

# MACHINE LEARNING **SUPERVISIONADO**

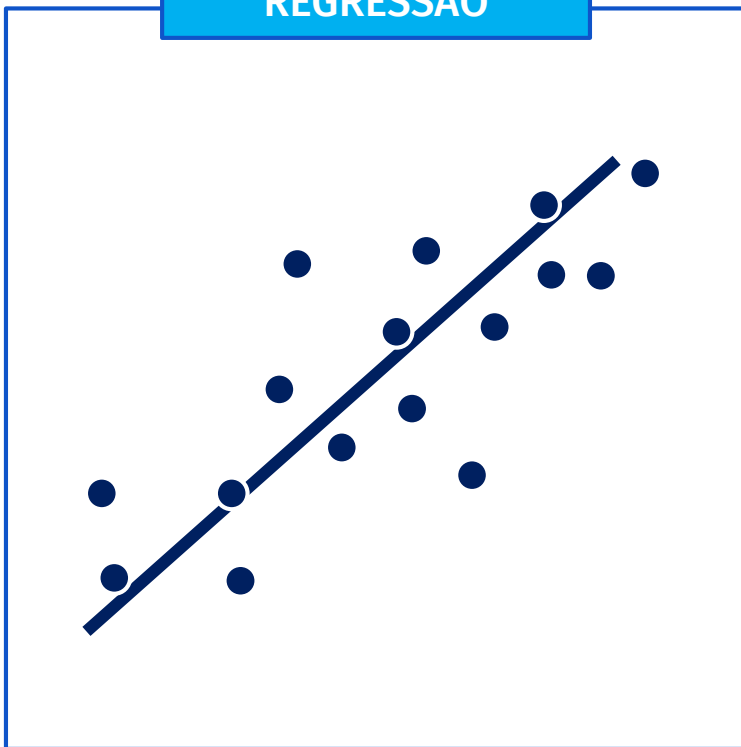
**UM MODELO PARA  
AJUSTAR**

**UMA MÉTRICA PARA  
OTIMIZAR**

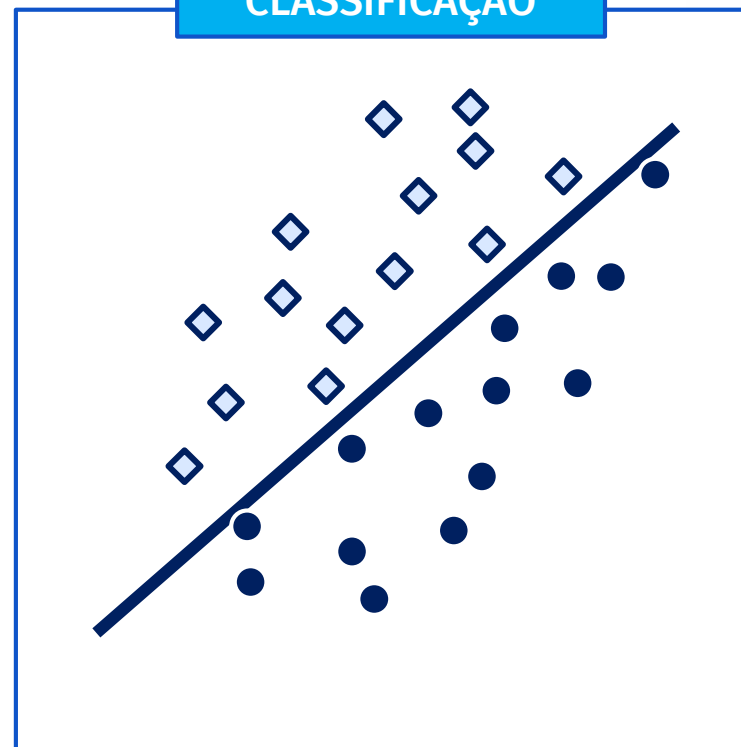
**UM MÉTODO PARA  
VALIDAR**

# Diferenças entre **CLASSIFICAÇÃO E REGRESSÃO**

REGRESSÃO



CLASSIFICAÇÃO



# REGRESSÃO LOGÍSTICA

A regressão logística atua com a seguinte relação funcional entre  $p$  e as variáveis explicativas:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta'_x$$

No geral, a regressão logística **mapeia um ponto  $X$**  no espaço de entrada  $p$ -dimensional para **um valor no intervalo de 0 a 1**

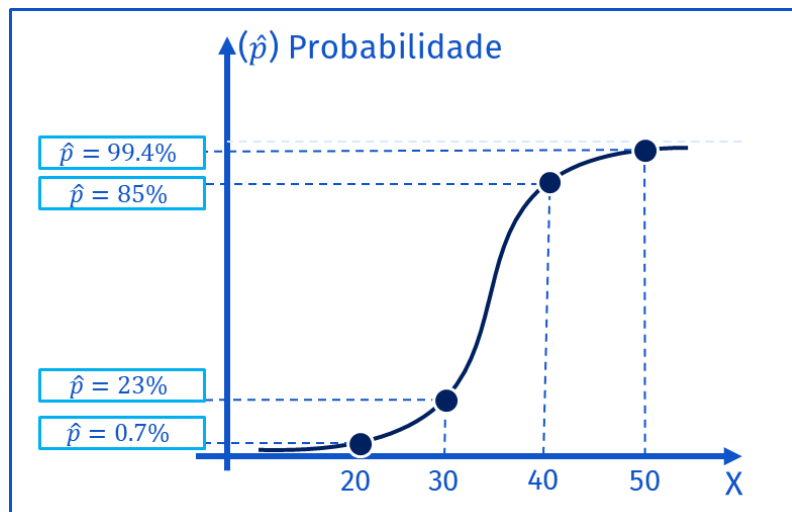
$$P(Y = 0|X = x) = \frac{\exp(-\beta_0 - \beta'_x)}{1 + \exp(-\beta_0 - \beta'_x)}$$

$$P(Y = 1|X = x) = \frac{1}{1 + \exp(-\beta_0 - \beta'_x)}$$

# ESTIMAÇÃO VIA MÁXIMA VEROSSIMILHANÇA

Os EMV de  $\beta_0$  e  $\beta$  são obtidos, em geral, via maximização direta da função de verossimilhança.

Métodos numéricos são frequentemente utilizados para encontrar as estimativas.



$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta x$$

# VANTAGENS DA UTILIZAÇÃO



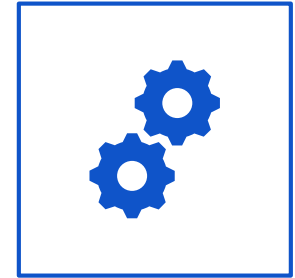
Rápido para  
estimar/treinar e  
classificar novas  
observações



Boa precisão para  
muitos conjuntos de  
dados simples



Resistente ao  
Overfitting



Pode-se interpretar  
os coeficientes do  
modelo como  
indicadores de  
importância

# **MÉTRICAS PARA CLASSIFICAÇÃO BINÁRIA**

# MATRIZ DE CONFUSÃO

	Preditos	Observados
0	1	1
1	0	0
2	0	0
3	1	0
4	0	0
...	...	...
1333	0	0
1334	0	0
1335	0	0
1336	0	0
1337	1	1

		RESPOSTA PREDITA	
		0	1
RESPOSTA OBSERVADA	0	VERDADEIRO NEGATIVO (VN)	FALSO POSITIVO (FP)
	1	FALSO NEGATIVO (FN)	VERDADEIRO POSITIVO (VP)



# ACURÁCIA

$$\text{Acurácia} = \frac{VN + VP}{VN + VP + FN + FP}$$

		RESPOSTA PREDITA	
		0	1
OBSERVADA	0	VN	FP
	1	FN	VP

Representa a **taxa de acerto** de um modelo.

# SENSIBILIDADE

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

		RESPOSTA PREDITA	
		0	1
OBSERVADA	0	VN	FP
	1	FN	VP

Também chamada de recall. Representa a **taxa de verdadeiro positivos**.

# PRECISÃO

$$\text{Valor Preditivo Positivo} = \frac{VP}{VP + FP}$$

		RESPOSTA PREDITA	
		0	1
OBSERVADA	0	VN	FP
	1	FN	VP

Também chamada de **Valor Preditivo Positivo**.



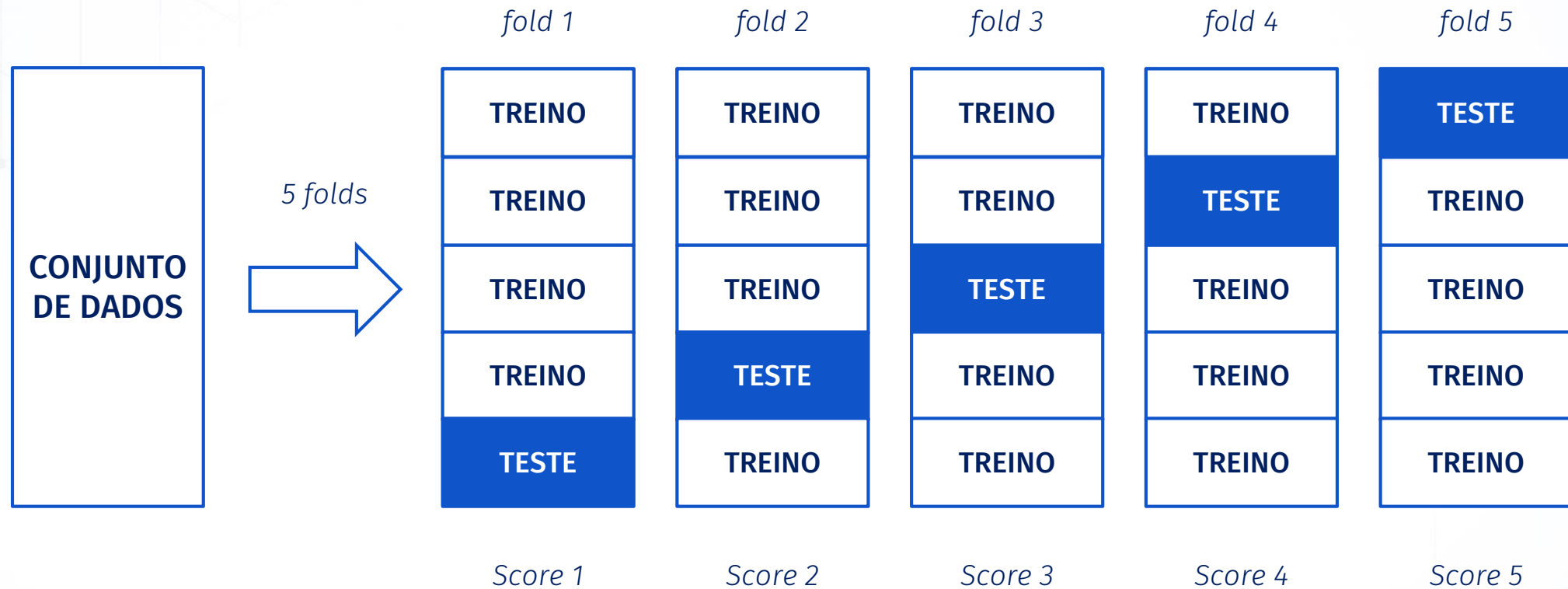
# **UPGRADE NA TÉCNICA DE VALIDAÇÃO**

# **VALIDAÇÃO HOLDOUT**



# **VALIDAÇÃO K-FOLD**

# VALIDAÇÃO CRUZADA K-FOLD



*Score Final – Média dos Scores Individuais*

# OBJETIVO DE UMA VALIDAÇÃO

## SELEÇÃO DE MODELOS



Selecionar os modelos ou abordagens de modelagens com melhor performance dentre uma lista de opções disponíveis.

## AVALIAÇÃO DE PERFORMANCE



Escolhido um modelo, deseja-se saber seu erro de generalização, isto é, o desempenho esperado do modelo em produção.