

INTELIGÊNCIA ARTIFICIAL

**APRENDIZADO DE MÁQUINA –
AVALIAÇÃO DE DESEMPENHO**

MATERIAL DE ESTUDO

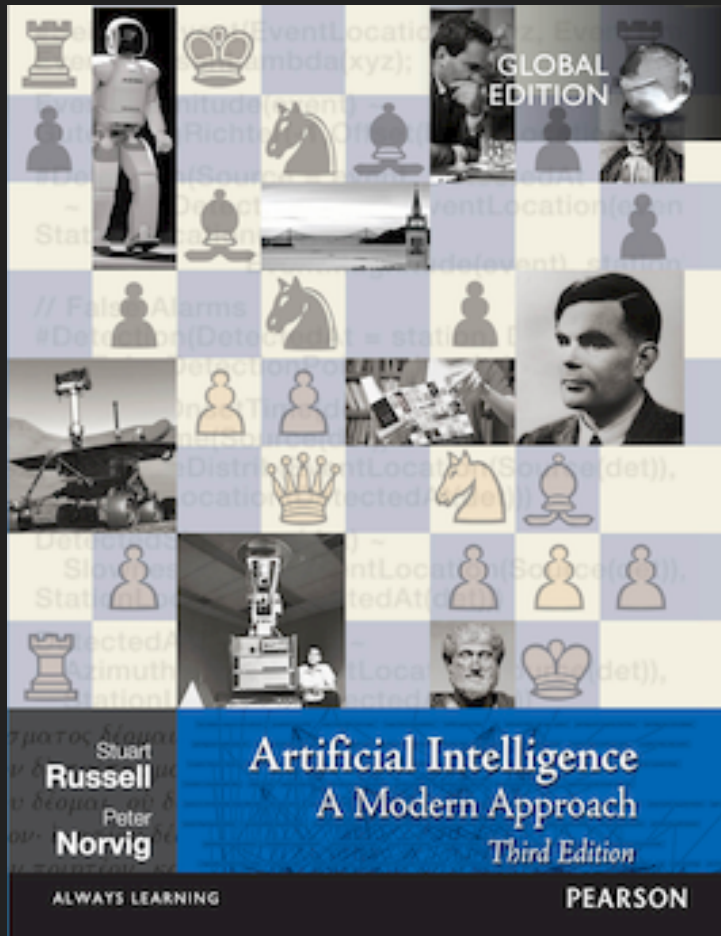
Artificial Intelligence: A modern approach

Cap. 18

Texto Complementar:

[What is Confusion Matrix and Advanced Classification Metrics?](#)

MATERIAL DE ESTUDO



Artificial Intelligence: A modern approach

Cap. 18

Texto Complementar:

[What is Confusion Matrix and Advanced Classification Metrics?](#)

SUMÁRIO

- ▶ Matriz de confusão (confusion matrix)
- ▶ Métricas de desempenho
- ▶ Avaliando e escolhendo as melhores hipóteses (modelos).

MATRIZ DE CONFUSÃO

MATRIZ DE CONFUSÃO

- ▶ Ferramenta simples para determinar o desempenho de uma hipótese.
- ▶ Usada em problemas de classificação (binária ou n-ária)

MATRIZ DE CONFUSÃO

Seja um problema de classificação binária temos a seguinte matriz de confusão:

		Predicted Class	
		Yes	No
True Class	Yes	TP	FN Type II Error
	No	FP Type I Error	TN

		Predicted Class	
		Yes	No
True Class	Yes	TP	FN Type II Error
	No	FP Type I Error	TN

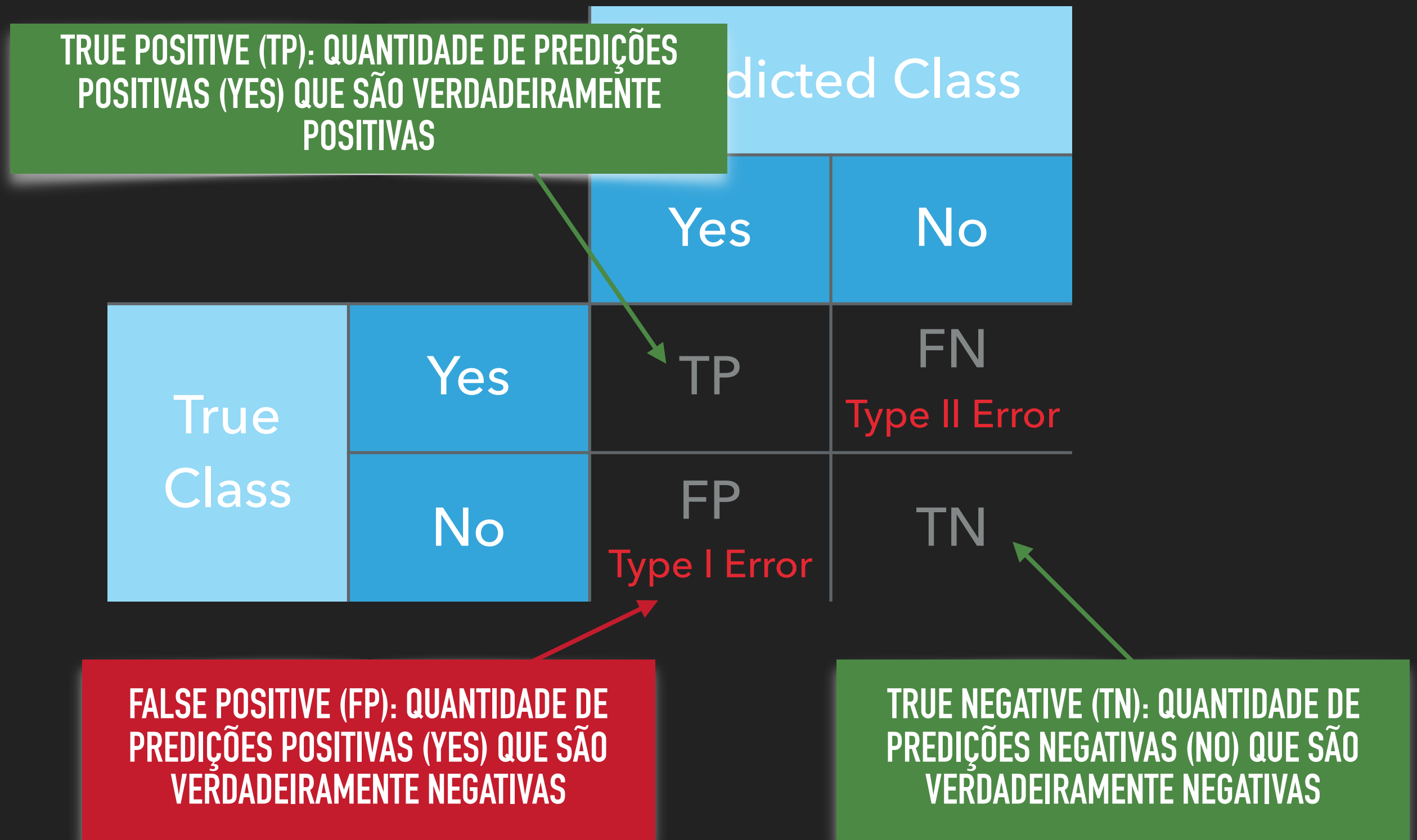
TRUE POSITIVE (TP): QUANTIDADE DE PREDIÇÕES POSITIVAS (YES) QUE SÃO VERDADEIRAMENTE POSITIVAS

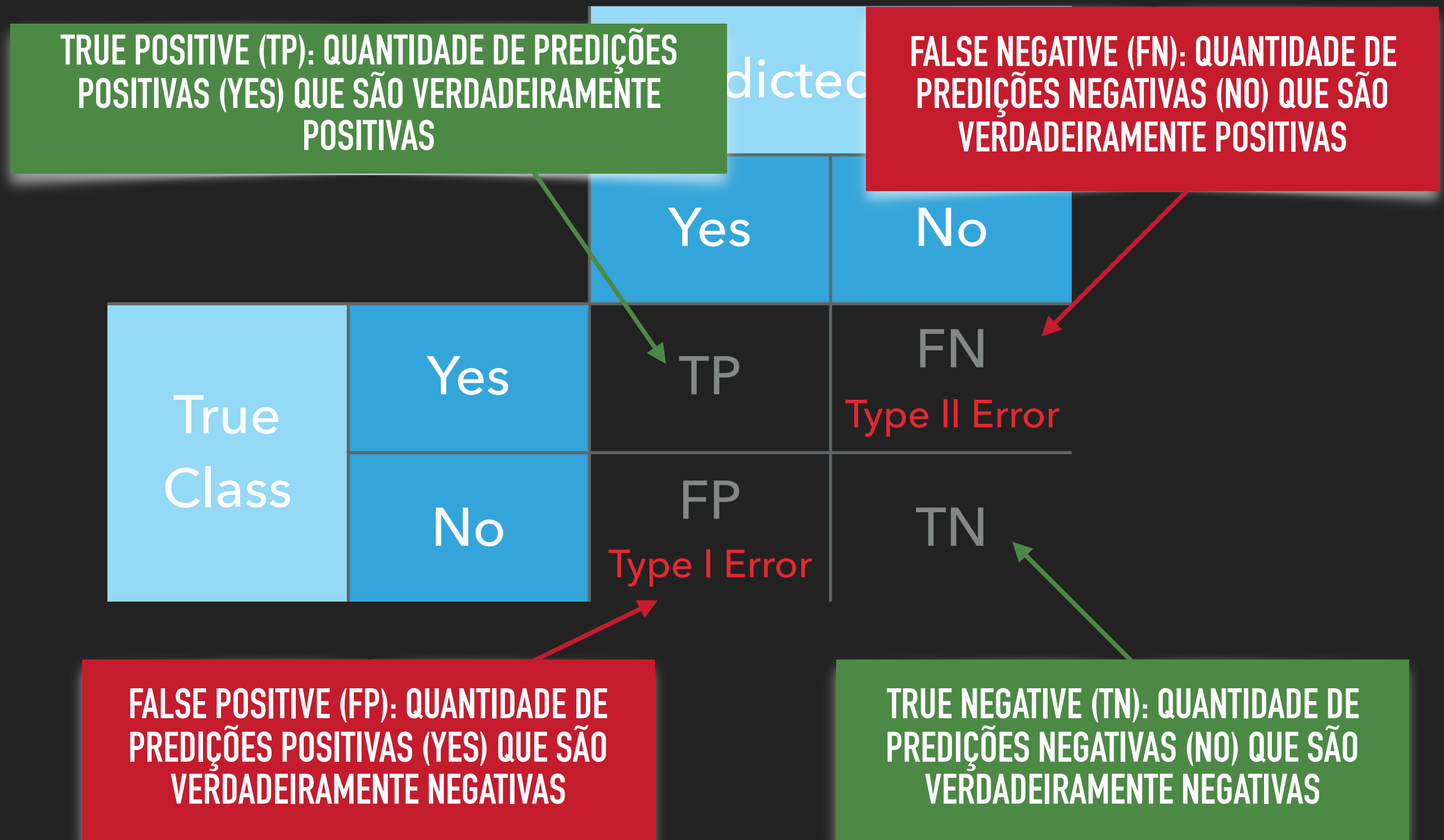
		Predicted Class	
		Yes	No
True Class	Yes	TP	FN Type II Error
	No	FP Type I Error	TN

		Predicted Class	
		Yes	No
True Class	Yes	TP	FN Type II Error
	No	FP Type I Error	TN

TRUE POSITIVE (TP): QUANTIDADE DE PREDIÇÕES POSITIVAS (YES) QUE SÃO VERDADEIRAMENTE POSITIVAS

TRUE NEGATIVE (TN): QUANTIDADE DE PREDIÇÕES NEGATIVAS (NO) QUE SÃO VERDADEIRAMENTE NEGATIVAS





MÉTRICAS

		Predicted Class	
		Yes	No
True Class	Yes	TP	FN Type II Error
	No	FP Type I Error	TN

MÉTRICAS

		Predicted Class		
		Yes	No	
True Class	Yes	TP	FN Type II Error	Sensitivity (Recall) $\frac{TP}{TP + FN}$
	No	FP Type I Error	TN	

MÉTRICAS

		Predicted Class		
		Yes	No	
True Class	Yes	TP	FN Type II Error	Sensitivity (Recall) $\frac{TP}{TP + FN}$
	No	FP Type I Error	TN	Specificity $\frac{TN}{TN + FP}$

MÉTRICAS

		Predicted Class		
		Yes	No	
True Class	Yes	TP	FN Type II Error	Sensitivity (Recall) $\frac{TP}{TP + FN}$
	No	FP Type I Error	TN	Specificity $\frac{TN}{TN + FP}$
		Precision $\frac{TP}{TP + FP}$		

MÉTRICAS

		Predicted Class		
		Yes	No	
True Class	Yes	TP	FN Type II Error	Sensitivity (Recall) $\frac{TP}{TP + FN}$
	No	FP Type I Error	TN	Specificity $\frac{TN}{TN + FP}$
		Precision $\frac{TP}{TP + FP}$		Acurácia $\frac{TP + TN}{TP + FP + TN + FN}$

MÉTRICAS

Predicted Class	
Yes	No

Balancamento entre Precision e Recall

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Precision

$$\frac{TP}{TP + FP}$$

Acurácia

$$\frac{TP + TN}{TP + FP + TN + FN}$$

AVALIANDO E ESCOLHENDO AS MELHORES HIPÓTESES

AVALIANDO E ESCOLHENDO AS MELHORES HIPÓTESES

- ▶ Queremos aprender uma hipótese (modelo) que melhor se encaixa nos dados futuros.

Queremos aprender uma hipótese (modelo) que melhor se encaixa nos dados futuros.

Queremos aprender uma hipótese (modelo) que melhor se encaixa nos dados futuros.

O QUE SÃO DADOS FUTUROS?

Queremos aprender uma hipótese (modelo) que melhor se encaixa nos dados futuros.

O QUE SÃO DADOS FUTUROS?

O QUE É MELHOR SE ENCAIXA?

DADOS FUTUROS

- ▶ Valores do domínio ainda não observados.
- ▶ Stationary Assumption: a distribuição de probabilidade dos exemplos é a mesma independentemente da ocorrência dos mesmos.

$$P(E_j | E_{j-1}, E_{j-2}, \dots, E_{j-n}) = P(E_j)$$

$$e_j = (x_j, y_j)$$

MELHOR SE ENCAIXA (BEST FIT)

- ▶ Taxa de erro (error rate): proporção erros cometidos pela hipótese em questão.

MELHOR SE ENCAIXA (BEST FIT)

- ▶ Taxa de erro (error rate): proporção erros cometidos pela hipótese em questão.

$$h(x) \neq y \forall (x, y)$$

MELHOR SE ENCAIXA (BEST FIT)

- ▶ Taxa de erro (error rate): proporção erros cometidos pela hipótese em questão.

$$h(x) \neq y \forall (x, y)$$

NÃO É PORQUE A TAXA DE ERRO É BAIXA A PARTIR DO CONJUNTO DE TREINAMENTO QUE A HIPÓTESE EM QUESTÃO GENERALIZARÁ BEM

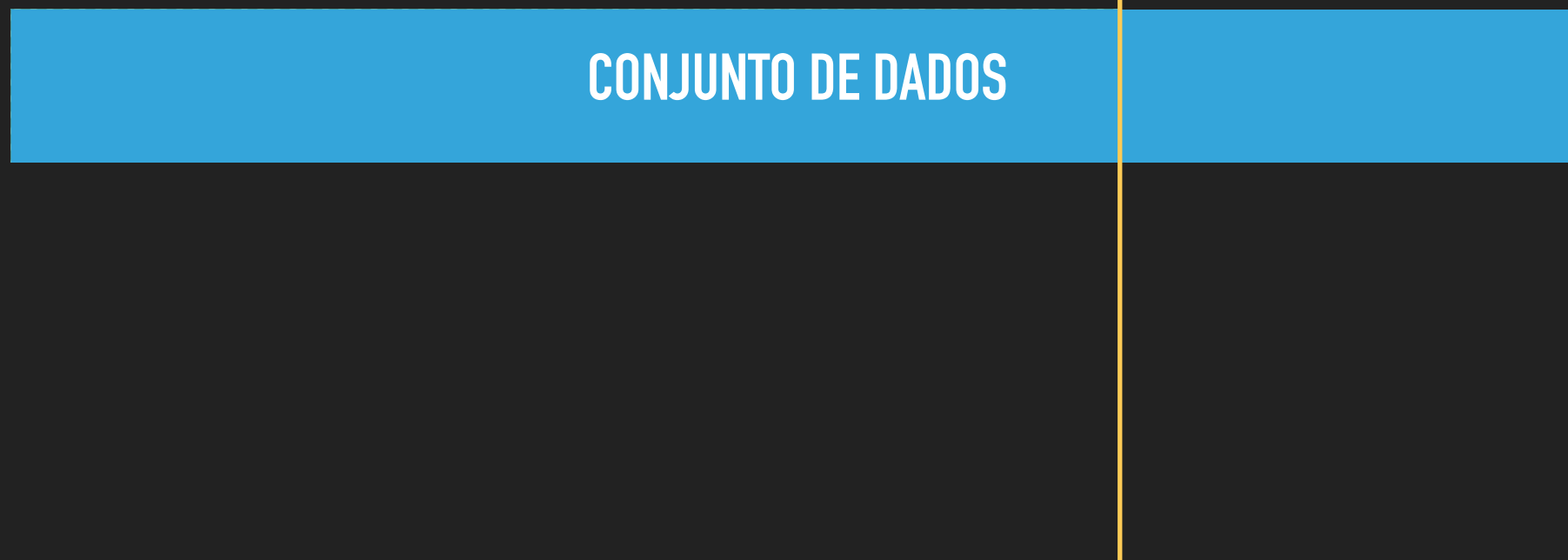
HOLDOUT CROSS-VALIDATION (TRAIN-TEST SPLIT)



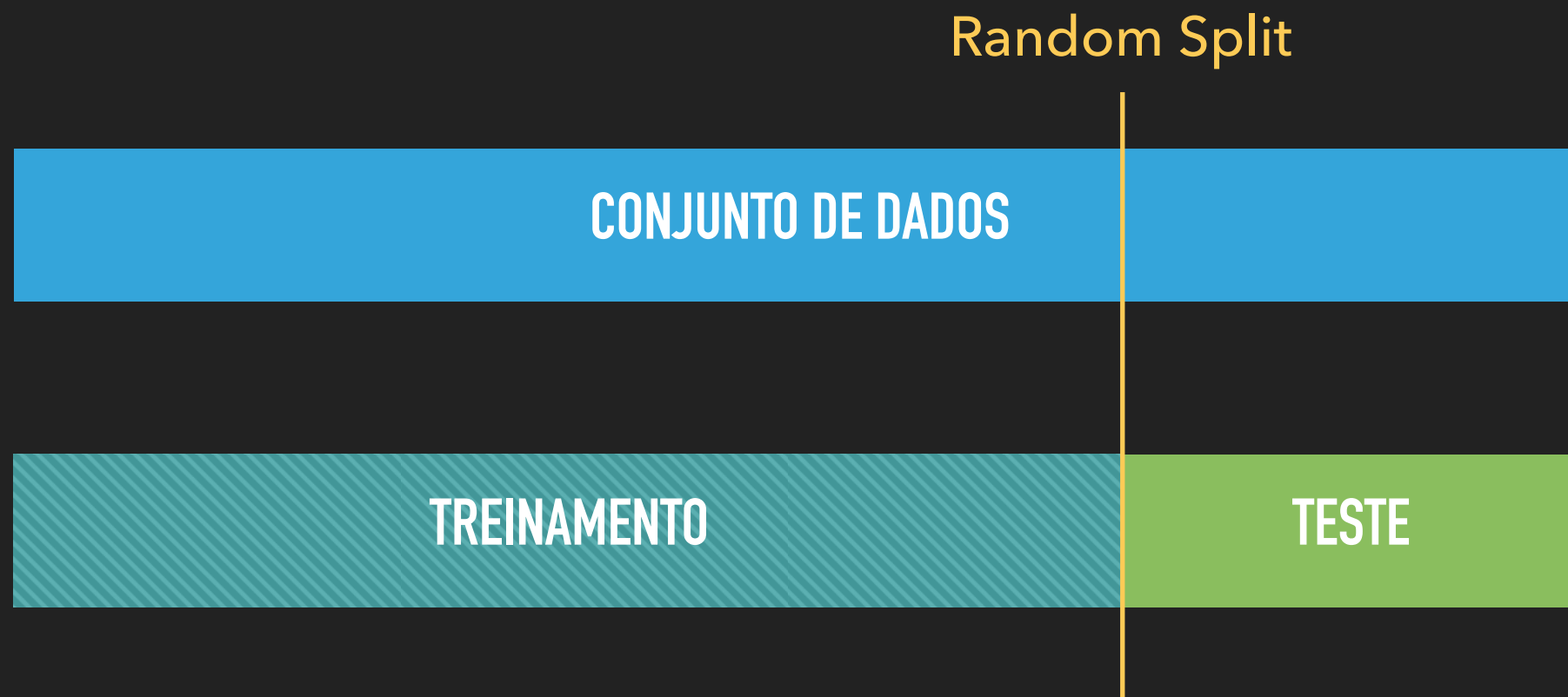
CONJUNTO DE DADOS

HOLDOUT CROSS-VALIDATION (TRAIN-TEST SPLIT)

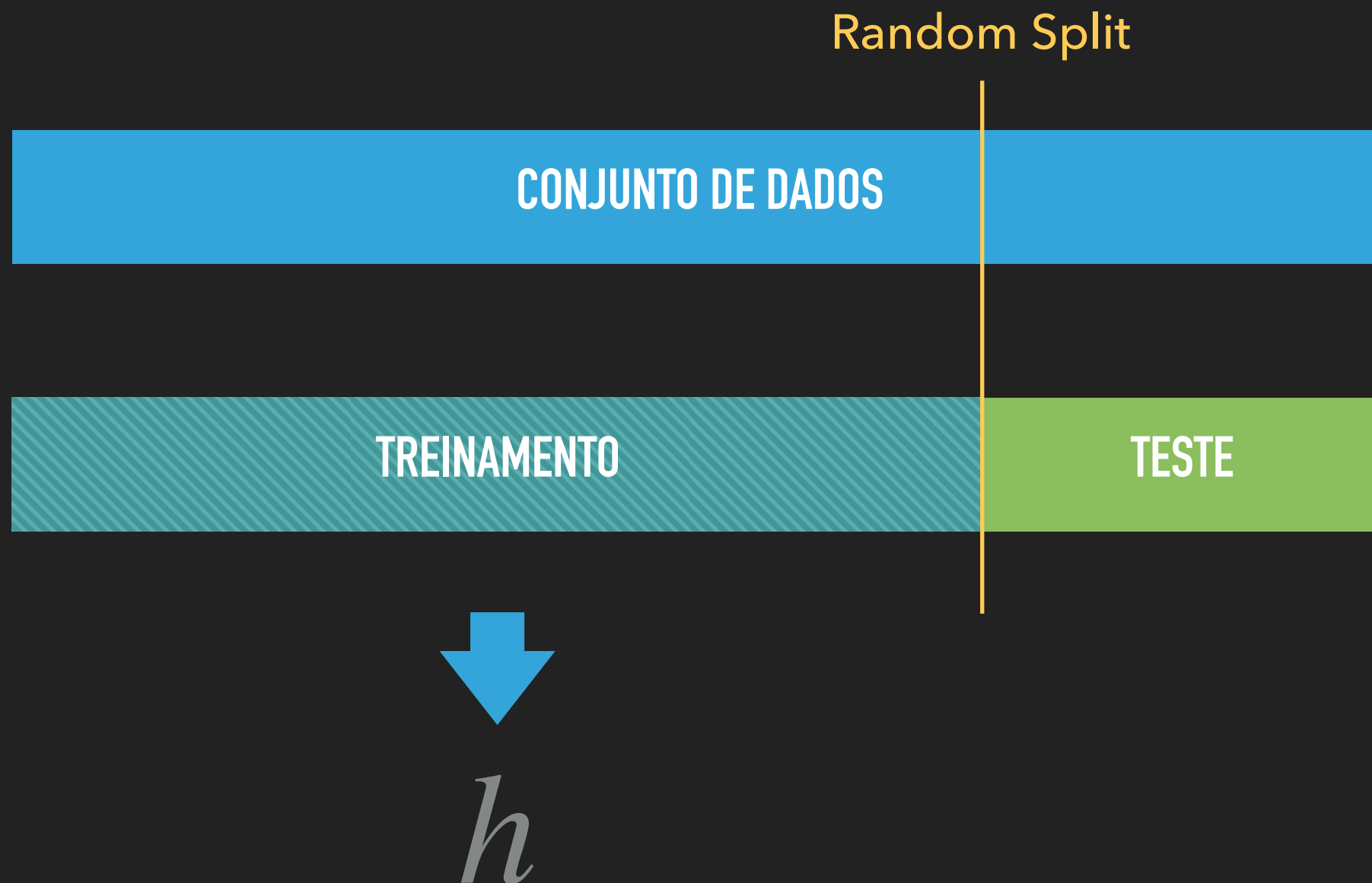
Random Split



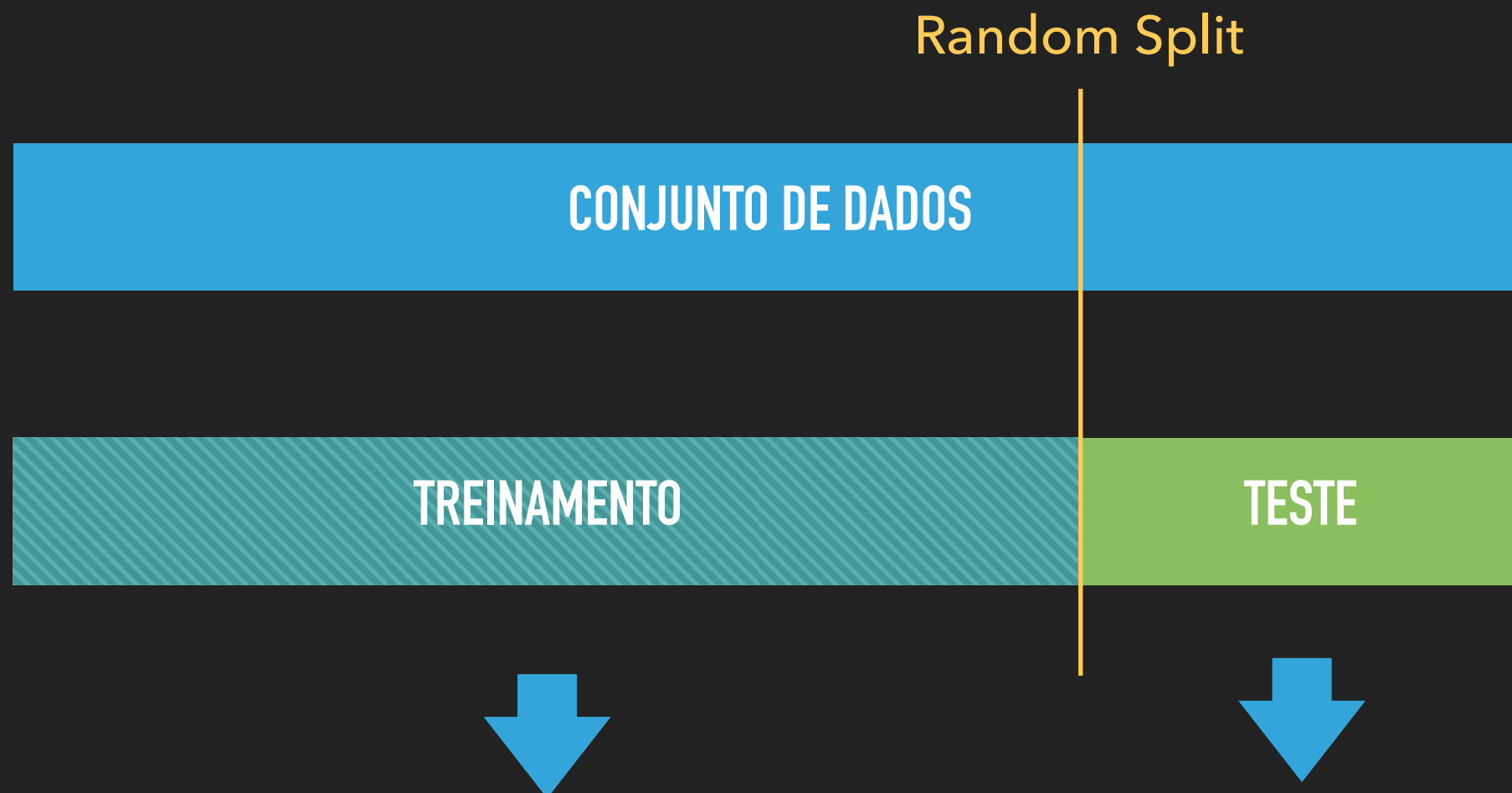
HOLDOUT CROSS-VALIDATION (TRAIN-TEST SPLIT)



HOLDOUT CROSS-VALIDATION (TRAIN-TEST SPLIT)



HOLDOUT CROSS-VALIDATION (TRAIN-TEST SPLIT)



$$h(x) \neq y \forall (x, y)$$

ERROR RATE

HOLDOUT CROSS-VALIDATION (TRAIN-TEST SPLIT)

- ▶ Problemas:

- ▶ Os dados não estarão, na sua totalidade, disponíveis para treinamento.
- ▶ Escolha do particionamento (split):
 - ▶ **alto** (i.e. 50%) pode deixar poucos dados para treinamento gerando um hipótese pobre.
 - ▶ **baixo** (i.e. 10%) pode deixar poucos dados para teste gerando um error rate não confiável.

K-FOLD CROSS-VALIDATION

TREINAMENTO

TESTE

CONJUNTO DE DADOS

K-FOLD CROSS-VALIDATION

TREINAMENTO

TESTE

$K = 5$

CONJUNTO DE DADOS



K-FOLD CROSS-VALIDATION

TREINAMENTO

TESTE

$K = 5$



K-FOLD CROSS-VALIDATION

TREINAMENTO

TESTE

$K = 5$



K-FOLD CROSS-VALIDATION

TREINAMENTO

TESTE

$K = 5$

	CONJUNTO DE DADOS				
1	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5
2	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5
3	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5

ERROR RATE 1

ERROR RATE 2

ERROR RATE 3

K-FOLD CROSS-VALIDATION

TREINAMENTO

TESTE

$K = 5$

	CONJUNTO DE DADOS					
1	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5	ERROR RATE 1
2	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5	ERROR RATE 2
3	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5	ERROR RATE 3
4	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5	ERROR RATE 4

K-FOLD CROSS-VALIDATION

TREINAMENTO

TESTE

K = 5

	CONJUNTO DE DADOS				
1	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5
2	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5
3	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5
4	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5
5	SPLIT 1	SPLIT 2	SPLIT 3	SPLIT 4	SPLIT 5
					ERROR RATE MÉDIO (+/-)

K-FOLD CROSS-VALIDATION

- ▶ Atenção:
 - ▶ A cada validação cruzada é necessário gerar uma nova hipótese com os mesmos parâmetros de aprendizado (evitar **peaking**).
 - ▶ Possui uma versão extrema chamada Leave-One-Out Cross-Validation: onde $k = n$, sendo n o tamanho do conjunto de dados disponível.

HANDS ON

