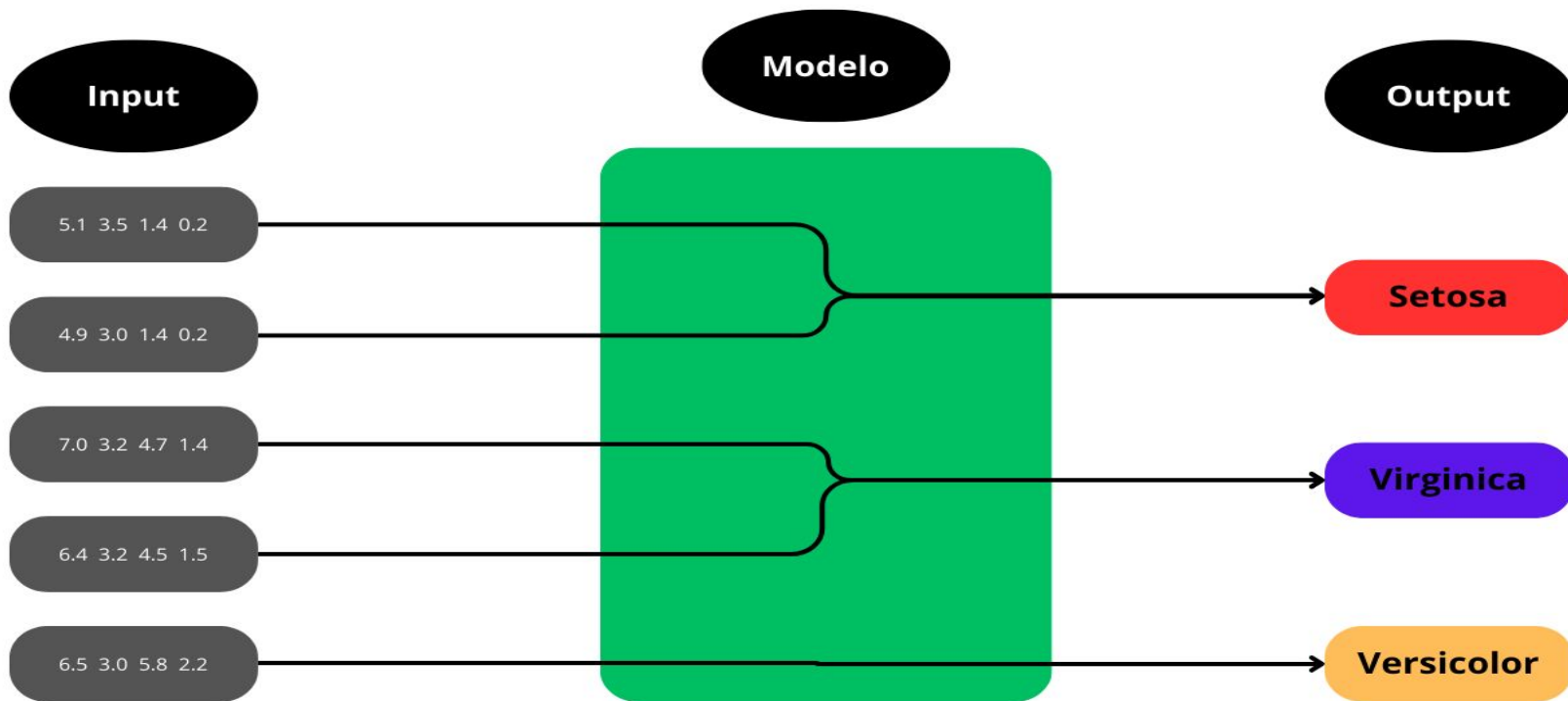




Introdução ao Machine Learning

Vitor A. Oliveira e Gabriel Rosa

Modelo



Treinamento

Dados de
treino

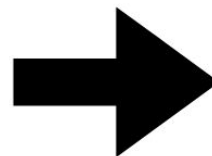
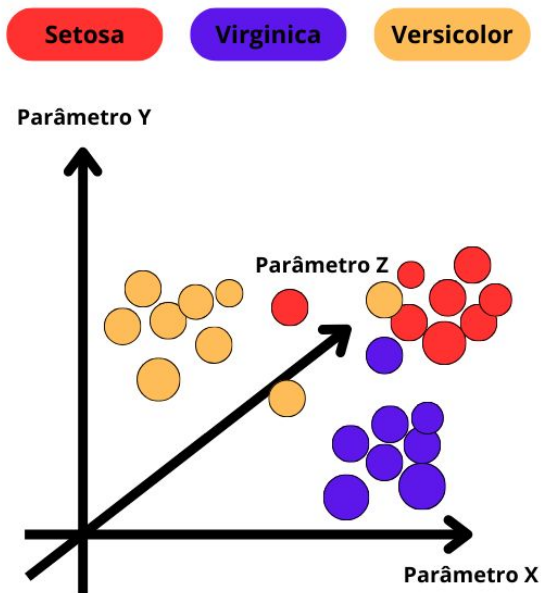
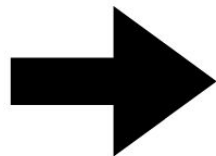
5.0 3.6 1.4 0.2 setosa

4.4 2.9 1.4 0.2 setosa

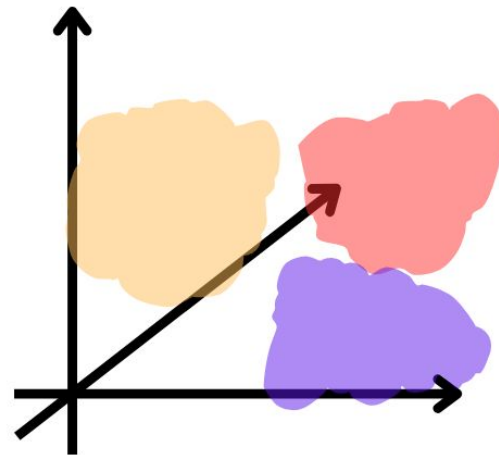
6.6 2.9 4.6 1.3 versicolor

6.2 2.8 4.8 1.8 virginica

•
•
•



Generalização



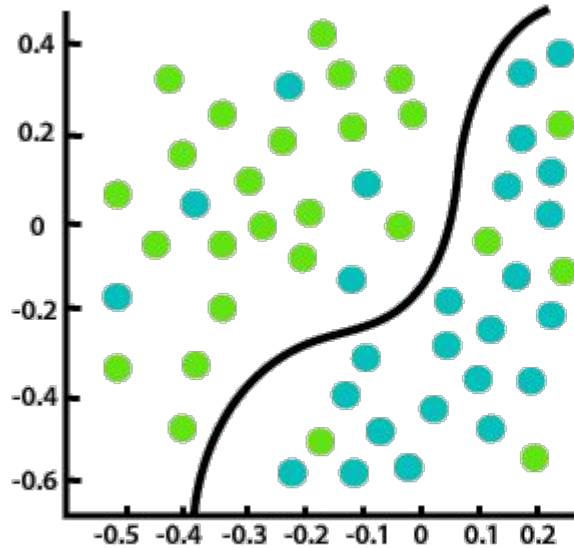
Aprendizado Supervisionado

```
iris = load_iris()
```

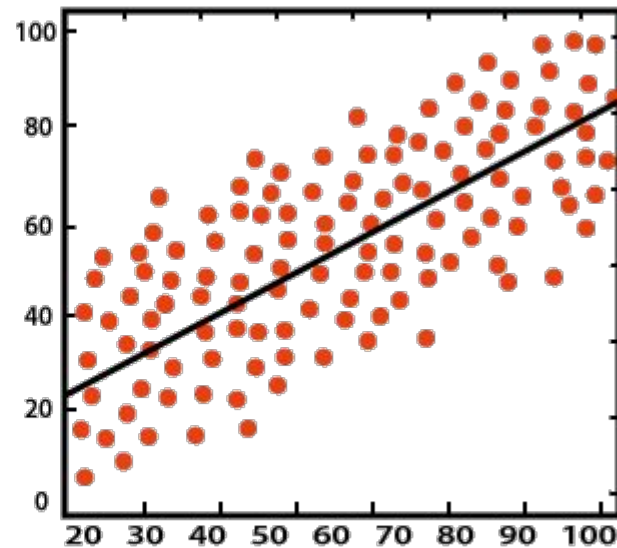
```
X = iris.data # Parâmetros
```

```
y = iris.target # Rótulos
```

Comprimento da sépala	Largura da sépala	Comprimento da pétala	Largura da pétala	Espécie
4.9	3.0	1.4	0.2	setosa
4.6	3.4	1.4	0.3	setosa
7.0	3.2	4.7	1.4	versicolor
5.7	2.8	4.5	1.3	versicolor
7.9	3.8	6.4	2.0	virginica



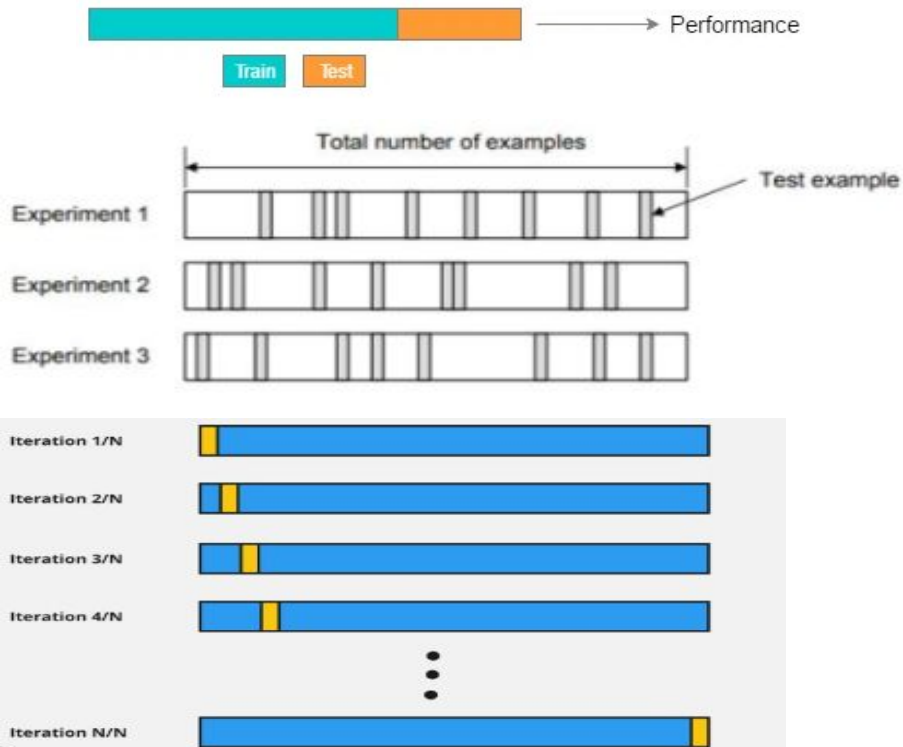
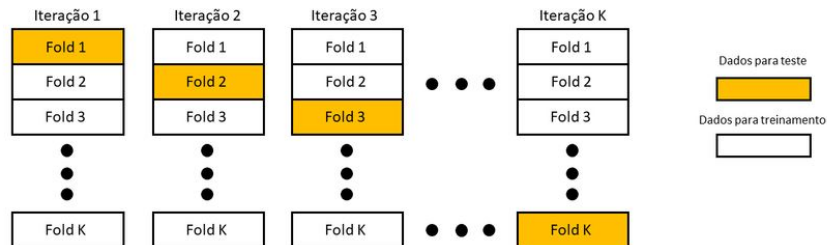
Classificação



Regressão

Técnicas de Validação de Modelos

Resubstituição - Utilização de todo o dataset para treinamento e testagem



Matriz de Confusão



Valores previstos

Valores reais

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Matriz de Confusão



Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Tabela Relacionada a Setosa

VP 5	

Valores reais

Matriz de Confusão



Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Valores reais

Tabela Relacionada a Setosa

VP 5	
	VN $3+1+1+5 = 10$

Matriz de Confusão



Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Tabela Relacionada a Setosa

VP 5	FN $0+0 = 0$
	VN $3+1+1+5 = 10$

Valores reais

Matriz de Confusão



Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Tabela Relacionada a Setosa

VP 5	FN $0+0 = 0$
FP $0+0 = 0$	VN $3+1+1+5 = 10$

Valores reais

Matriz de Confusão



Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Valores reais

Tabela Relacionada a Versicolor

VP 3	FN $0+1 = 1$
FP $0+1 = 1$	VN $5+0+0+5 = 10$

Matriz de Confusão



Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Tabela Relacionada a Virginica

VP 5	FN $0+1 = 1$
FP $0+1 = 1$	VN $5+0+0+3 = 8$

Valores reais

Métricas de Avaliação de Modelos



Acurácia -> *O quão certo o modelo está? $(VP+VN)/Total$*

Precisão -> *De todos os dados classificados, quais estão realmente corretos? $VP/(VP+FP)$*

Revocação -> *Qual a porcentagem de dados classificados positivos comparado a quantidade real de positivos da amostra? $VP/(VP+FN)$*

F1-Score -> *Une a precisão e a revocação para determinar um valor que represente a qualidade geral do modelo. $2 * Precisão * Revocação / (Precisão + Revocação)$*

Acurácia

Valores previstos

Valores reais

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

VPs/

13/

Acurácia

Valores previstos

Valores reais

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

VPs/Total

$$13/15 = 0.833$$

Valores reais

Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Precisão Versicolor

$$VP/VP+FP$$

$$3/3+1 = 0.75$$

Tabela Relacionada a Versicolor

VP 3	FN 0+1 = 1
FP 0+1 = 1	VN 5+0+0+5 = 10

Valores reais

Valores previstos

	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Revocação Versicolor

$$VP/VP+FN$$

$$3/3+1 = 0.75$$

Tabela Relacionada a Versicolor

VP 3	FN 0+1 = 1
FP 0+1 = 1	VN 5+0+0+5 = 10



F1-score Versicolor

$$\frac{2 * \text{precisão} * \text{revocação}}{\text{precisão} + \text{revocação}}$$

$$\frac{2 * 0.75 * 0.75}{0.75 + 0.75} = 0.75$$



Valores reais

Valores previstos

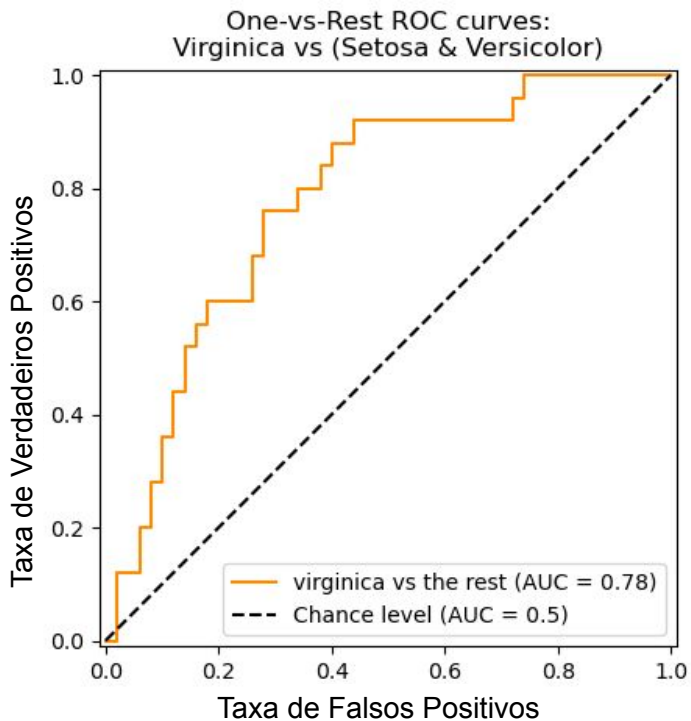
	Setosa	Versicolor	Virginica
Setosa	5	0	0
Versicolor	0	3	1
Virginica	0	1	5

Taxa de verdadeiros positivos

$$\text{VP} / \text{VP} + \text{FN}$$

Taxa de falsos positivos

$$\text{VN} / \text{VN} + \text{FP}$$



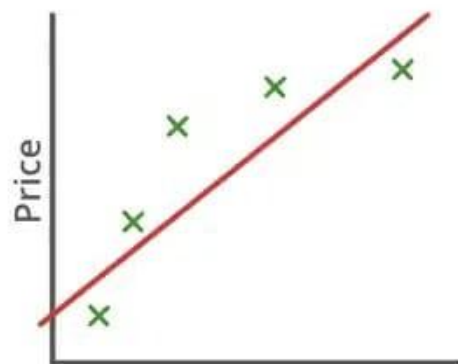
Overfitting e Underfitting



Overfitting é o treinamento excessivo de um modelo em um dataset ao ponto que ele começa a identificar os “ruídos” específicos daquele dataset.

Underfitting é o oposto do Overfitting, ocorre quando o modelo treina pouco em qualquer base de dados, seja por redução de complexidade ou por interromper o treinamento antecipadamente.

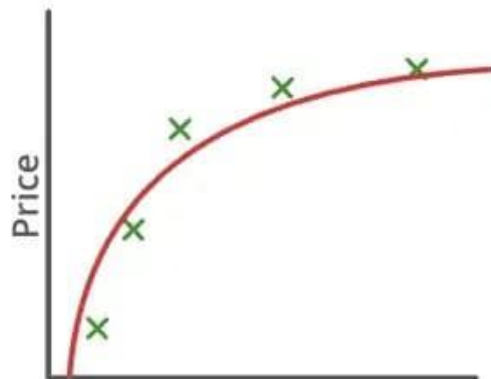
Generalização de modelo é a capacidade do modelo identificar os padrões da base de dados e aplicá-los a novos dados, expandindo seu conhecimento adquirido dos dados de treino para situações novas e desconhecidas.



Size

$$\theta_0 + \theta_1 x$$

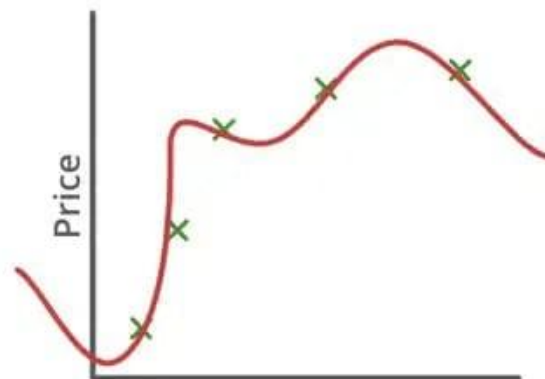
High Bias
(Underfitting)



Size

$$\theta_0 + \theta_1 x + \theta_2 x^2$$

Low Bias, Low Variance
(Goodfitting)



Size

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

High Variance
(Overfitting)

