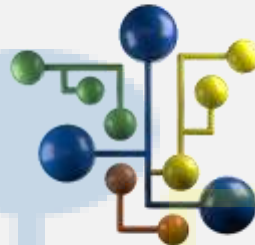
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Machine Learning



Data Science
Academy

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

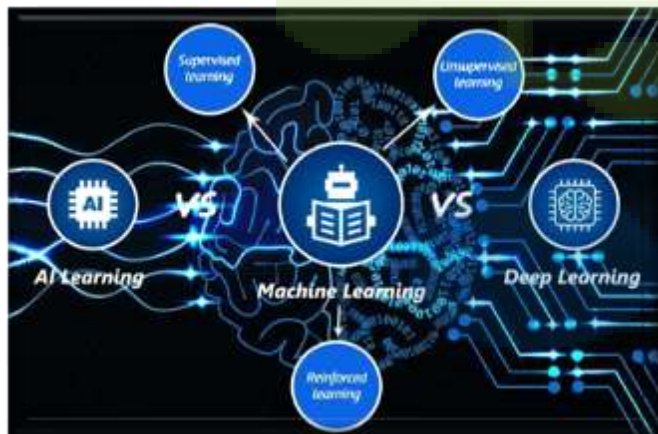


**Data Science
Academy**

Seja muito bem-vindo(a)!



Como Funciona a Aprendizagem de Máquina



Processo de Aprendizagem

Como os Algoritmos Aprendem

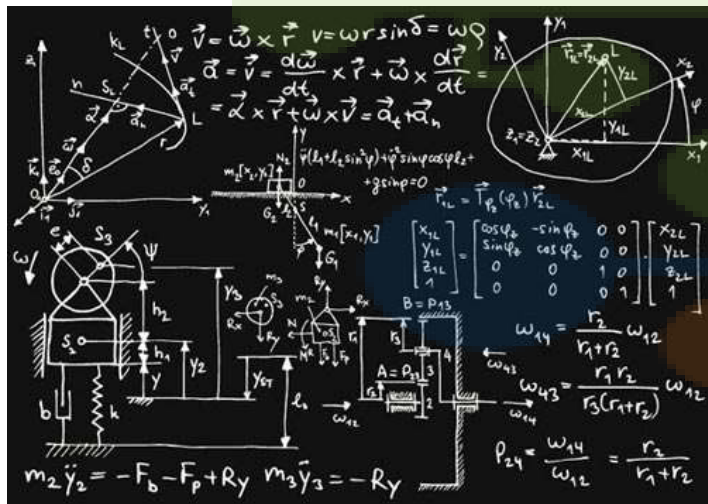




Processo de Aprendizagem



Processo de Aprendizagem



Processo de Aprendizagem

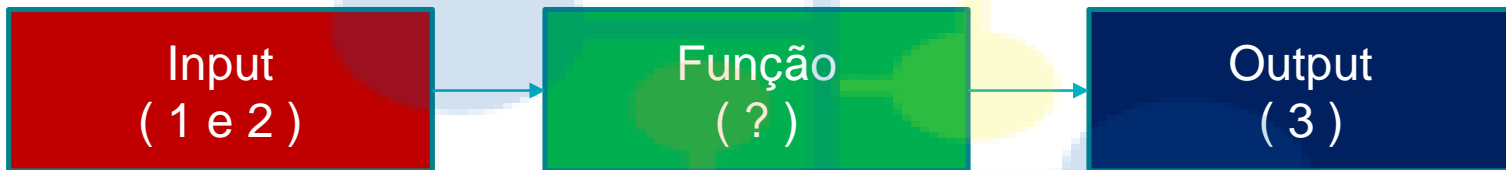


Processo de Aprendizagem




Processo de Aprendizagem

Machine Learning



E como um algoritmo encontra a função matemática que descreve este relacionamento?

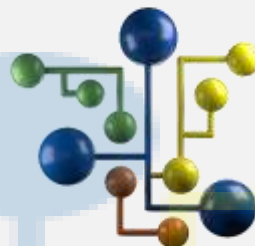
A faint, stylized diagram in the background consisting of several colored circles (blue, green, yellow, orange) connected by lines, resembling a network or a flowchart.

Isso é o que vamos responder ao longo
dos próximos vídeos!



Data Science
Academy

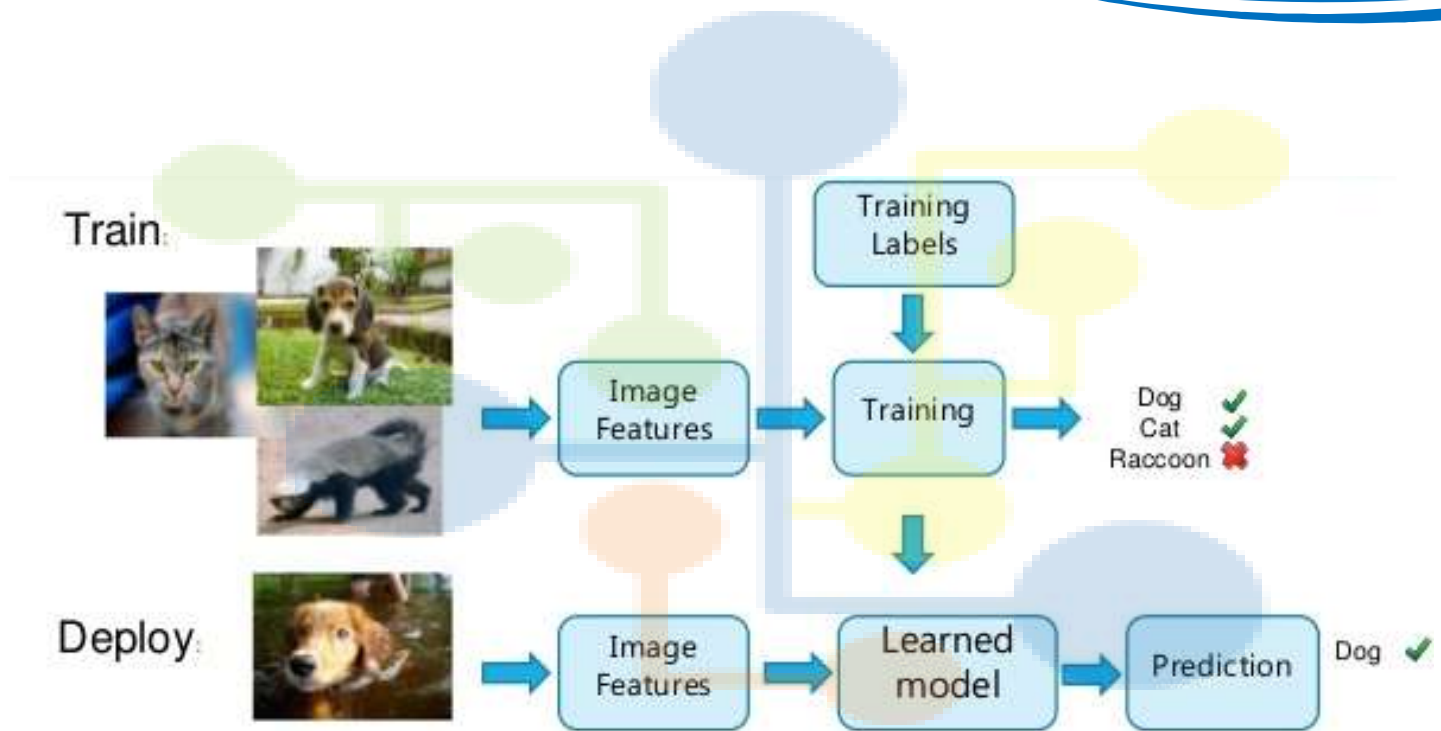
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

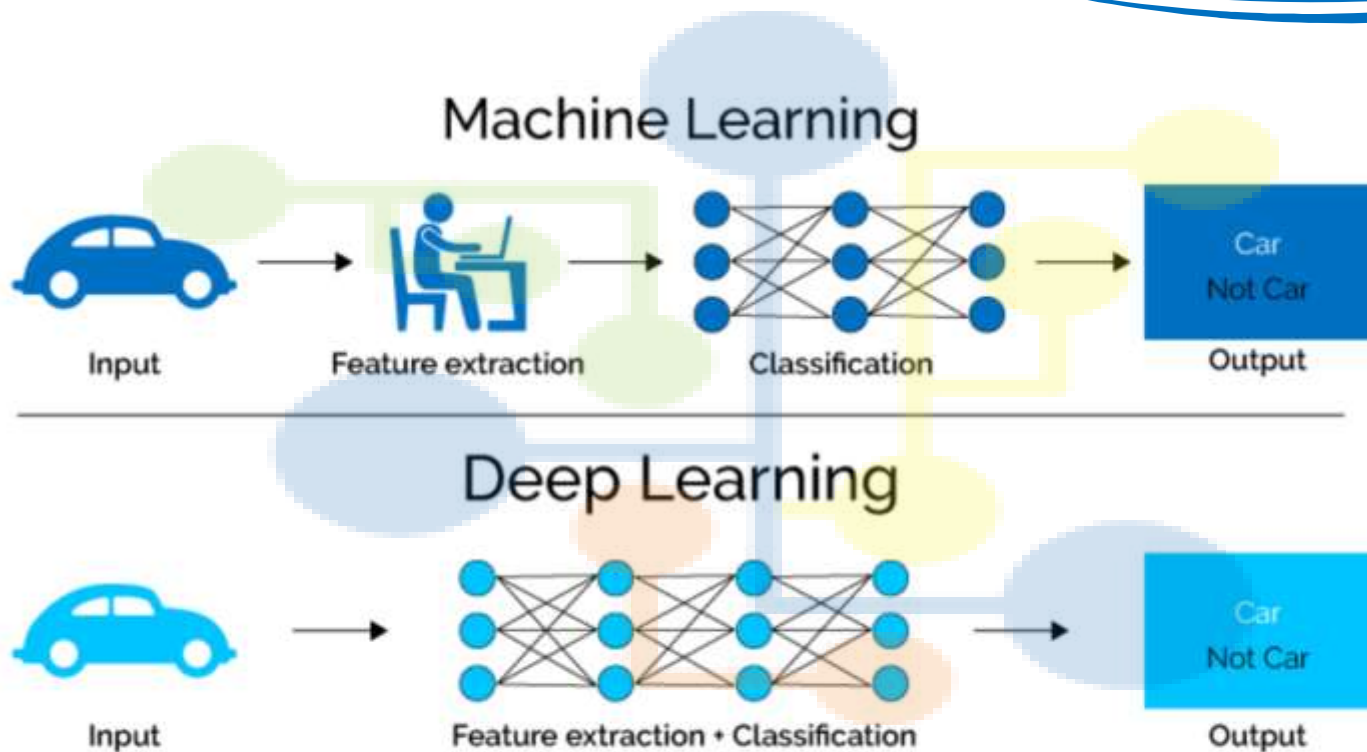


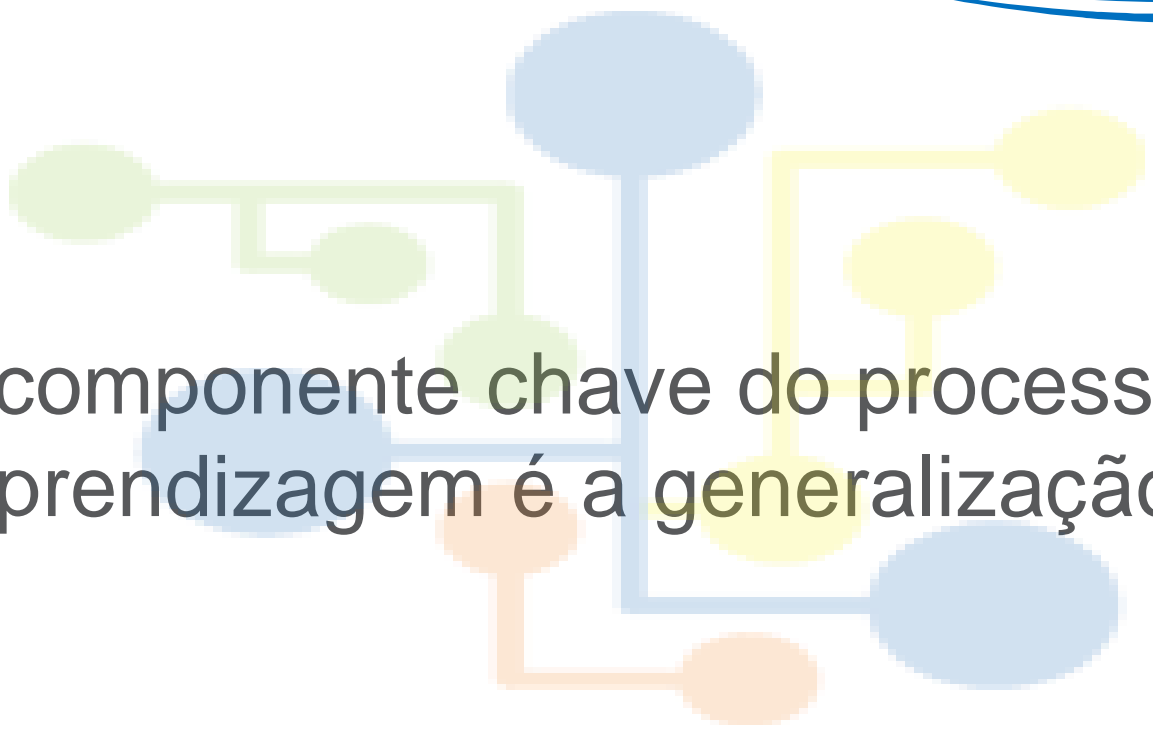
**Data Science
Academy**

Processo de Aprendizagem





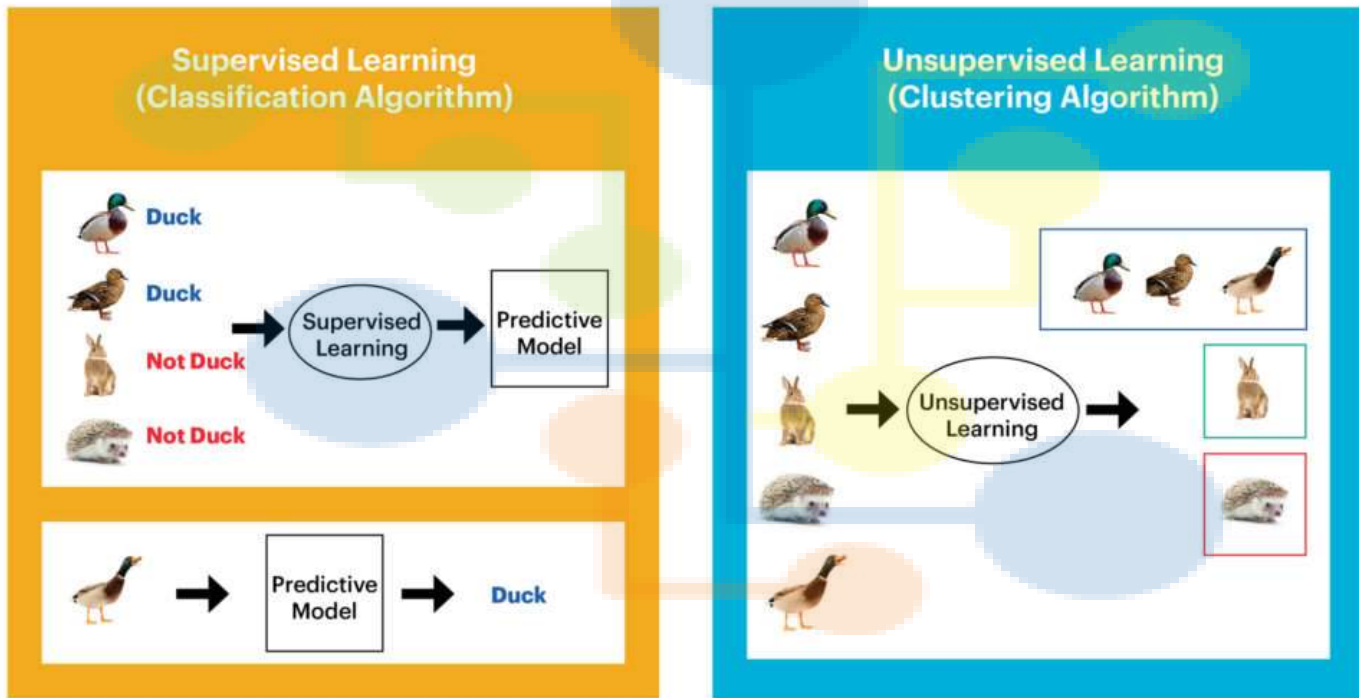


An abstract network diagram in the background, consisting of several circular nodes connected by lines. The nodes are colored in shades of blue, green, yellow, and orange. The lines connecting them are also in these colors, creating a complex web-like structure.

Um componente chave do processo de aprendizagem é a generalização.



Se um algoritmo de Machine Learning não for capaz de generalizar uma função matemática que faça previsões sobre novos conjuntos de dados, ele não está aprendendo nada e sim memorizando os dados, o que é bem diferente.



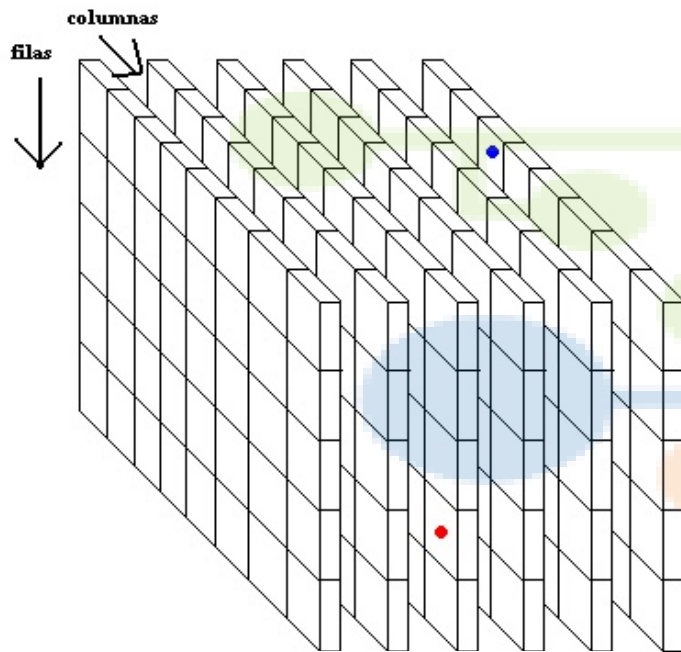


E para poder generalizar a função que melhor resolve o problema, os algoritmos de Machine Learning se baseiam em 3 componentes:

Representação

Avaliação

Otimização

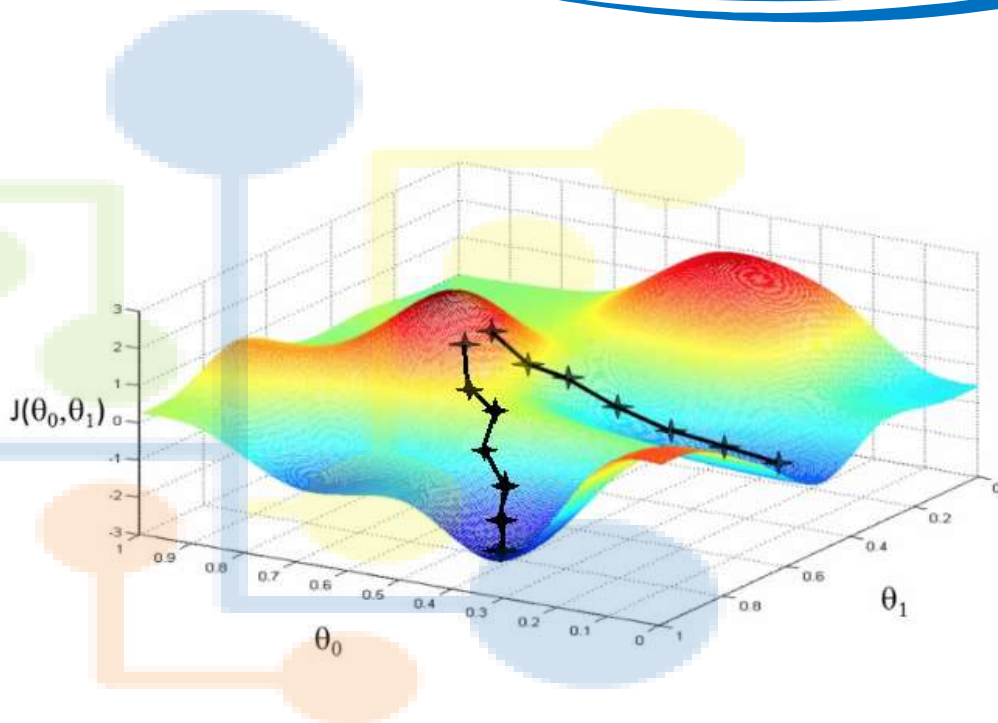


Os algoritmos de aprendizagem possuem diversos parâmetros internos.



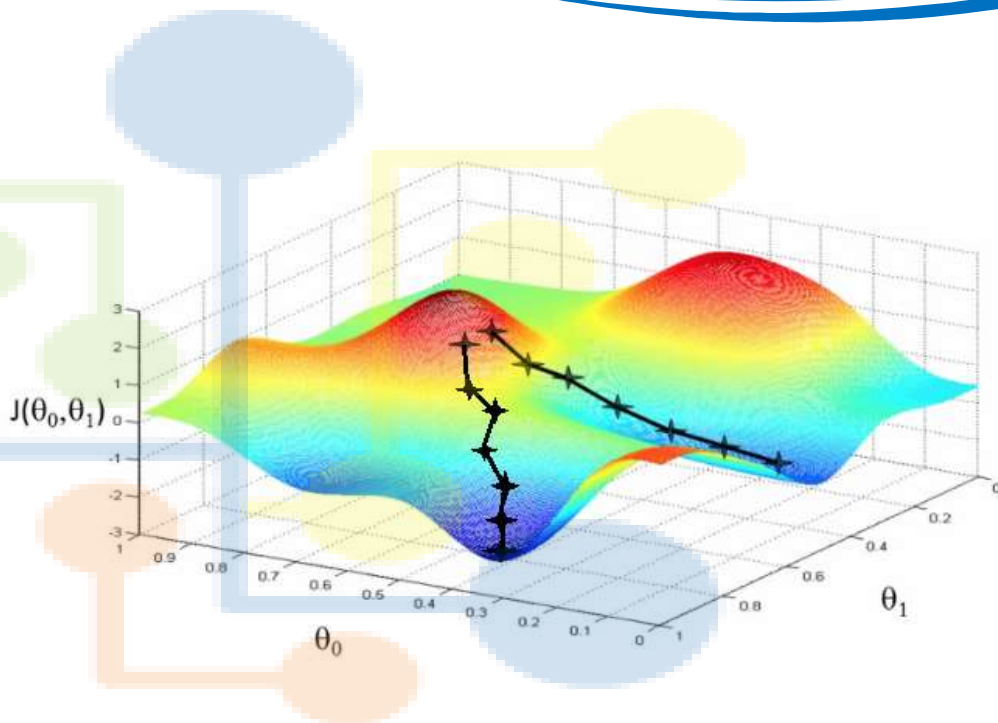


Otimização





Espaço de Hipótese





Nenhum algoritmo único ou uma combinação de algoritmos é 100% preciso o tempo todo.

Pelo menos não ainda!!



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.



Hypothesis:

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

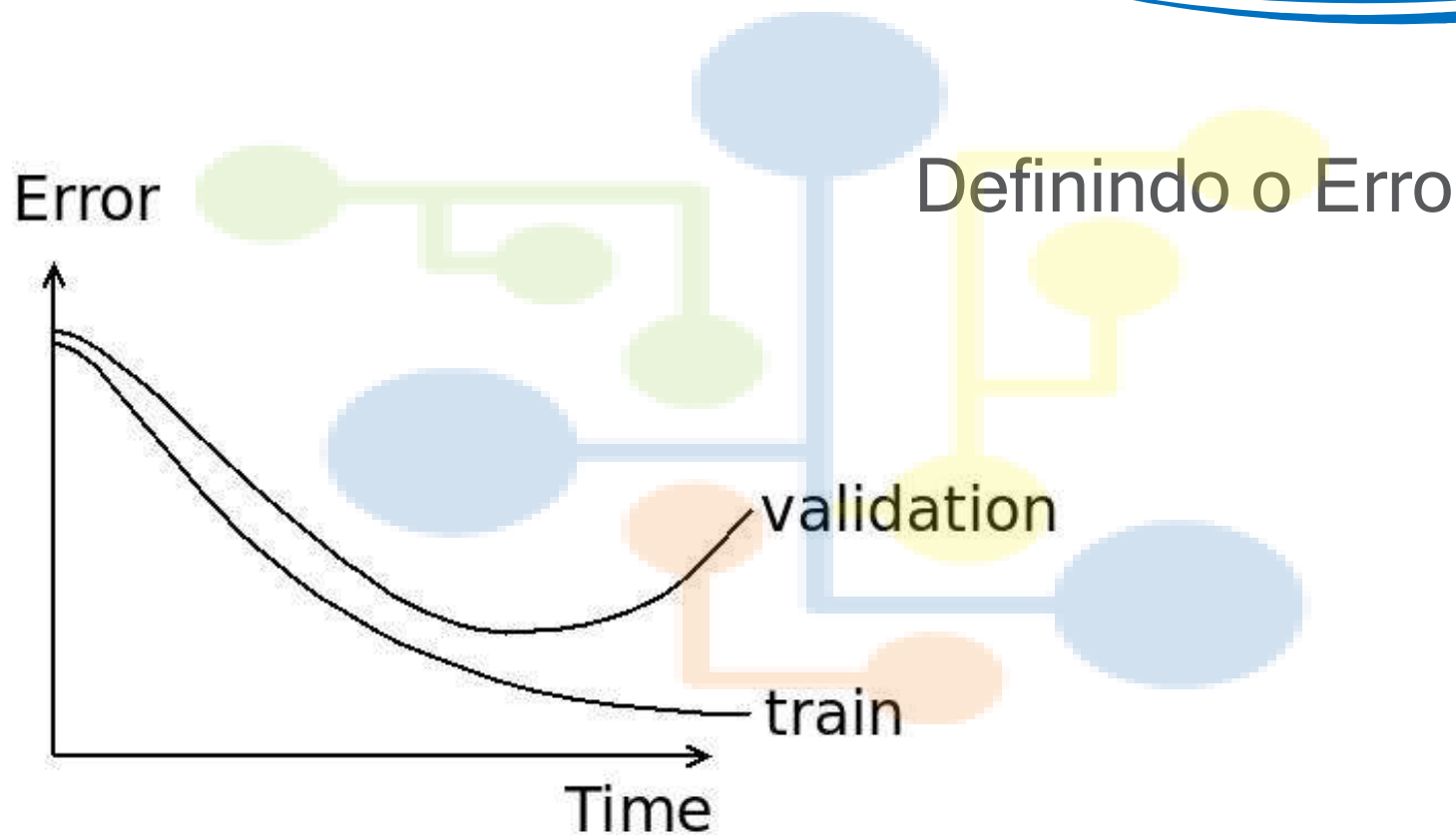
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Goal:

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

Cost Function

A decorative background diagram consisting of a network of 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, light yellow, and light orange, matching the nodes. The diagram is positioned on the right side of the slide, behind the 'Cost Function' text.





Cost Function → Nível de erro

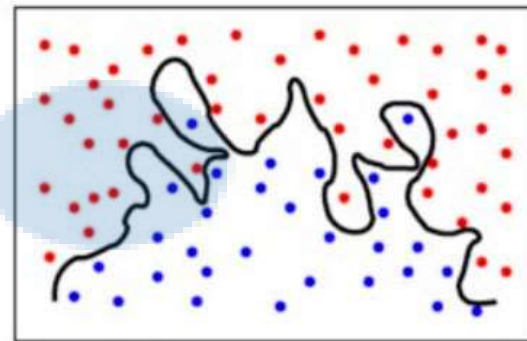
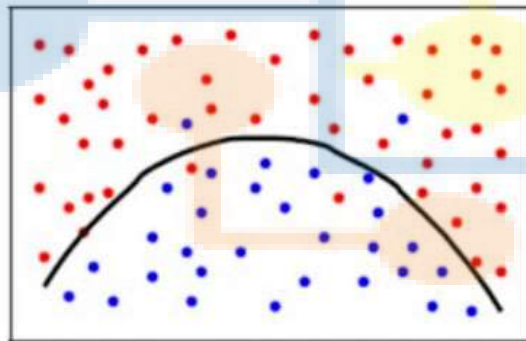
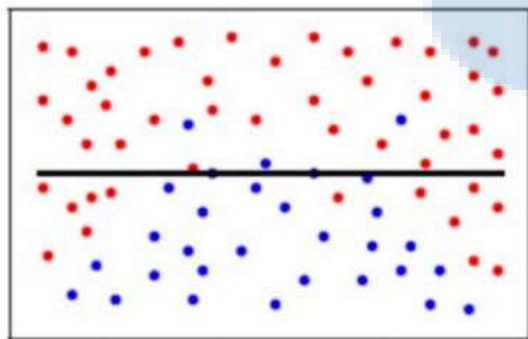


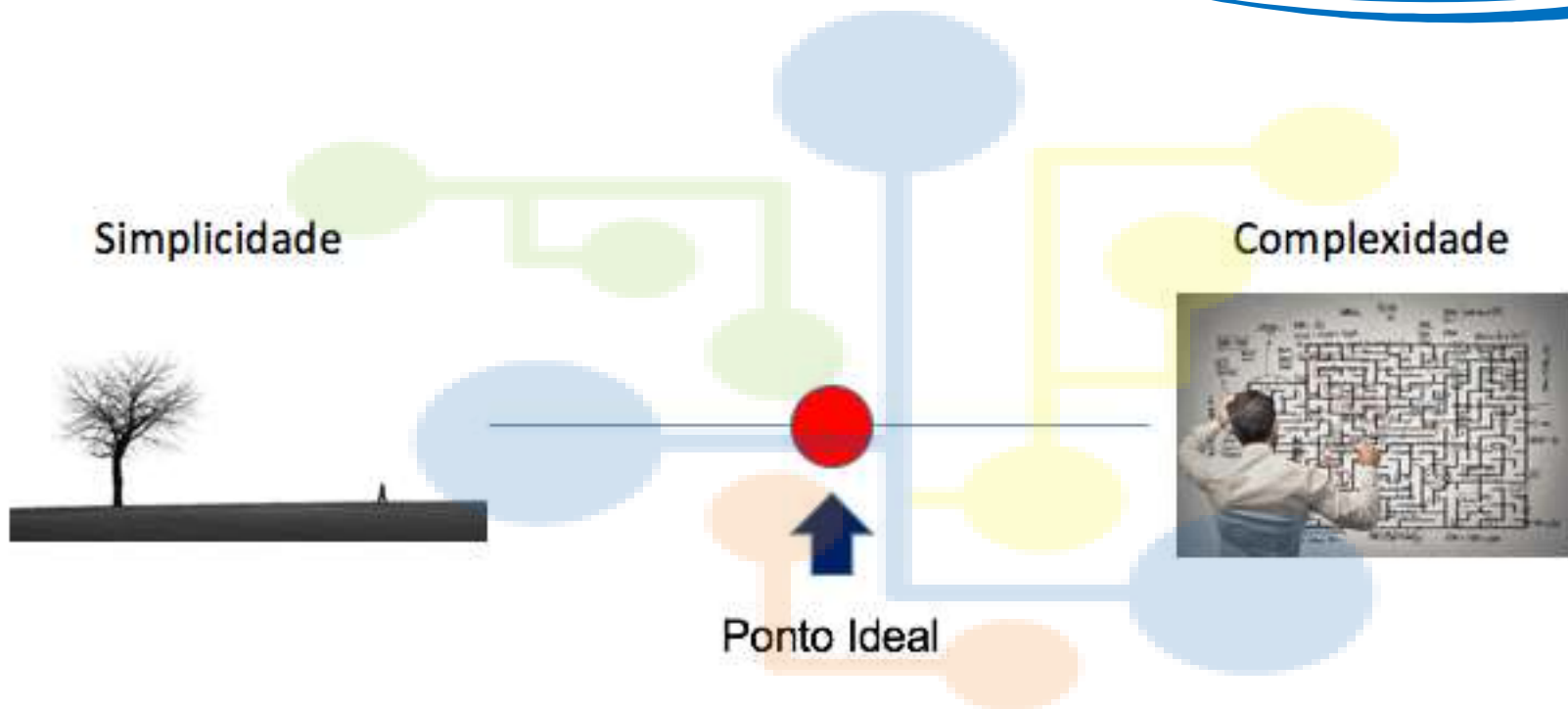
Underfitting x Overfitting

Underfitting

Ok

Overfitting







Elementos do Processo de Aprendizagem



Data Science
Academy

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

Machine Learning



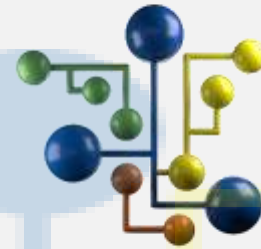


Elementos do Processo de Aprendizagem

Um padrão existe

Não há um único
modelo matemático que
explique esse padrão

Dados estão disponíveis



**Data Science
Academy**

Componentes do Processo de Aprendizagem




Elementos do Processo de Aprendizagem

Um padrão existe

Não há um único
modelo matemático que
explique esse padrão

Dados estão disponíveis

An abstract 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. The diagram is centered on the slide, behind the main title.

Componentes do Processo de Aprendizagem



Aprovação de Crédito



Dados Históricos de Clientes do Banco:

Sexo	Idade	Salário Mensal (R\$)	Anos no Emprego Atual	Anos de Residência	Saldo Bancário (R\$)	Recebeu Crédito
M	35	8.000	6	2	120.000	Sim
F	29	8.500	4	6	78.000	Sim
M	32	7.200	2	7	45.000	Não
M	31	9.600	7	8	9.000	Sim
F	46	10.400	1	2	0	Não

Entrada (variáveis preditoras)

Saída (Label)



Componentes do Processo de Aprendizagem

Input

x

Output

y

Função alvo

$f: x \rightarrow y$

Dados

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

Hipótese

$g: x \rightarrow y$

{Dados do cliente}

{Decisão \rightarrow Crédito: Sim/Não}

{Representação do relacionamento}
{Fórmula matemática desconhecida}

{Dados históricos}

{Faz parte do espaço de hipóteses do algoritmo}



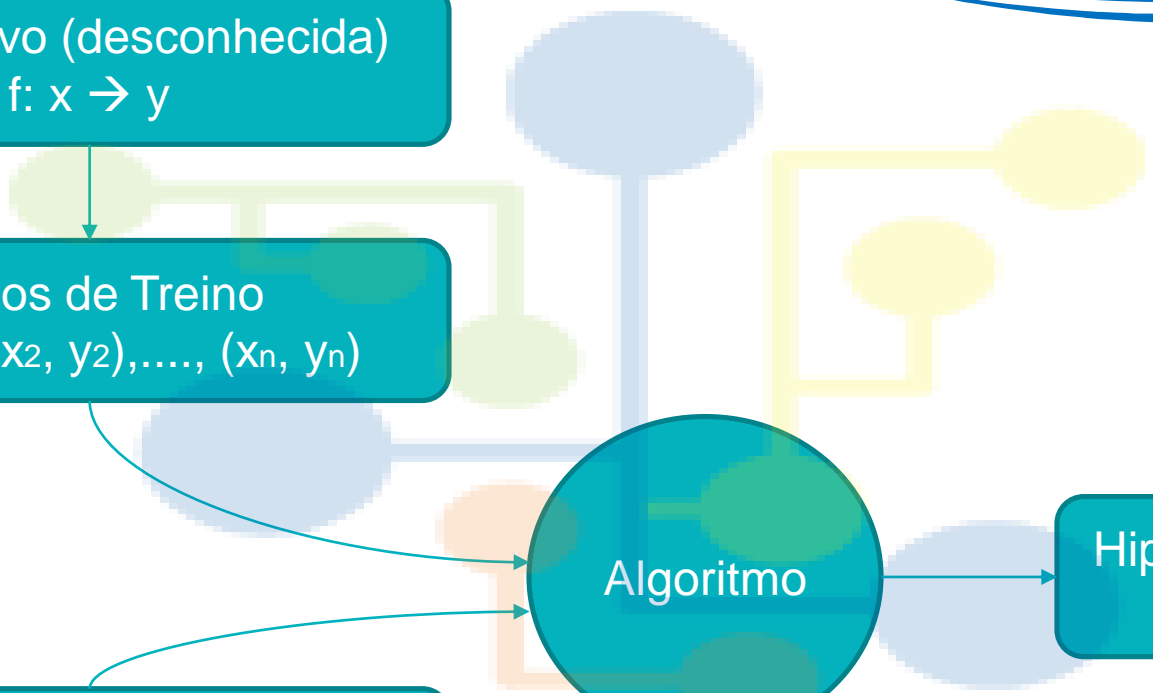
Função alvo (desconhecida)
 $f: x \rightarrow y$

Dados de Treino
 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

Espaço de Hipóteses

Algoritmo

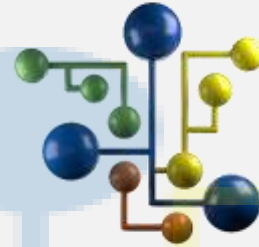
Hipótese Final
 $g \approx f$





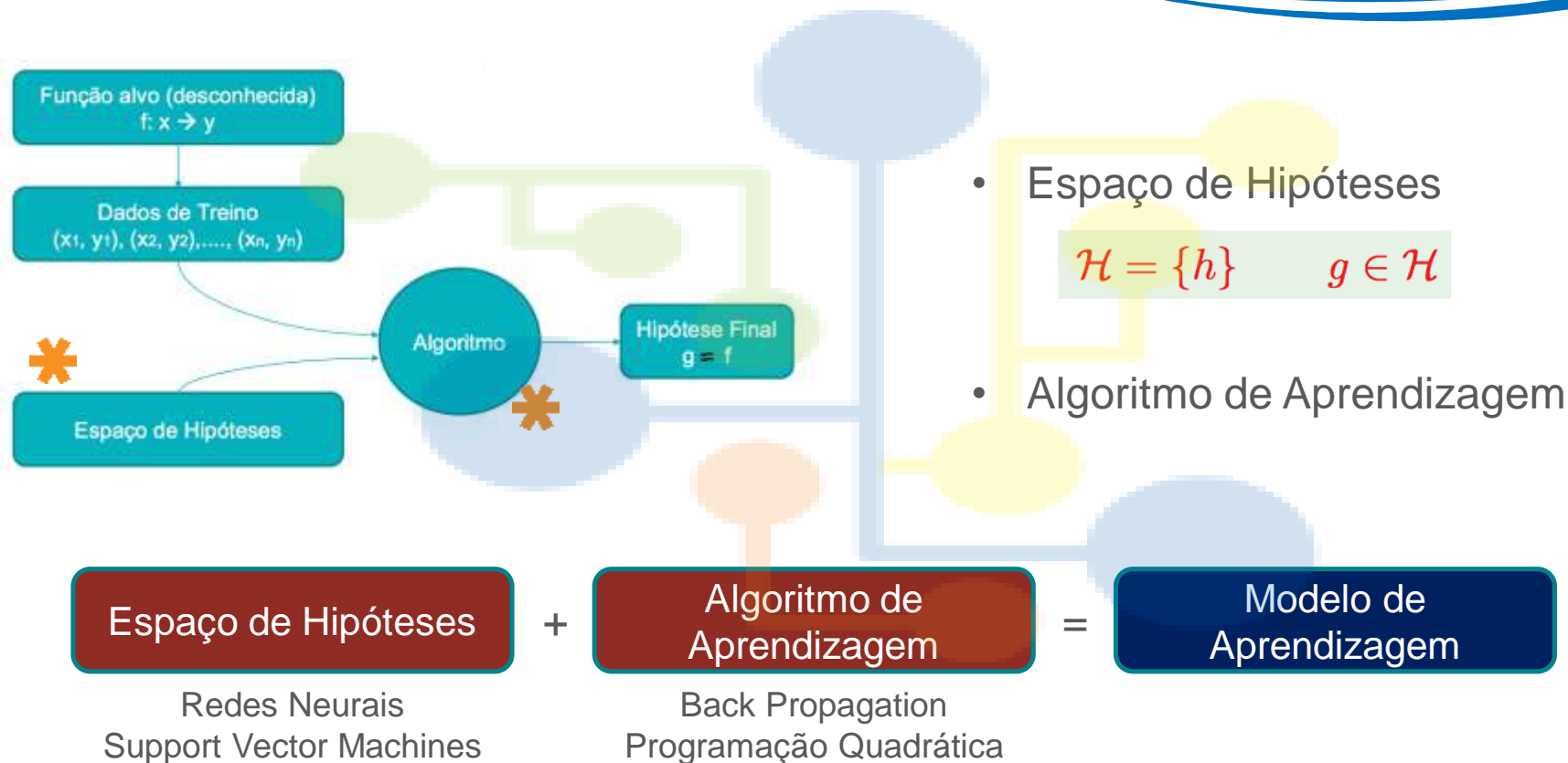
Data Science
Academy

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



**Data Science
Academy**

Modelo de Aprendizagem

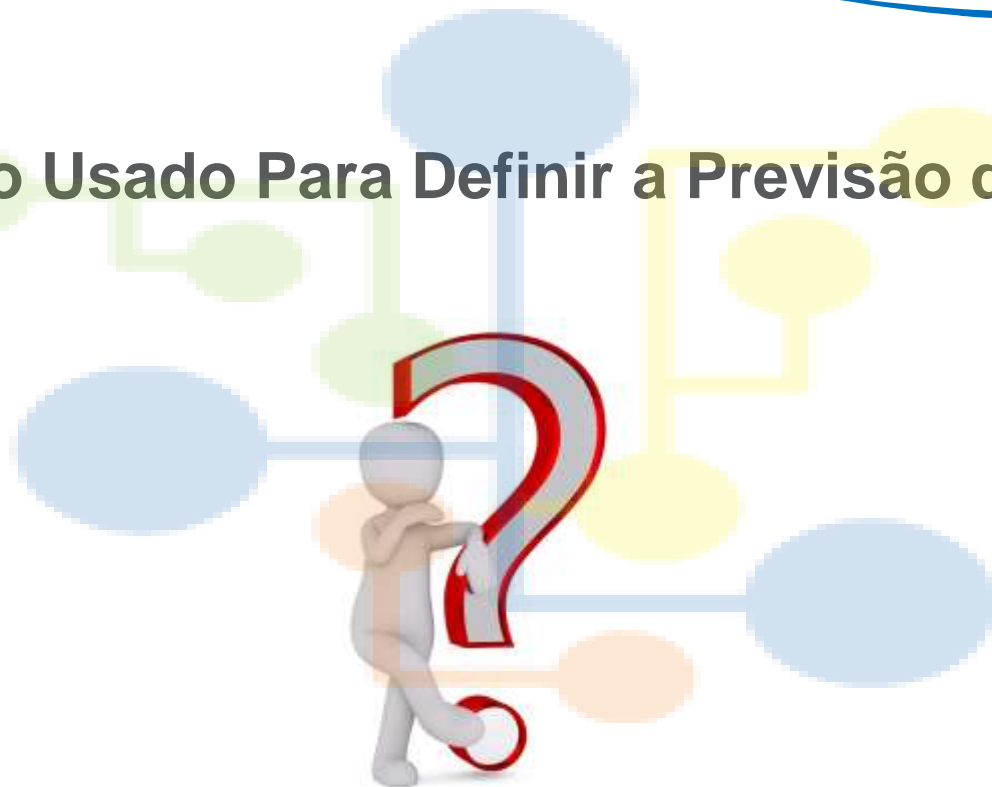




O Espaço de Hipóteses contém os recursos com os quais podemos trabalhar. O Algoritmo de Aprendizagem recebe os dados e navega pelo Espaço de Hipóteses a fim de encontrar a melhor hipótese que gera o resultado desejado.

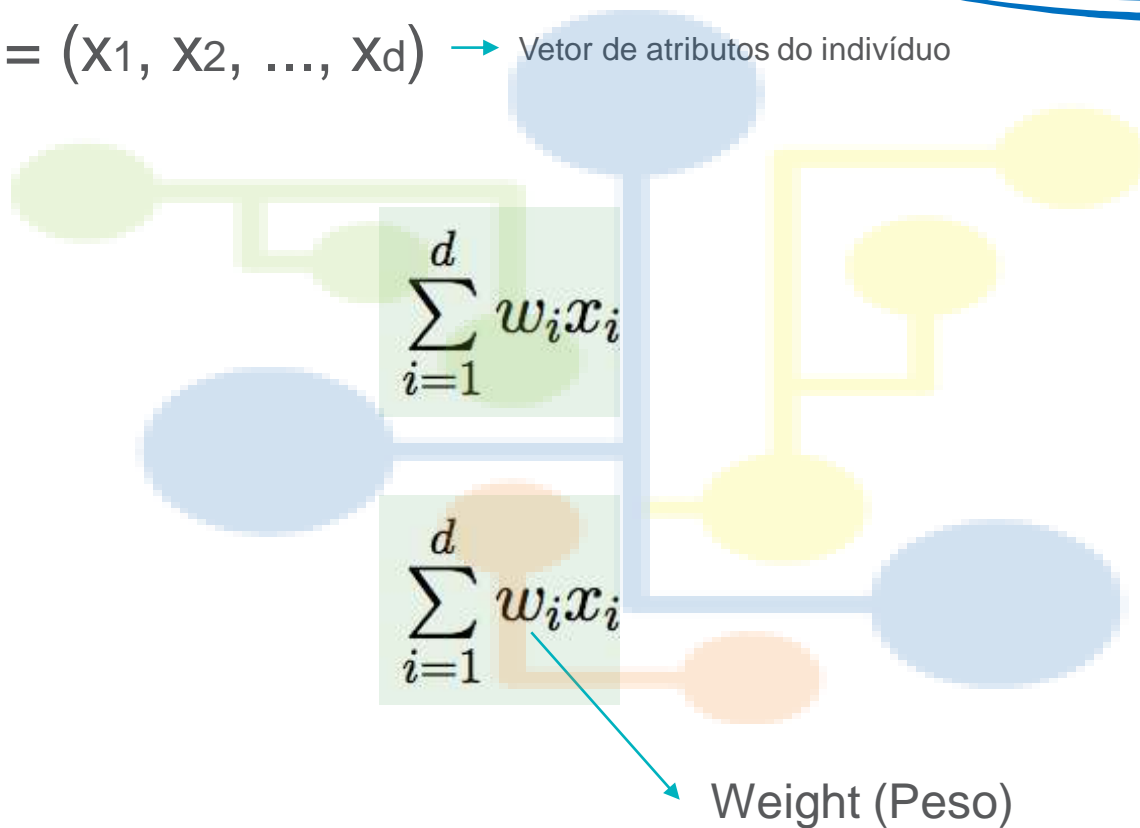


Qual o Critério Usado Para Definir a Previsão do Modelo?





Input $\rightarrow X = (x_1, x_2, \dots, x_d)$ \rightarrow Vetor de atributos do indivíduo





Input $\rightarrow X = (x_1, x_2, \dots, x_d)$

Crédito é **aprovado** se

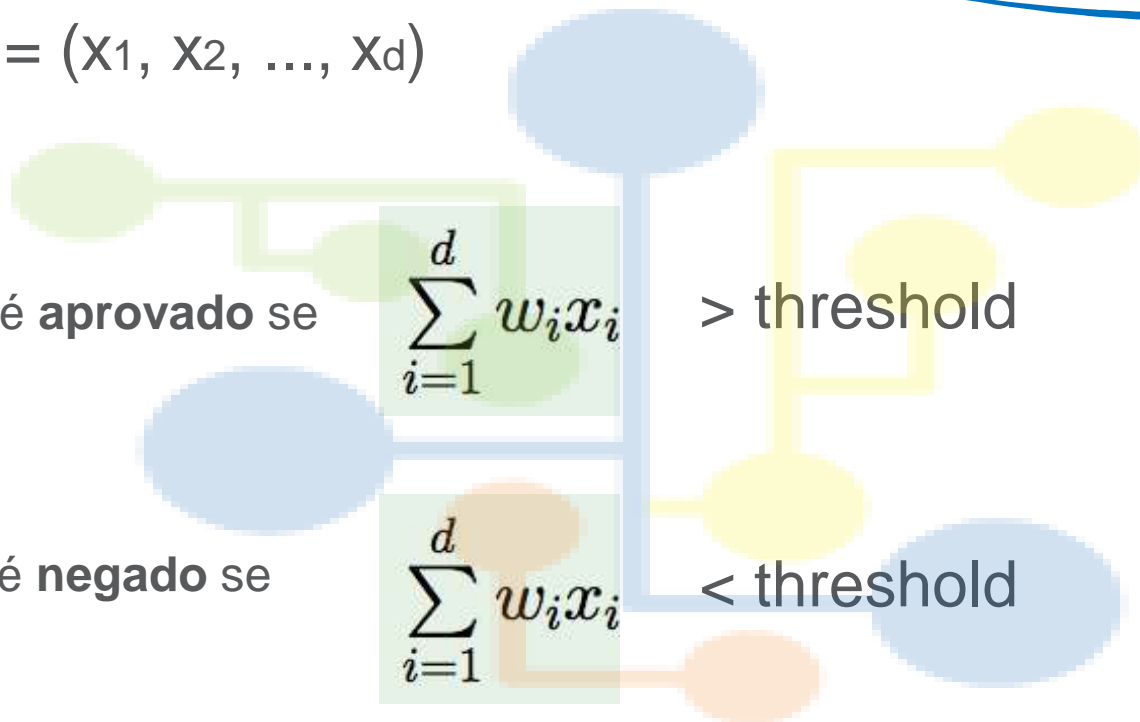
$$\sum_{i=1}^d w_i x_i$$

$> \text{threshold}$

Crédito é **negado** se

$$\sum_{i=1}^d w_i x_i$$

$< \text{threshold}$

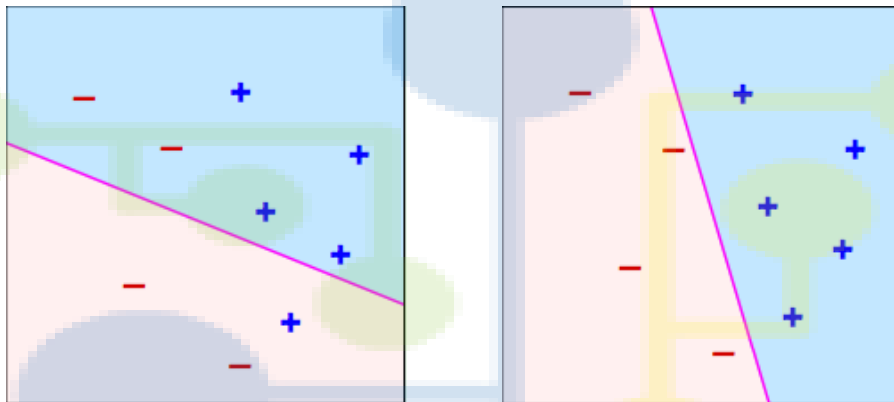




Fórmula que Define as Hipóteses no Espaço de Hipóteses

$$h(\mathbf{x}) = \text{sign} \left(\left(\sum_{i=1}^d w_i x_i \right) - \text{threshold} \right)$$

As diferentes combinações weight/threshold vão formar diferentes hipóteses



As duas linhas em rosa nas imagens acima representam os modelos. Os dois modelos classificam os dados, mas um faz isso melhor do que o outro.

E nosso trabalho, como Cientistas de Dados, é encontrar o melhor modelo possível.



Algoritmo de Aprendizagem

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$$

Dados de Treino

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

O Que Acontece Quando o Modelo Comete Um Erro de Classificação?



Algoritmo de Aprendizagem

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$$

Dados de Treino

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

Erro de Classificação

$$\text{sign}(\mathbf{w}^T \mathbf{x}_n) \neq y_n$$

Ajuste

$$\mathbf{w} \leftarrow \mathbf{w} + y_n \mathbf{x}_n$$



Iteração 1

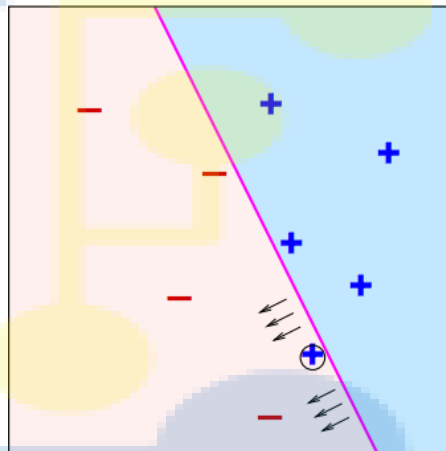
$$\mathbf{w} \leftarrow \mathbf{w} + y_n \mathbf{x}_n$$

Iteração 2

$$\mathbf{w} \leftarrow \mathbf{w} + y_n \mathbf{x}_n$$

•
•
•
•

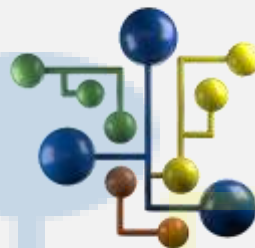
Se os dados forem linearmente separáveis, o algoritmo fará diversas iterações até encontrar a linha que realmente separa as duas classes





Data Science
Academy

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



**Data Science
Academy**

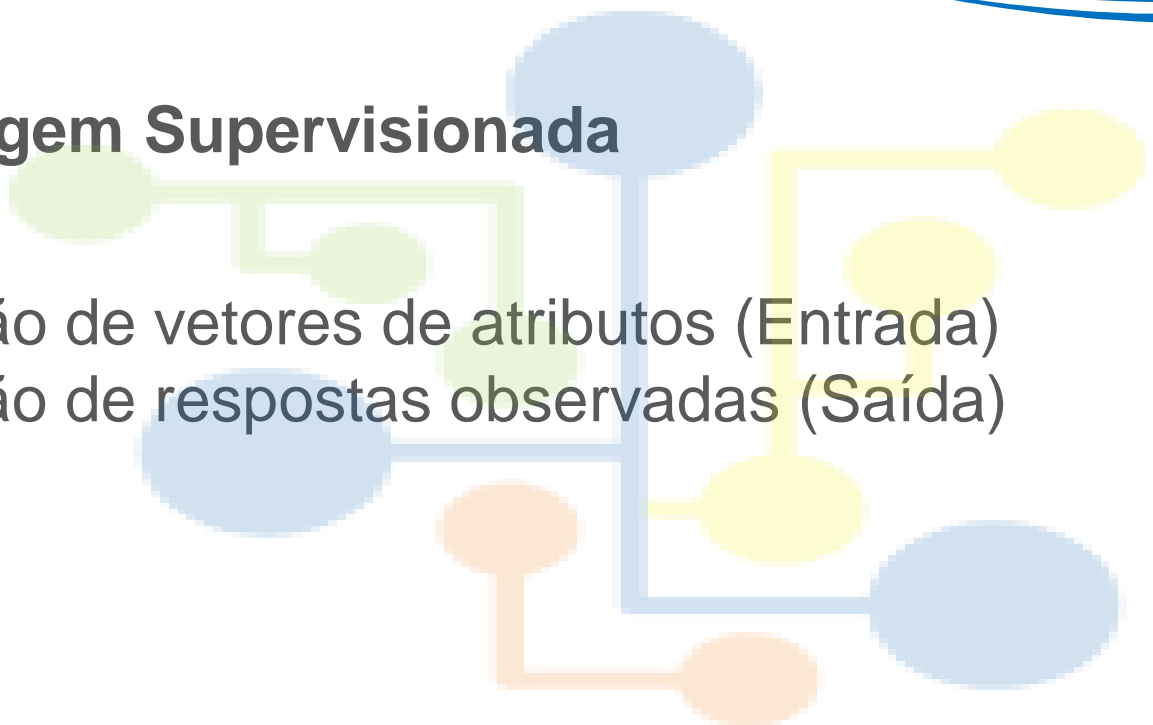
Cost Function (Função de Custo)



Aprendizagem Supervisionada

Coleção de vetores de atributos (Entrada)

Coleção de respostas observadas (Saída)





Aprendizagem Supervisionada

Coleção de vetores de atributos

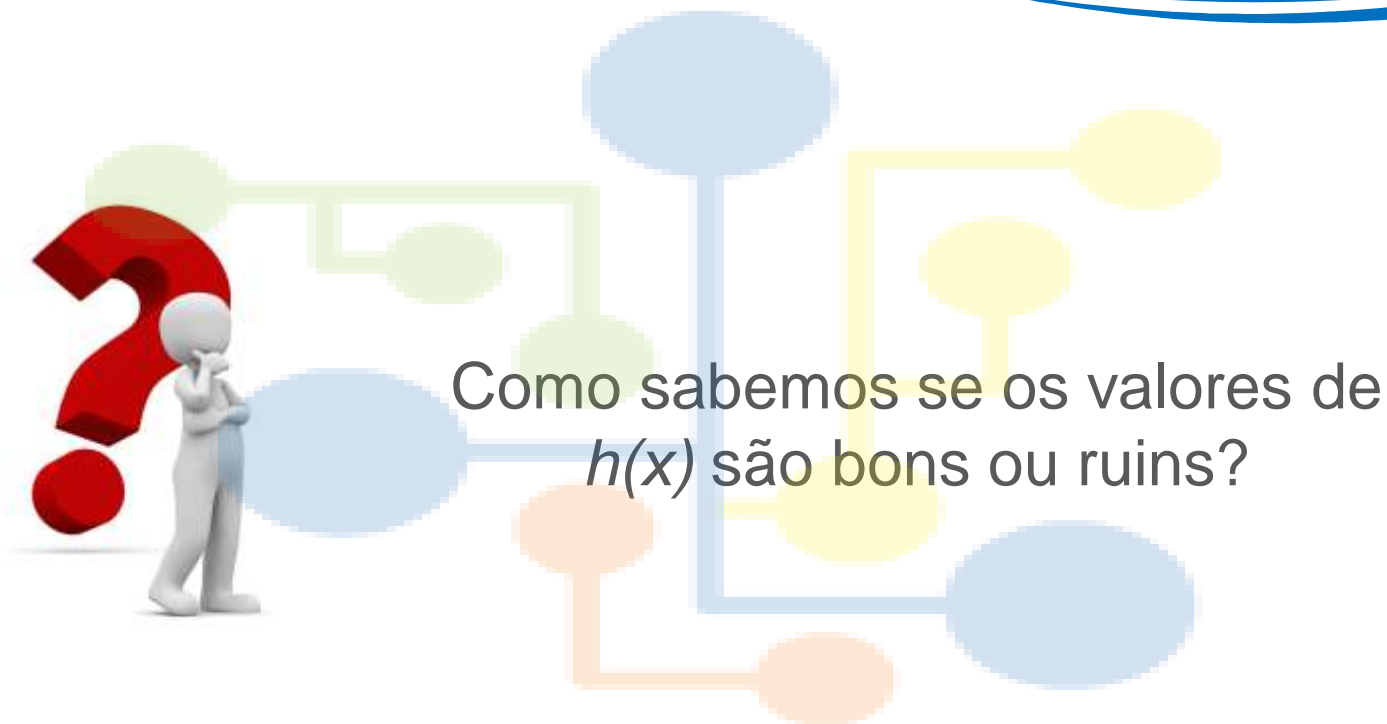
$$\{x_i\}, i = 1, n$$

Coleção de respostas observadas

$$\{y_i\}, i = 1, n$$

Durante o treinamento, construímos uma área de respostas (espaço de hipóteses)

$$h(x)$$



Como sabemos se os valores de $h(x)$ são bons ou ruins?

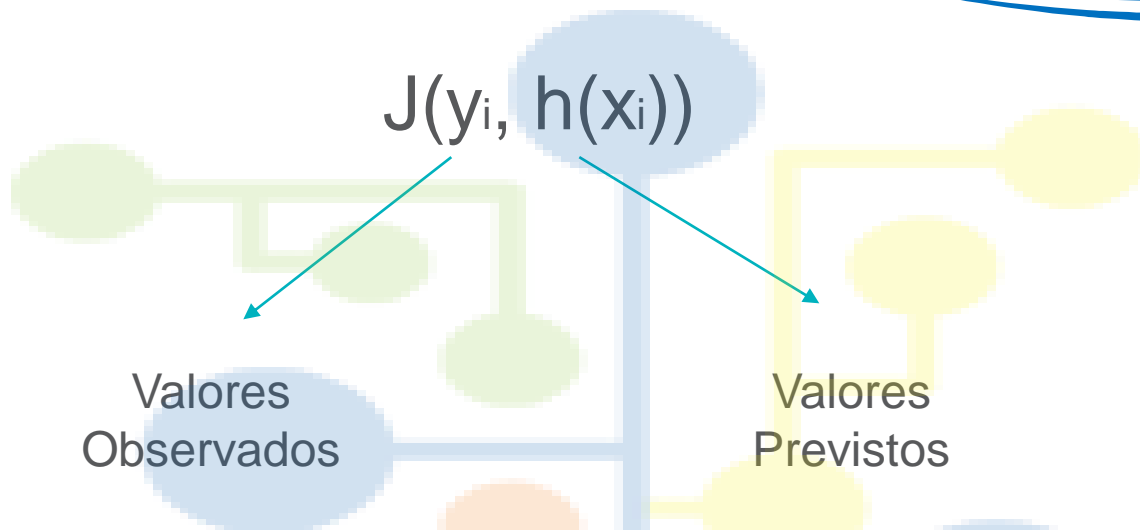


Cost Function

Descreve quão bem resposta na área de respostas (espaço de hipóteses) se encaixa no conjunto de dados que está sendo analisado.

$h(x)$

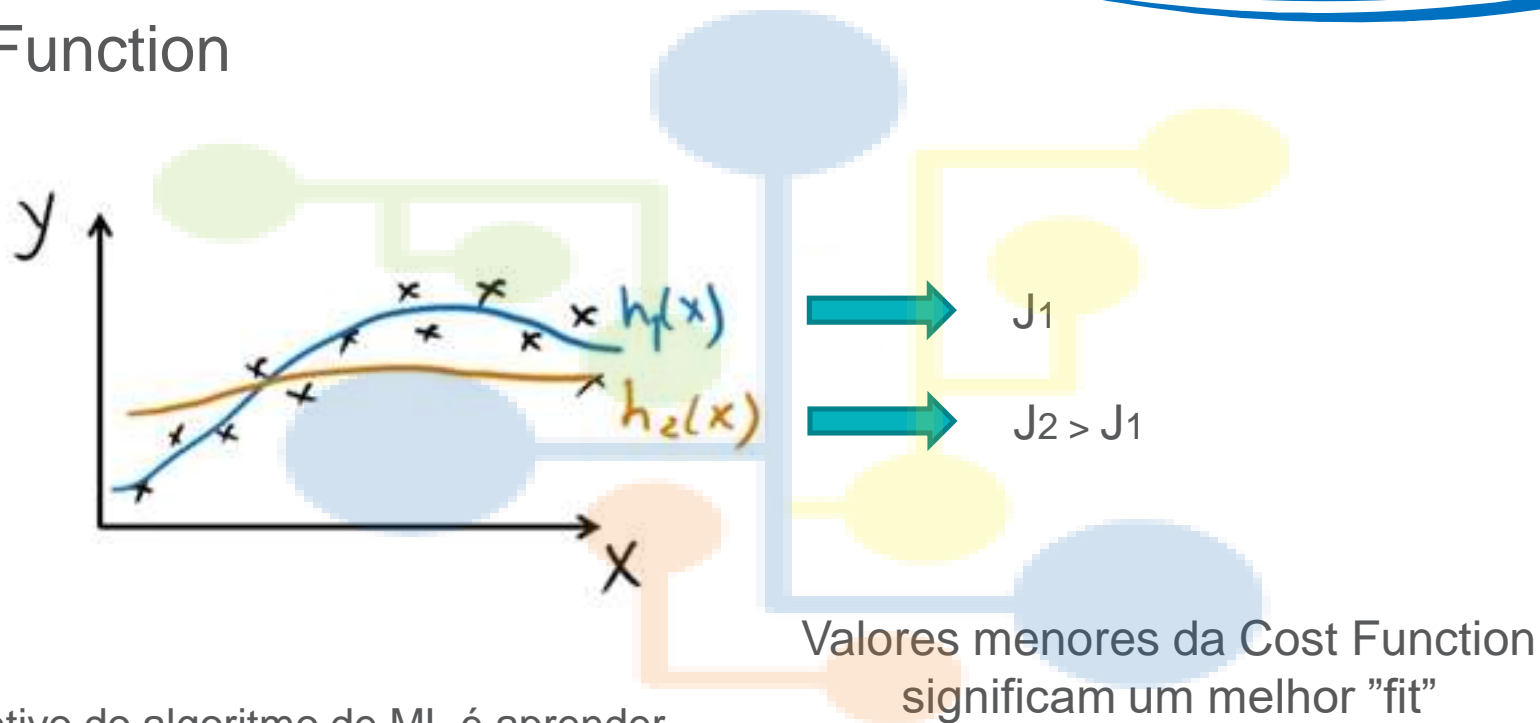
$J(y_i, h(x_i))$



A Cost Function é um número que melhor representa a relação entre valores observados e valores previstos. Em outras palavras: é a diferença entre o que deveria ser previsto pelo modelo e aquilo que ele realmente previu!



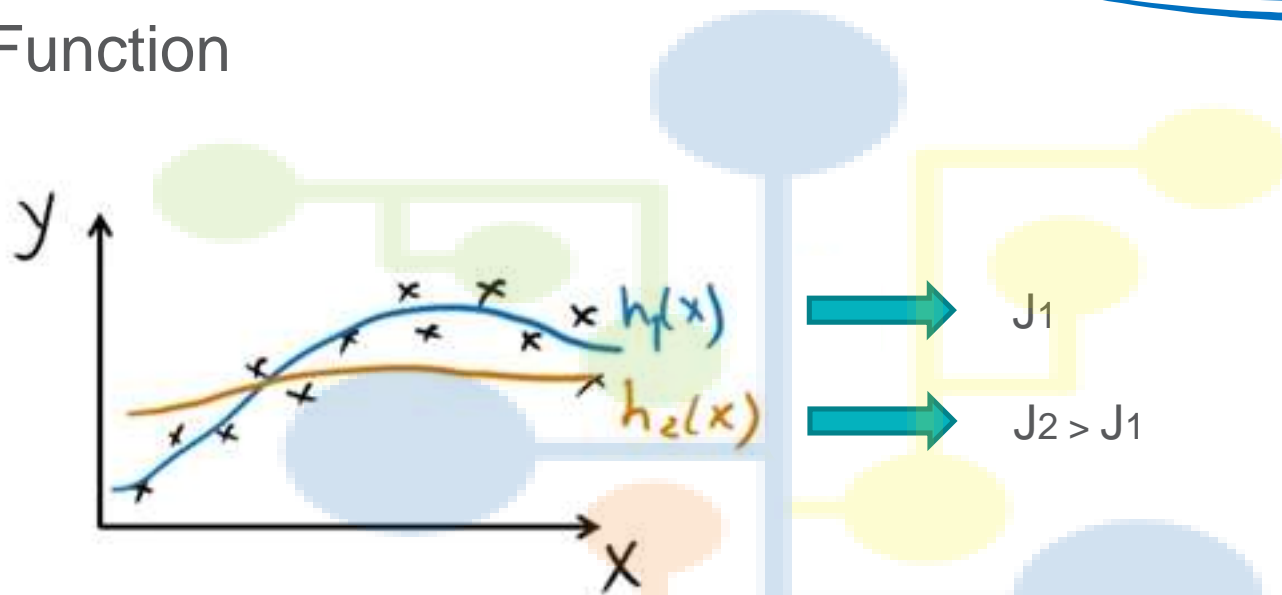
Cost Function



O objetivo do algoritmo de ML é aprender um modelo que minimize os erros



Cost Function



Um dos objetivos em Machine Learning é construir $h(x)$ de modo que o valor de J seja minimizado



Em problemas de regressão, $h(x)$ é normalmente interpretada diretamente como a resposta a ser prevista.



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

Parameters: θ_0, θ_1

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

Goal: minimize $J(\theta_0, \theta_1)$
 θ_0, θ_1



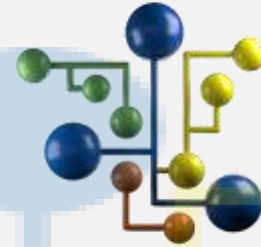
Comparando uma previsão contra o seu valor real, usando uma cost function, determinamos o nível de erro do modelo.

Por ser uma formulação matemática, a cost function expressa o nível de erro em uma forma numérica. A cost function transmite o que é realmente importante e significativo para seus propósitos com o algoritmo de aprendizagem.



Data Science
Academy

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



**Data Science
Academy**

Gradiente Descendente



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

Objetivo:

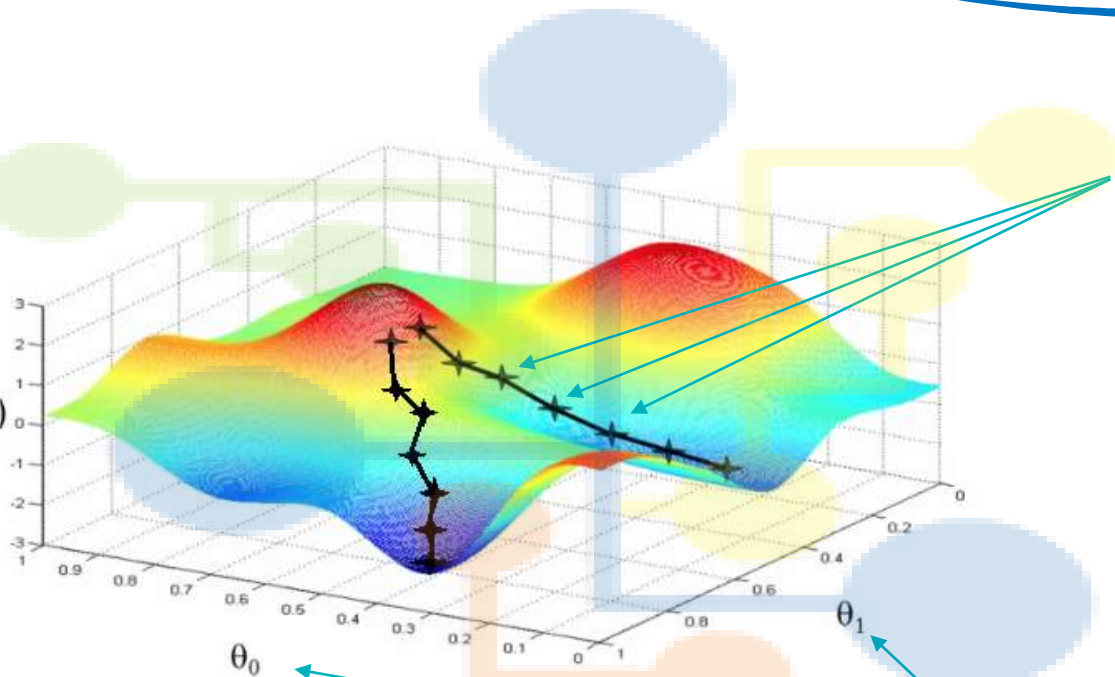
$$\underset{\theta_0, \theta_1}{\text{minimize}} J(\theta_0, \theta_1)$$



Cost Function

$J(\theta_0, \theta_1)$

Gradiente
Descendente



Quanto melhor os valores de
parâmetros, menor o valor de J.

Parâmetros da
Cost Function

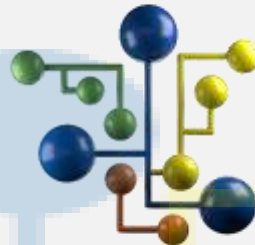


Gradiente Descendente é ideal quando os parâmetros não podem ser calculados analiticamente (por exemplo, usando álgebra linear) e devem ser pesquisados por um algoritmo de otimização.



Data Science
Academy

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



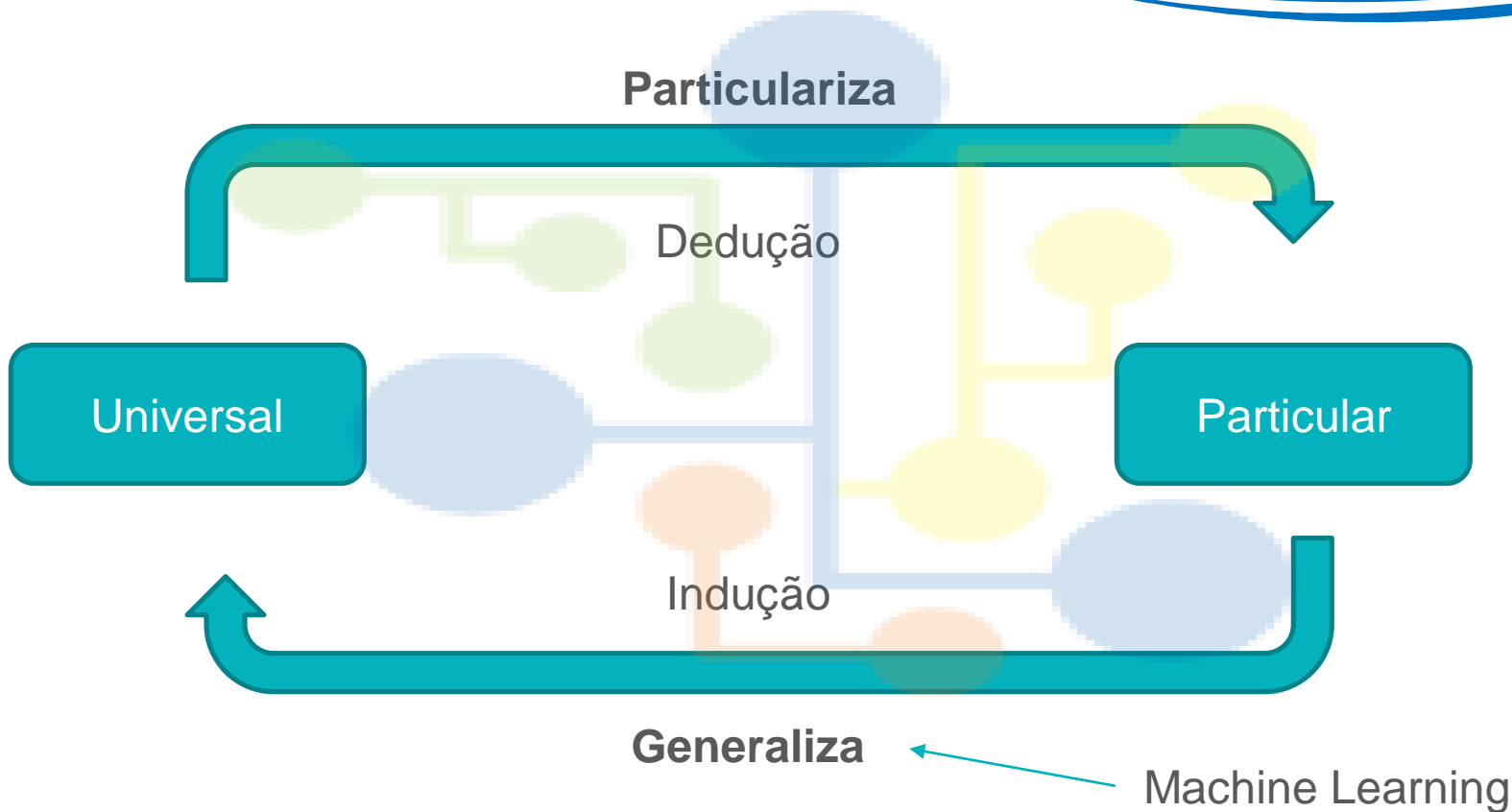
**Data Science
Academy**

Overfitting x Underfitting



Aprendizagem Supervisionada

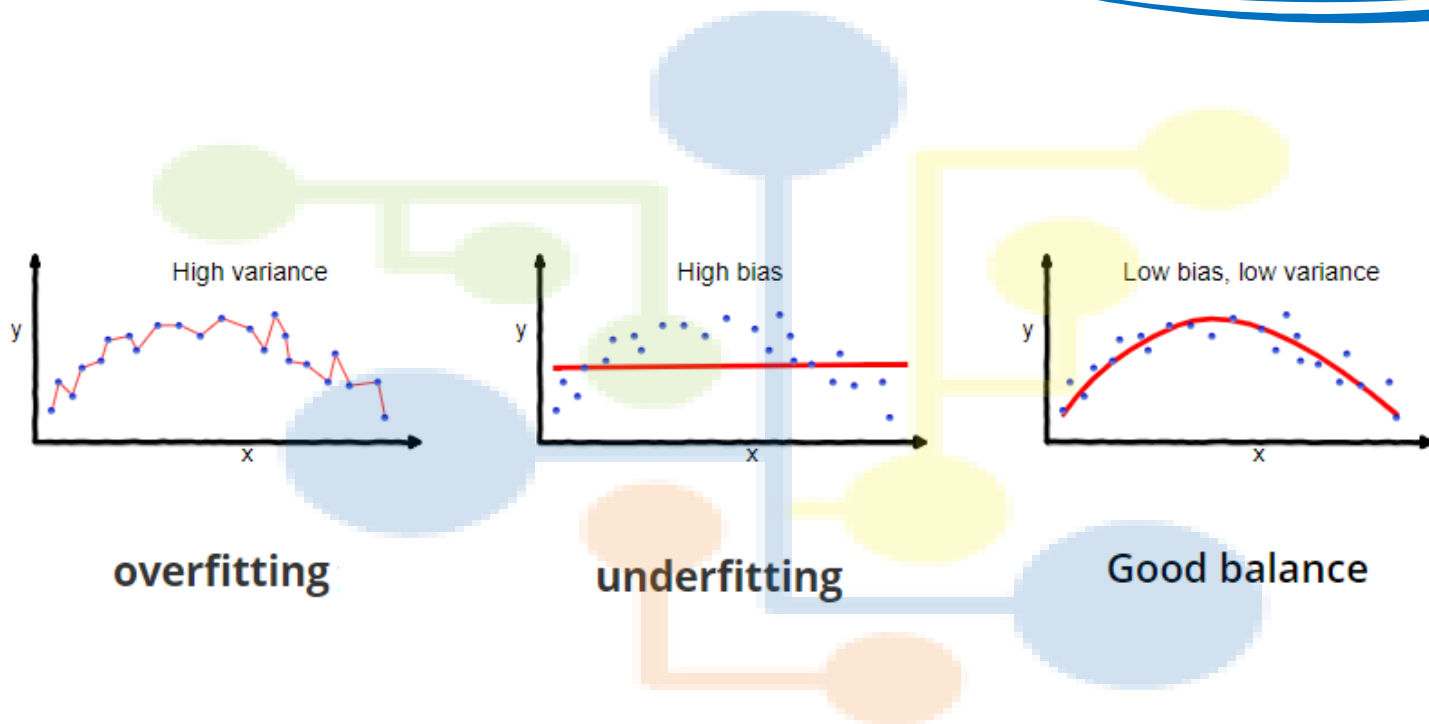
$$Y = f(X)$$





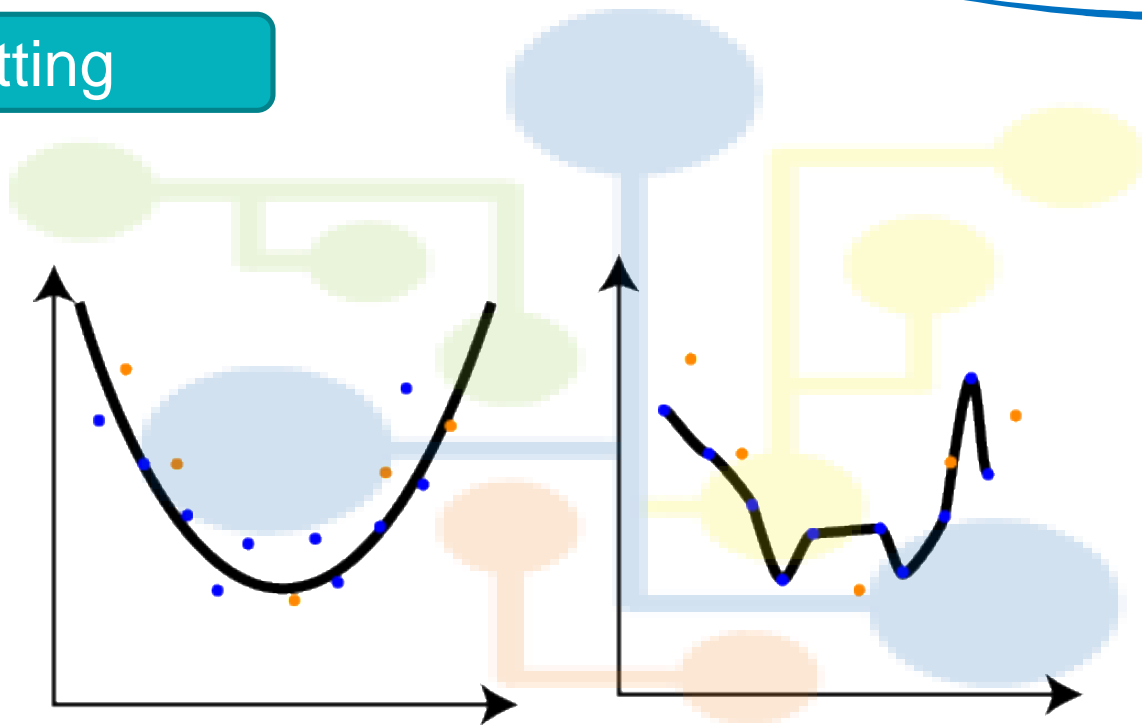
Generalização refere-se a quão bem os conceitos aprendidos por um modelo de aprendizado de máquina se aplicam a exemplos específicos não vistos pelo modelo durante o processo de aprendizado.





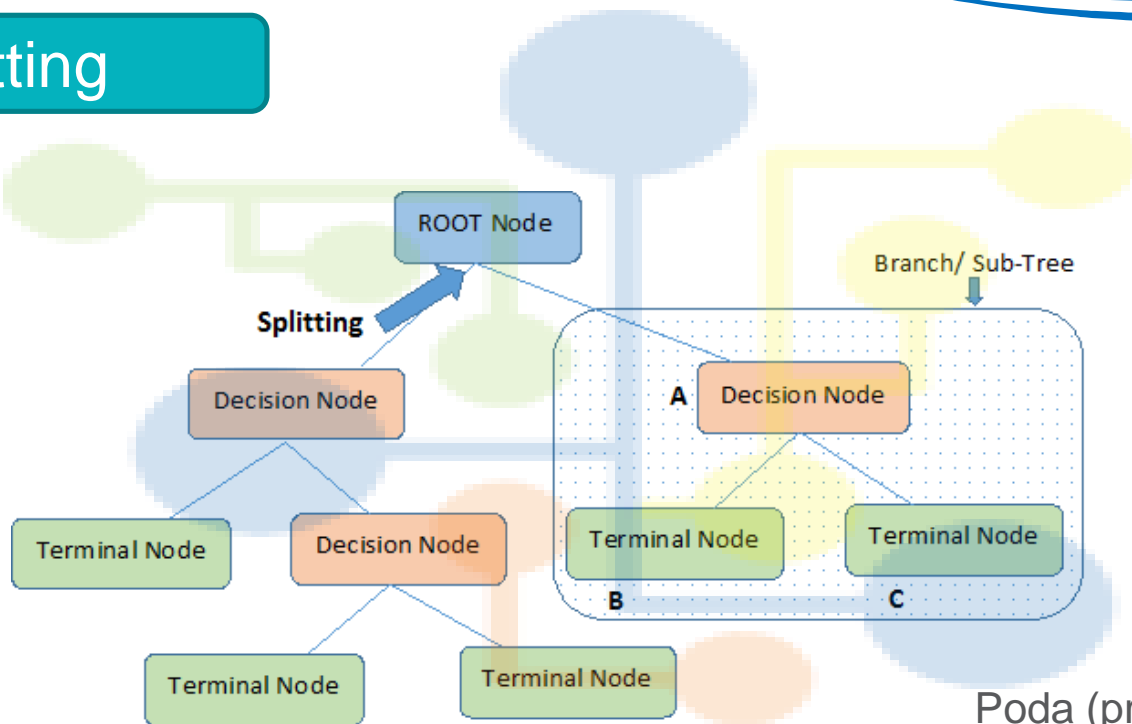


Overfitting





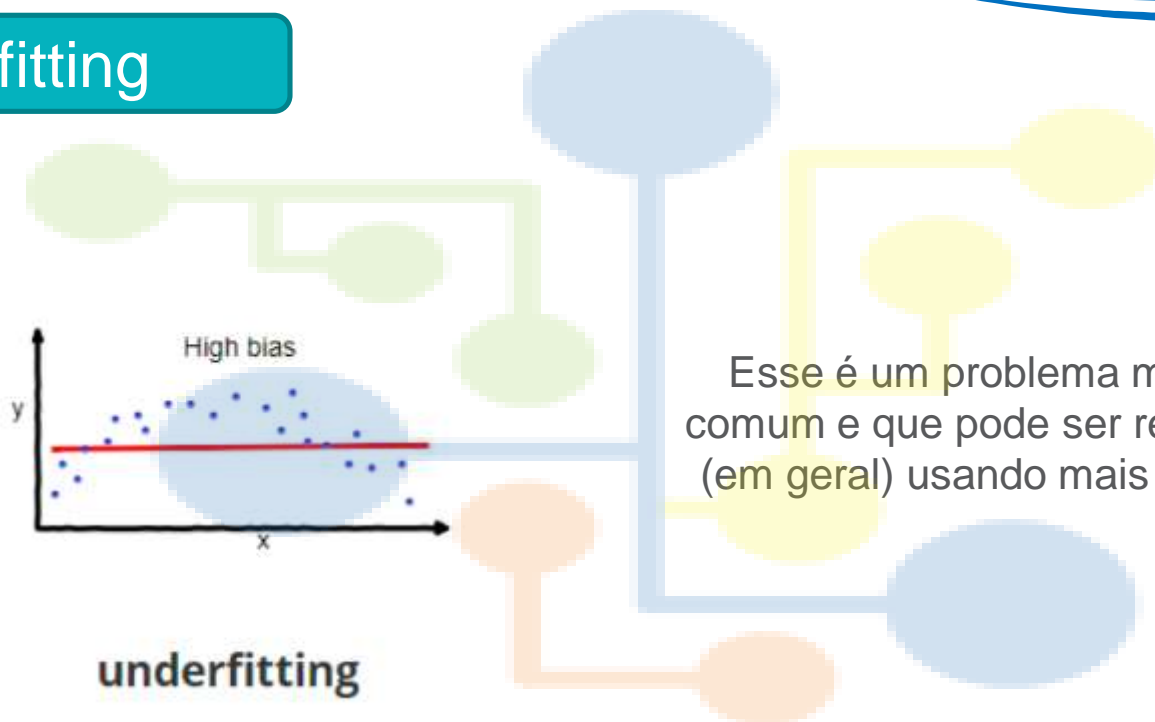
Overfitting



Poda (pruning) da árvore de decisão para evitar overfitting.

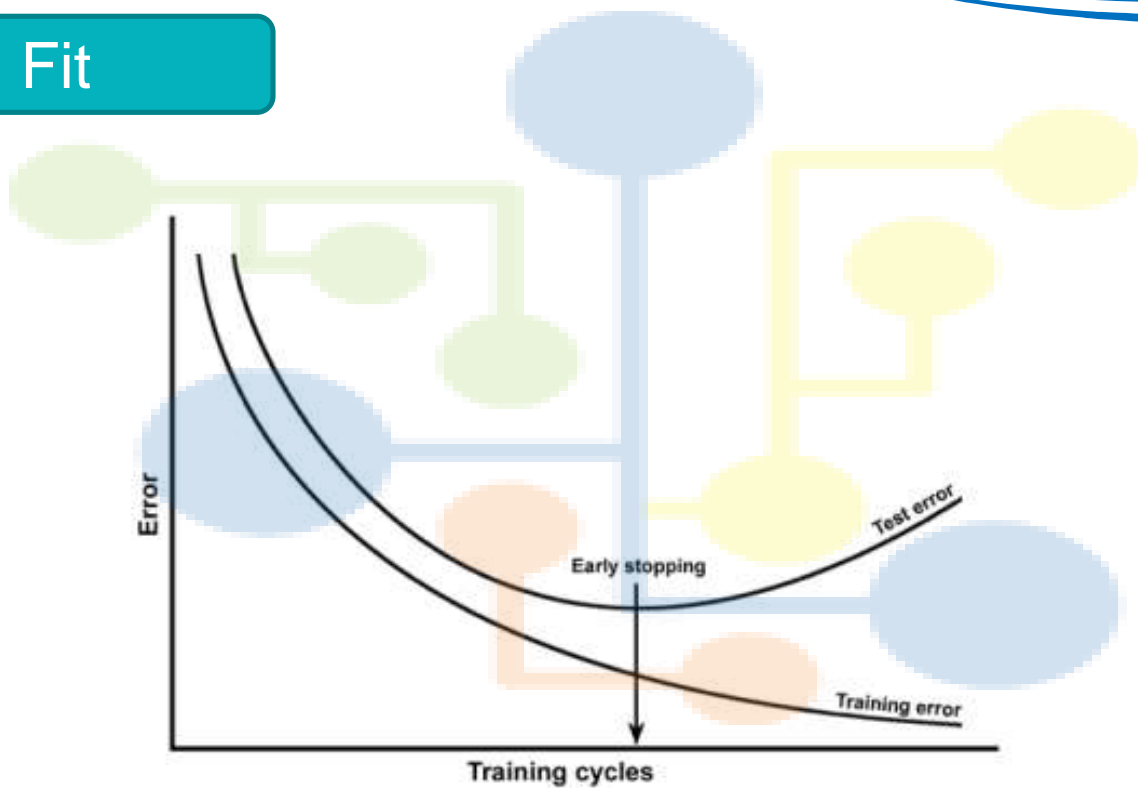


Underfitting



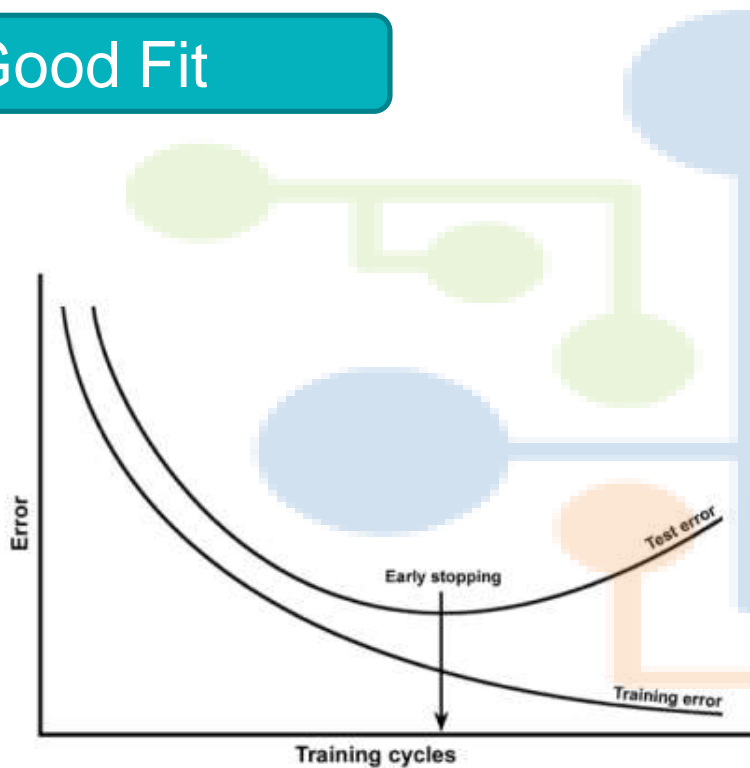


Good Fit





Good Fit



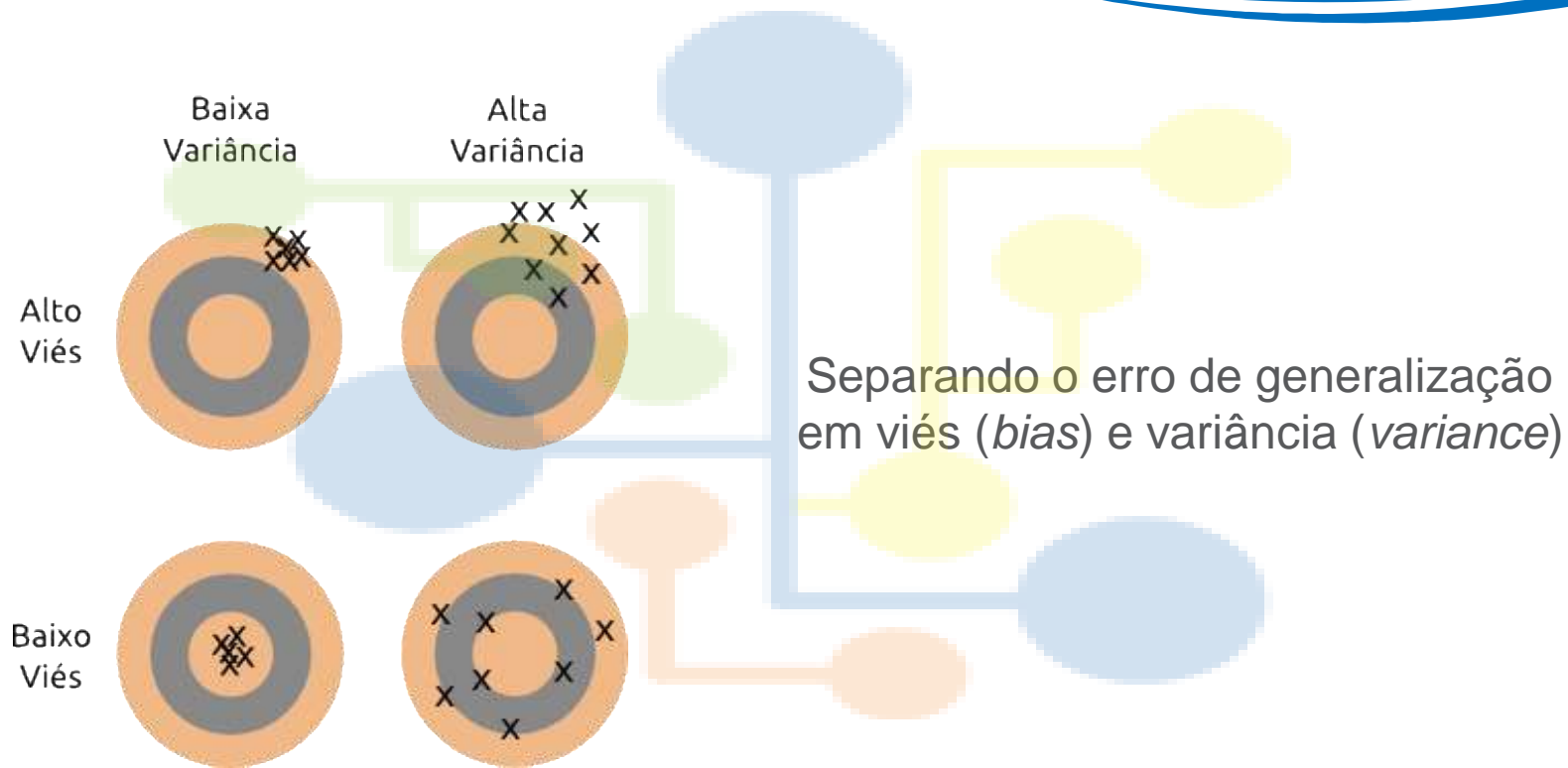
- Reamostragem
- Conjuntos de Dados de Validação
- Early-Stopping

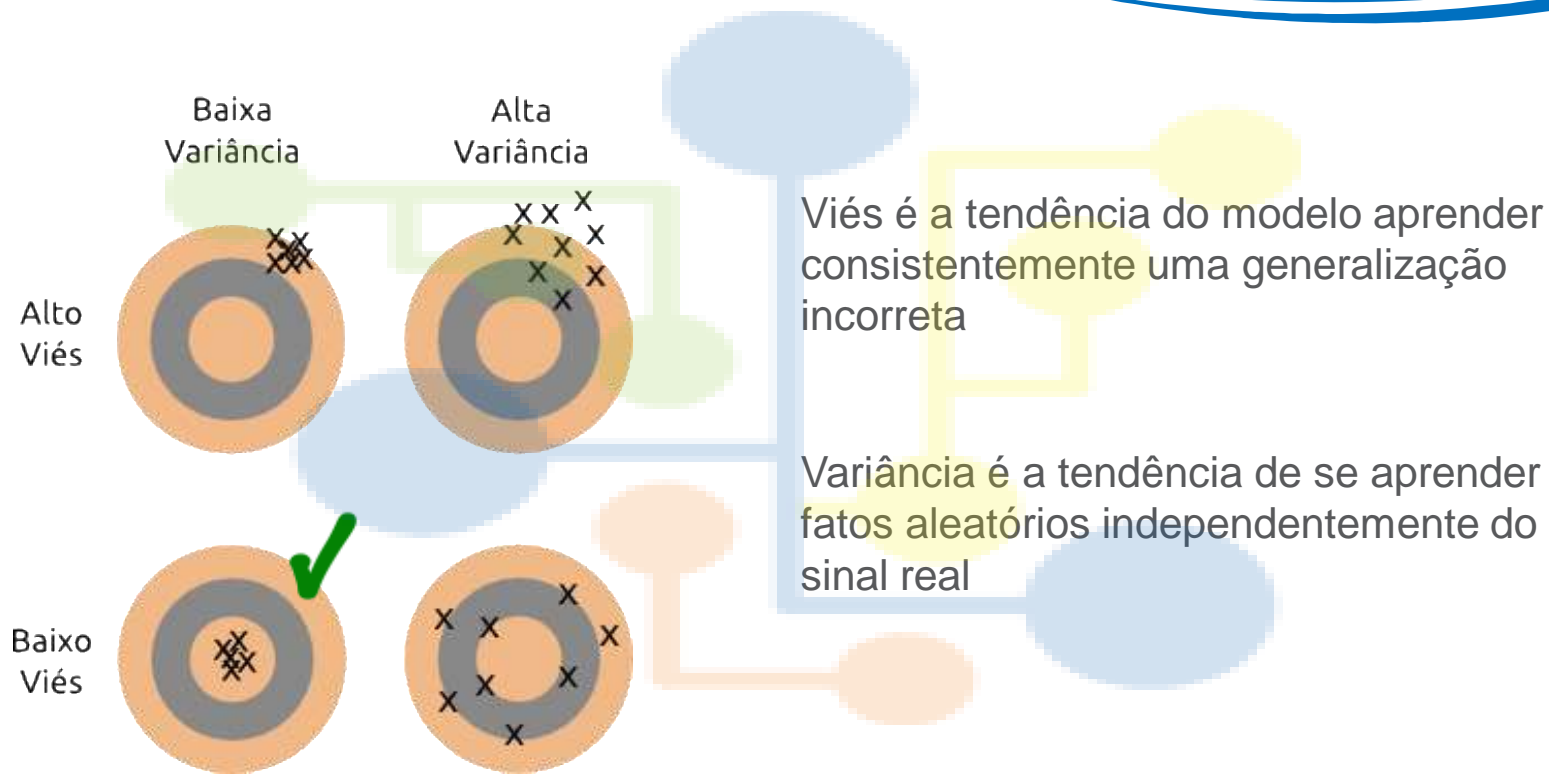


Data Science
Academy

Bias (Viés) e Variância









Erro de Previsão de um Modelo

$$E[(y - \hat{f}(x))^2] = Bias[\hat{f}(x)]^2 + Var[\hat{f}(x)] + \sigma^2$$

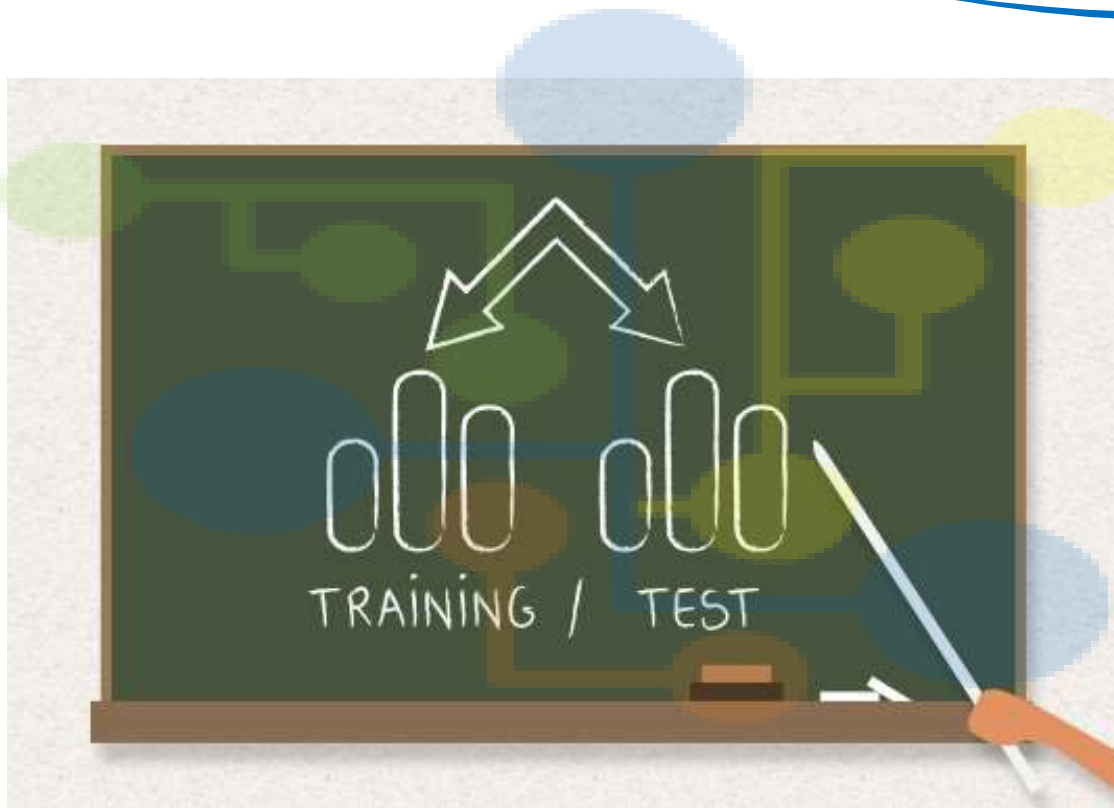
Bias

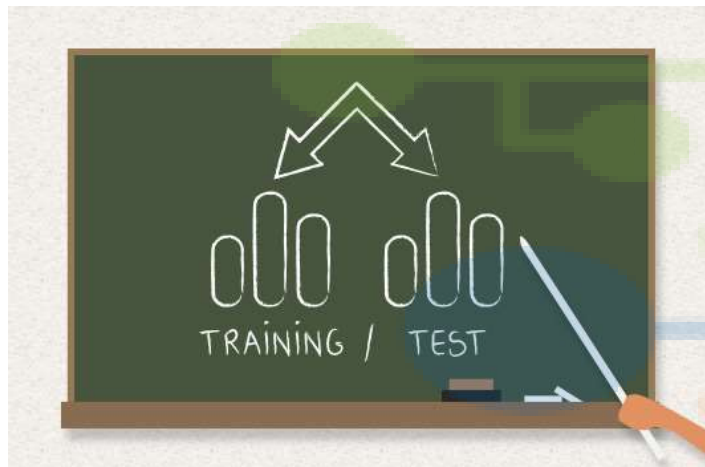
$$Bias[\hat{f}(x)] = E[\hat{f}(x) - f(x)]$$

Variance

$$Var[\hat{f}(x)] = E[\hat{f}(x)^2] - E[\hat{f}(x)]^2$$



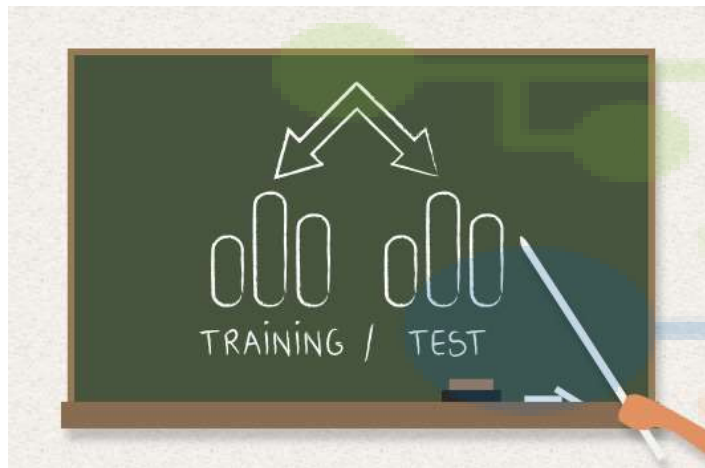




Utilizar um modelo complexo que é capaz de reduzir consideravelmente o erro de previsão no dataset de treino, mas ao mesmo tempo não é tão generalizável a ponto de apresentar um bom resultado no dataset de teste

Viés

Variância



Utilizar um modelo simples que é bem generalizável, mas não reduz consideravelmente o erro de previsão no dataset de treino

Viés

Variância





Tradeoff

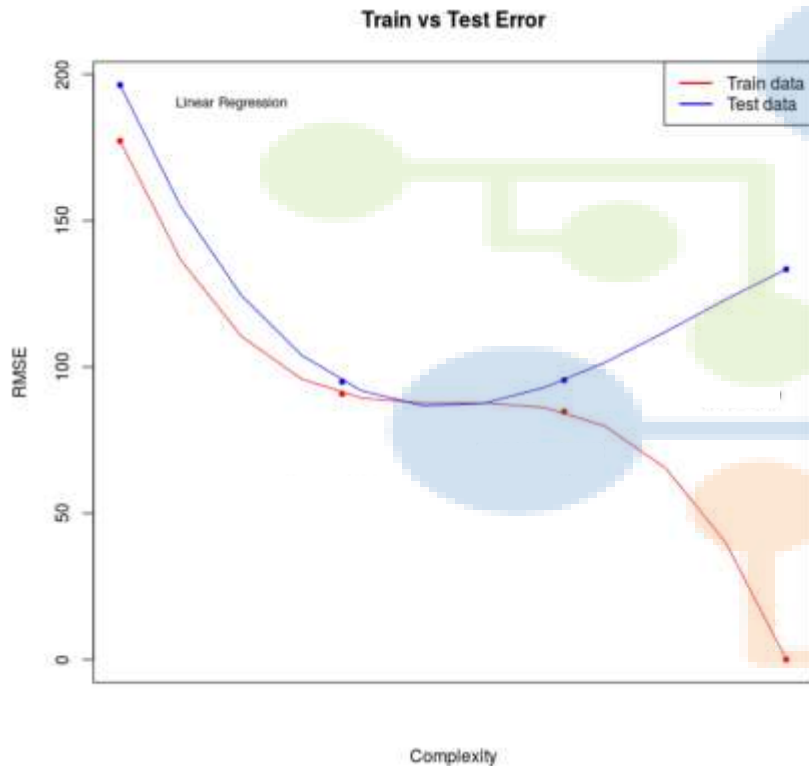




Modelos mais simples tem
viés alto mas variância
baixa (underfitting)

Modelos mais complexos
tem viés baixo mas
variância alta (overfitting)





A tarefa essencial de previsão é selecionar um modelo que se aproxime do ponto mínimo da curva de erro do dataset de teste





Continue Trilhando uma Excelente Jornada de Aprendizagem!

Muito Obrigado!