

MÉTODOS E MÉTRICAS DE AVALIAÇÃO

Prof. Julio Cesar dos Reis

jreis@ic.unicamp.br

www.ic.unicamp.br/~jreis

CT-0611

Objetivos da aula

2

- Estudar Métodos e Métricas de Avaliação
- Medir a qualidade de classificadores
- Aprender o uso correto das métricas para garantir modelos mais confiáveis e adequados ao contexto do problema.

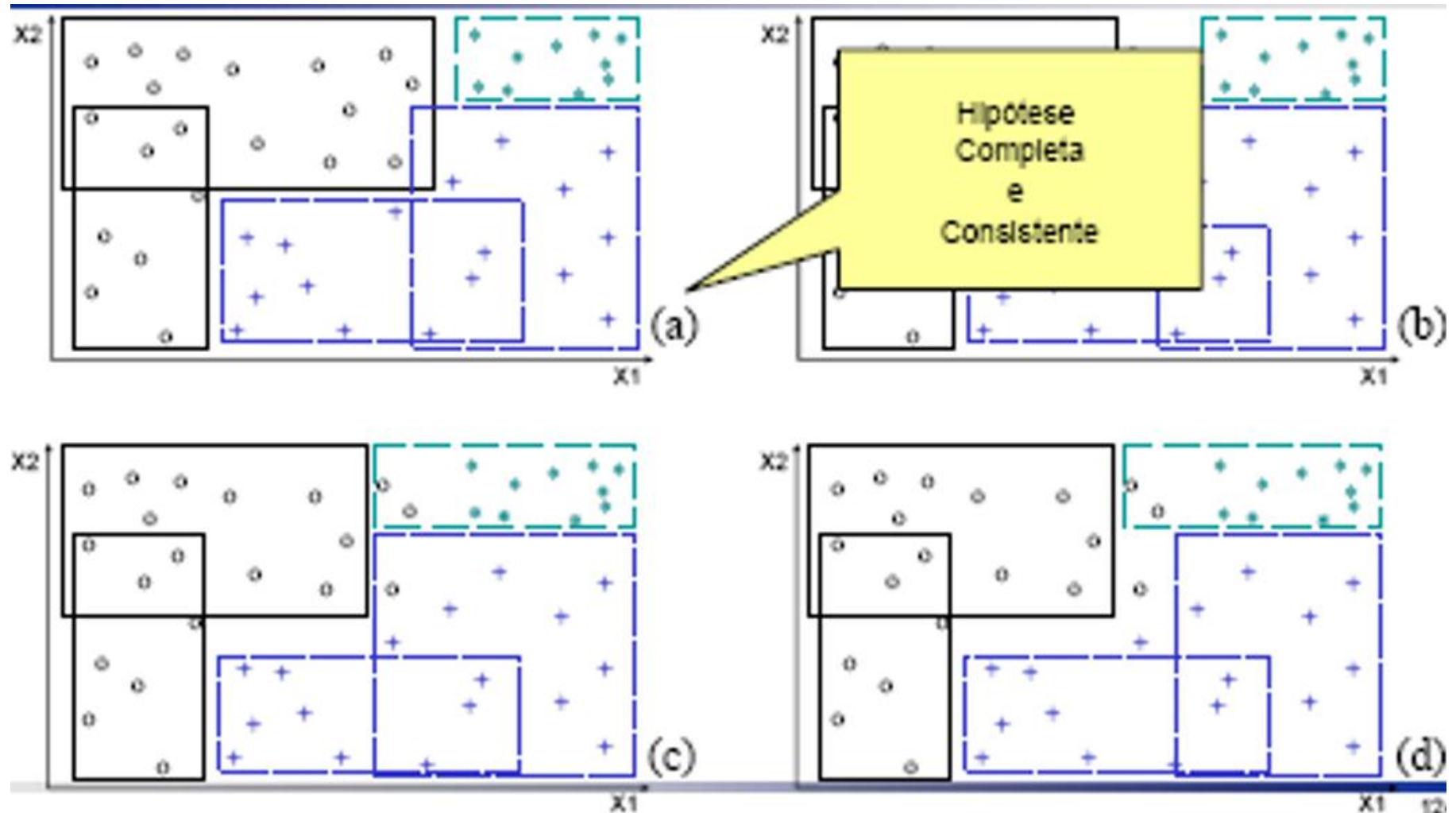
Consistência e completude

3

- Depois de induzida, uma hipótese pode ser avaliada em relação à:
 - Consistência
 - ▣ se classifica corretamente os exemplos
 - Completude
 - ▣ se classifica todos os exemplos

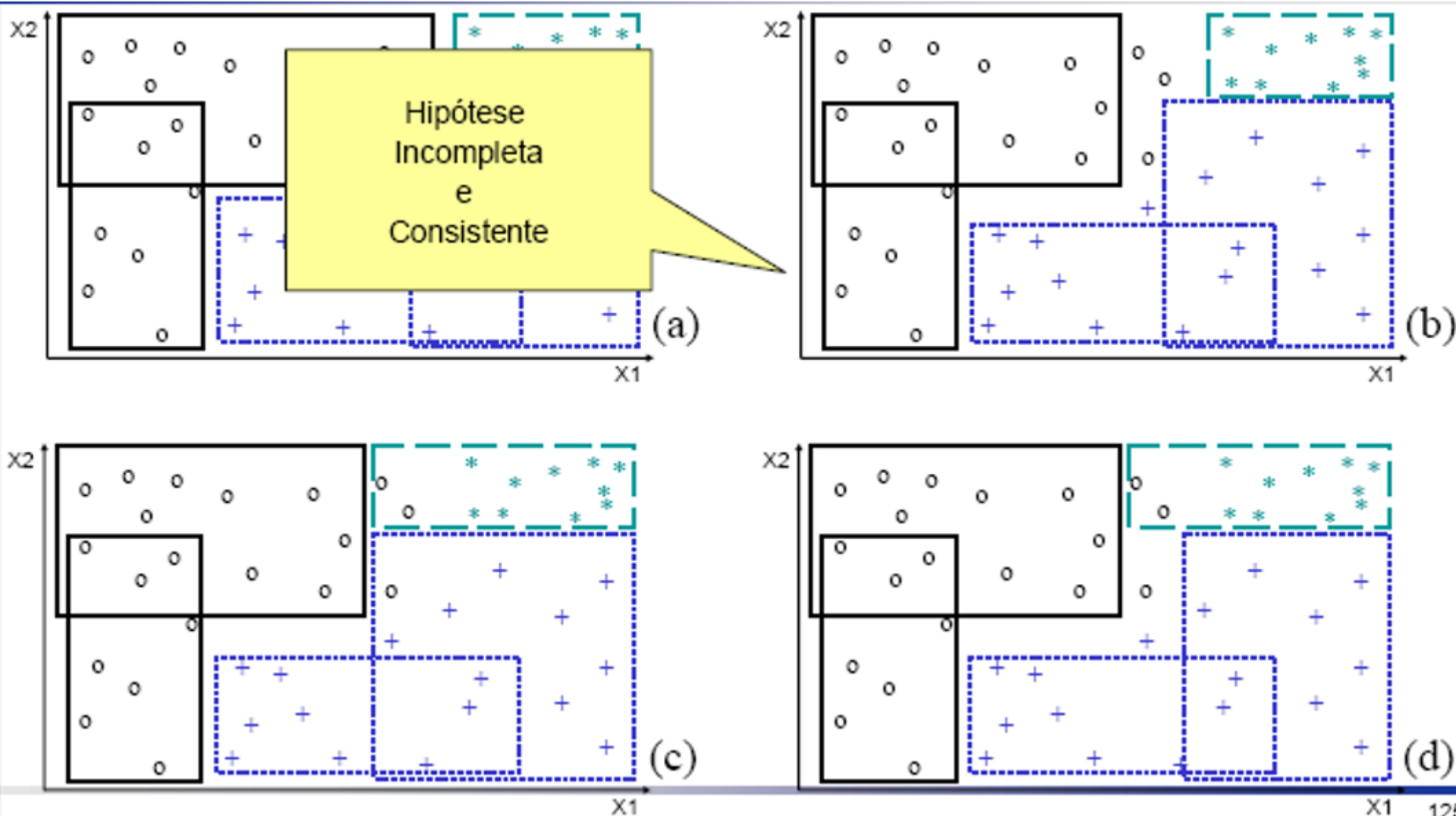
Relação entre completude e consistência

4



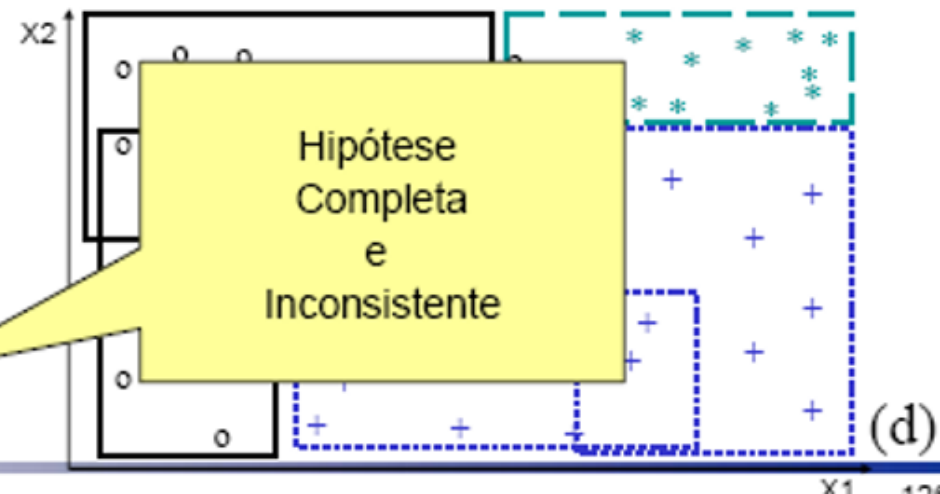
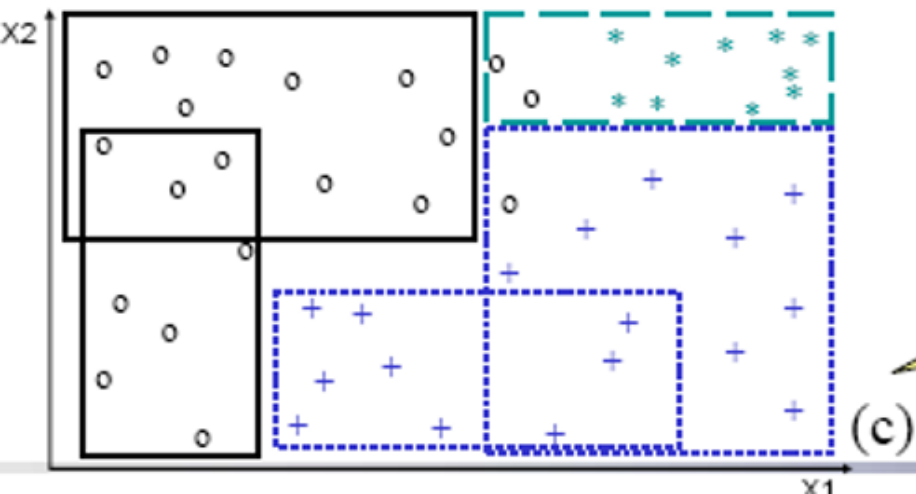
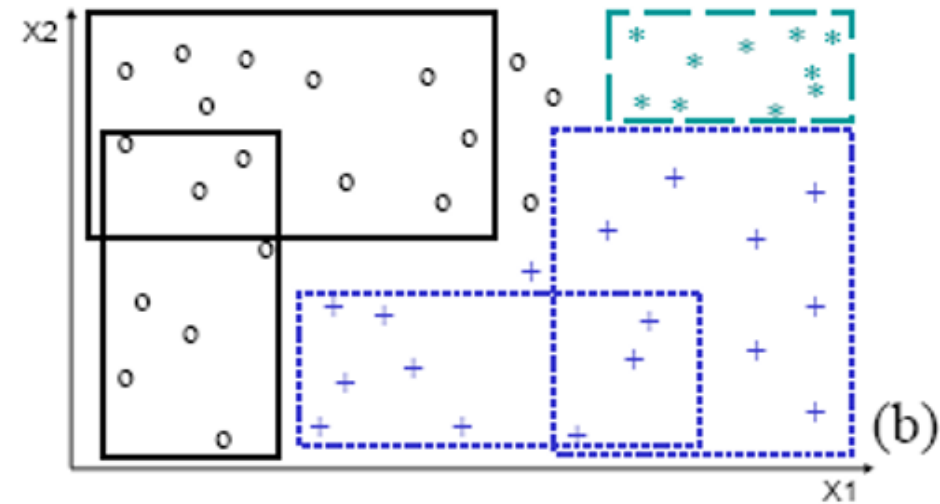
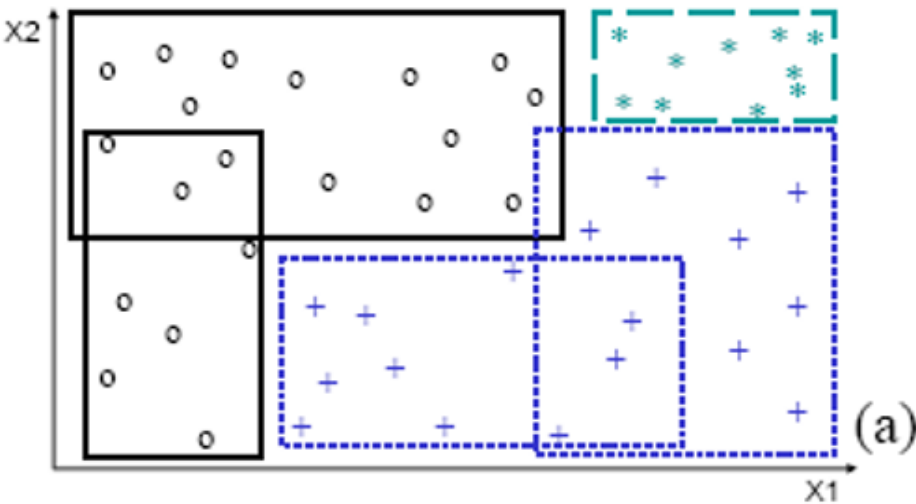
Relação entre completude e consistência

5



Relação entre completude e consistência

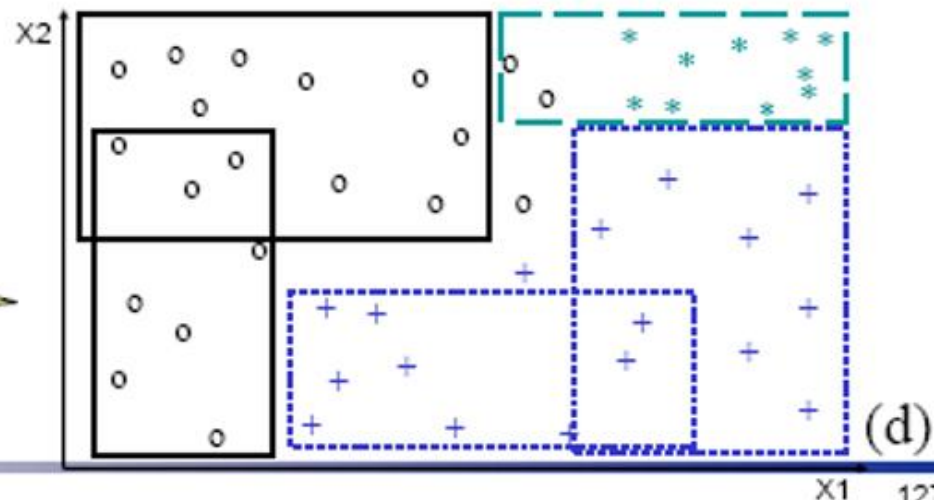
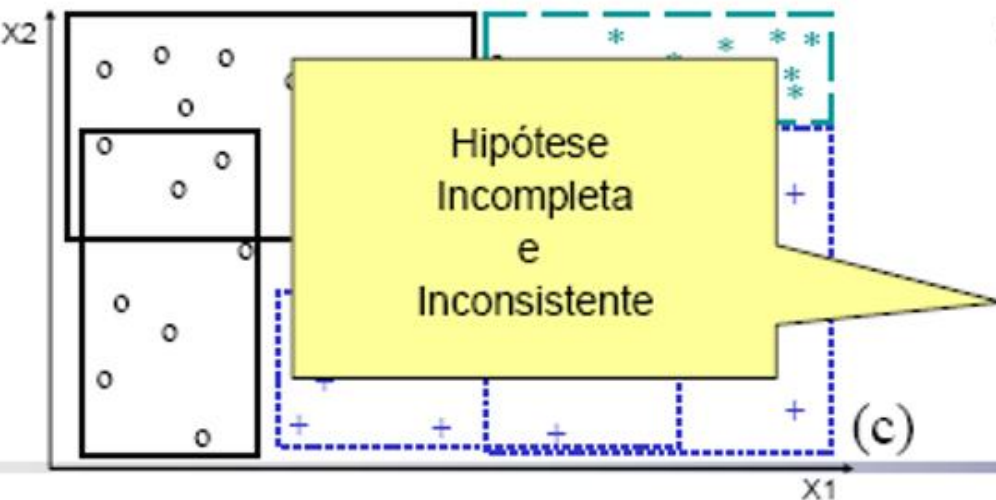
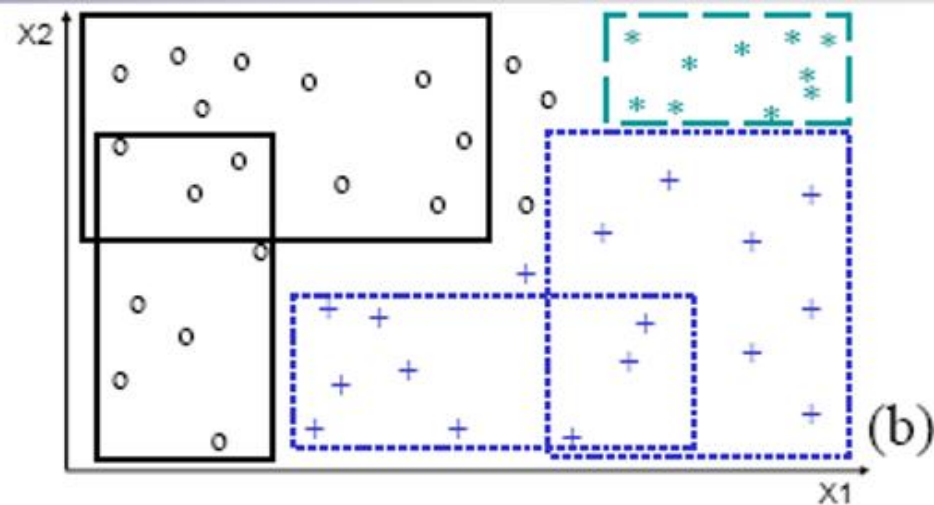
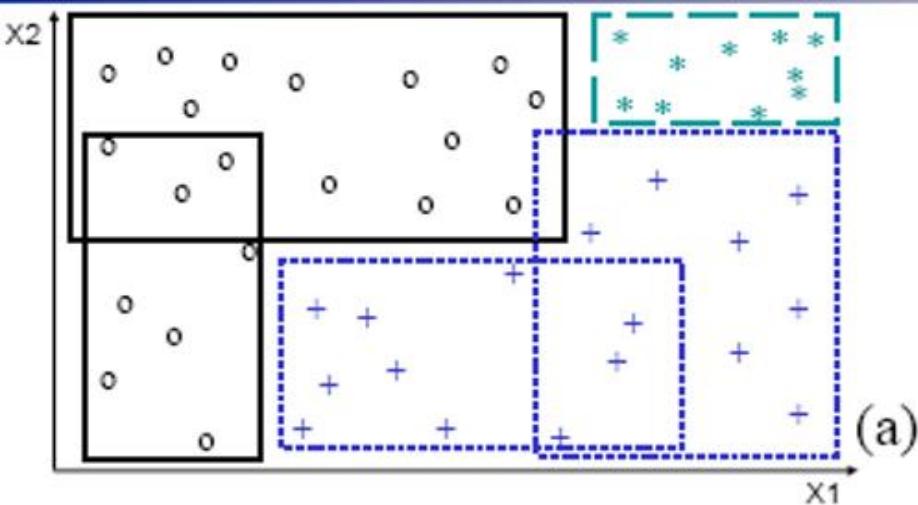
6



Hipótese
Completa
e
Inconsistente

Relação entre completude e consistência

7



Medindo a qualidade da predição

8

- Precisão, compreensível e interessante
- Acurácia = classificados corretamente / total de exemplos
- Erro = $1 - \text{Acurácia}$

Matriz de confusão

9

- Uma matriz quadrada que indica as classificações corretas e erradas
 - ▣ A classe que está sendo analisada aparece na linha
 - ▣ As classificações encontradas aparecem nas colunas
 - ▣ A **diagonal** da matriz corresponde às **classificações corretas**

```
=== Confusion Matrix ===
```

```
a b  <-- classified as  
9 0 | a = Sim  
0 5 | b = Não
```

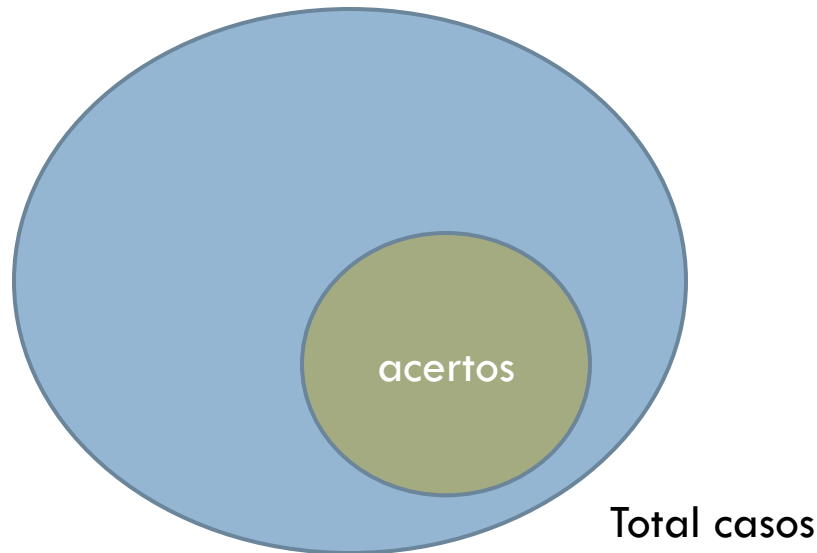
```
=== Confusion Matrix ===
```

```
a b  <-- classified as  
0 8 | a = Sim  
0 3 | b = Não
```

Qualidade da predição

10

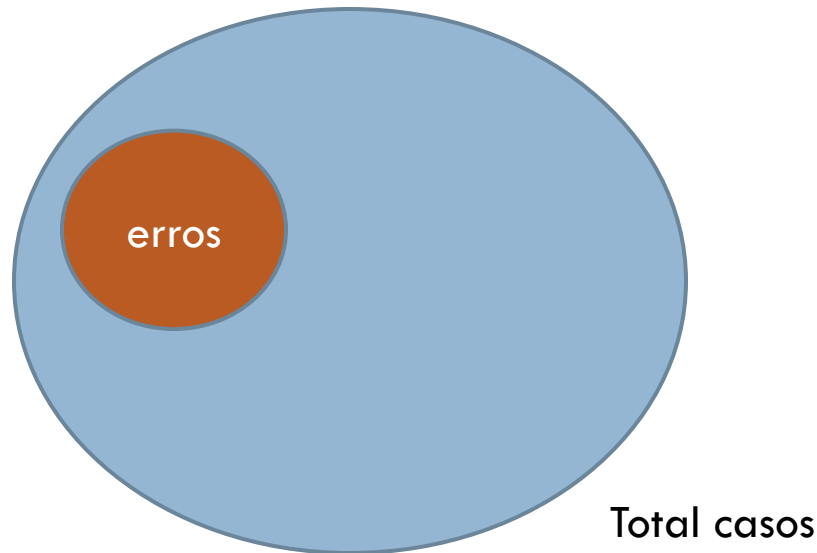
- ❑ **Taxa de Acertos:** número de casos classificados corretamente dividido pelo número total de casos
 - $\text{taxa_de_acertos} = N_{\text{acertos}} / N_{\text{total}}$



Qualidade da predição

11

- **Taxa de Erros:** número de casos classificados incorretamente dividido pelo número total de casos
 - $\text{taxa_de_erros} = N_{\text{erros}} / N_{\text{total}}$



Qualidade da predição

12

- $\text{proporção_de_acertos} = N_{\text{acertos}} / N_{\text{total}}$
 $\text{proporção_de_acertos} = 5 / 7 = 71,4\%$
- $\text{proporção_de_erros} = N_{\text{erros}} / N_{\text{total}}$
 $\text{proporção_de_erros} = 2 / 7 = 28,6\%$

caso	montante	idade	salário	conta	empréstimo	empréstimo (predito)
15	médio	sênior	alto	sim	sim	<i>sim</i>
16	médio	sênior	alto	não	não	<i>sim</i>
17	baixo	jovem	alto	sim	sim	<i>sim</i>
18	baixo	sênior	alto	não	não	<i>sim</i>
19	alto	média	alto	não	não	<i>não</i>
20	alto	jovem	alto	sim	sim	<i>sim</i>
21	médio	jovem	alto	sim	sim	<i>sim</i>

Matriz de confusão

13

- A matriz de confusão de uma hipótese h oferece uma medida efetiva do modelo de classificação
- Mostrar o número de classificações corretas versus as classificações preditas para cada classe, sobre um conjunto de exemplos T

Matriz de confusão

14

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Matriz de confusão

16

- O número de acertos, para cada classe, se localiza na diagonal principal $M(C_i, C_i)$ da matriz
- Os demais elementos $M(C_i, C_j)$, para $i \neq j$, representam erros na classificação
 - ▣ A matriz de confusão de um classificador ideal possui todos esses elementos iguais a zero uma vez que ele não comete erros

Matriz de confusão

17

Modelo adequado

Actual	Positive	Negative
	23	7
Positive	10	60
Predicted		

Modelo inadequado

Actual	Positive	Negative
	23	7
Positive	58	12
Predicted		

Matriz de confusão

18

□ Múltiplas classes

Actual	Elephant	25	3	0	2
	Monkey	3	53	2	3
	Fish	2	1	24	2
	Lion	1	58	2	13
		Elephant	Monkey	Fish	Lion
		Predicted			

Actual	Elephant	TN	FP	TN	
	Monkey	TN	FP	TN	
	Fish	FN	TP	FN	
	Lion	TN	FP	TN	
		Elephant	Monkey	Fish	Lion
		Predicted			

Acurácia

19

		Classe prevista	
		Sim	Não
Classe real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Acurácia: $\frac{VP + VN}{n} \times 100\%$

$$n = VP + VN + FP + FN$$

Erro

20

		Classe prevista	
		Sim	Não
Classe real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

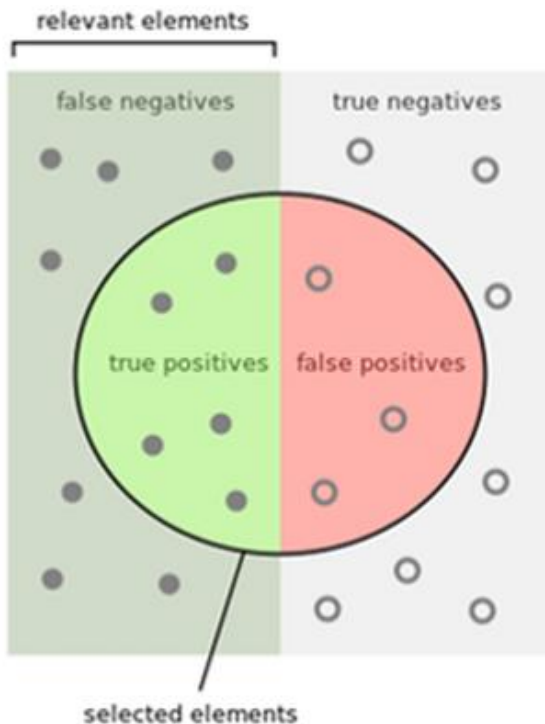
$$\text{Erro: } \frac{FP + FN}{n} \times 100\%$$

$$n = VP + VN + FP + FN$$

Cobertura

21

$$\frac{VP}{VP + FN} \times 100\%$$



How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

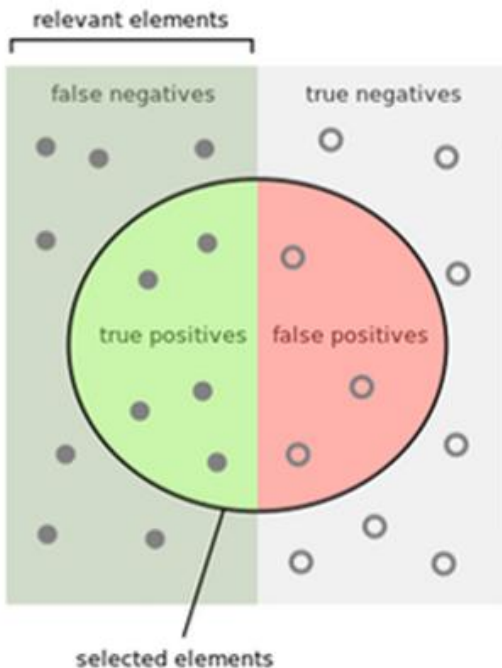
		Classe prevista	
		Sim	Não
Classe real	Sim	VP	FN
	Não	FP	VN

Precisão

22

$$\frac{VP}{VP + FP} \times 100\%$$

		Classe prevista	
		Sim	Não
Classe real	Sim	VP	FN
	Não	FP	VN



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

F-measure

24

- Média harmônica entre cobertura e precisão

$$\frac{2 \cdot \textit{precisão} \cdot \textit{recall}}{\textit{precisão} + \textit{recall}}$$

AUC-ROC

25

- Métrica usada para avaliar o desempenho de **modelos de classificação binária**
 - ▣ Mede a capacidade do modelo de distinguir entre duas classes
 - ▣ ROC: relação entre **taxa de verdadeiros positivos (TPR)** e **taxa de falsos positivos (FPR)** para diferentes limiares de decisão.
 - ▣ **AUC (Área Under the Curve)** →
Representa a área sob essa curva e indica a capacidade do modelo de classificar corretamente as classes.



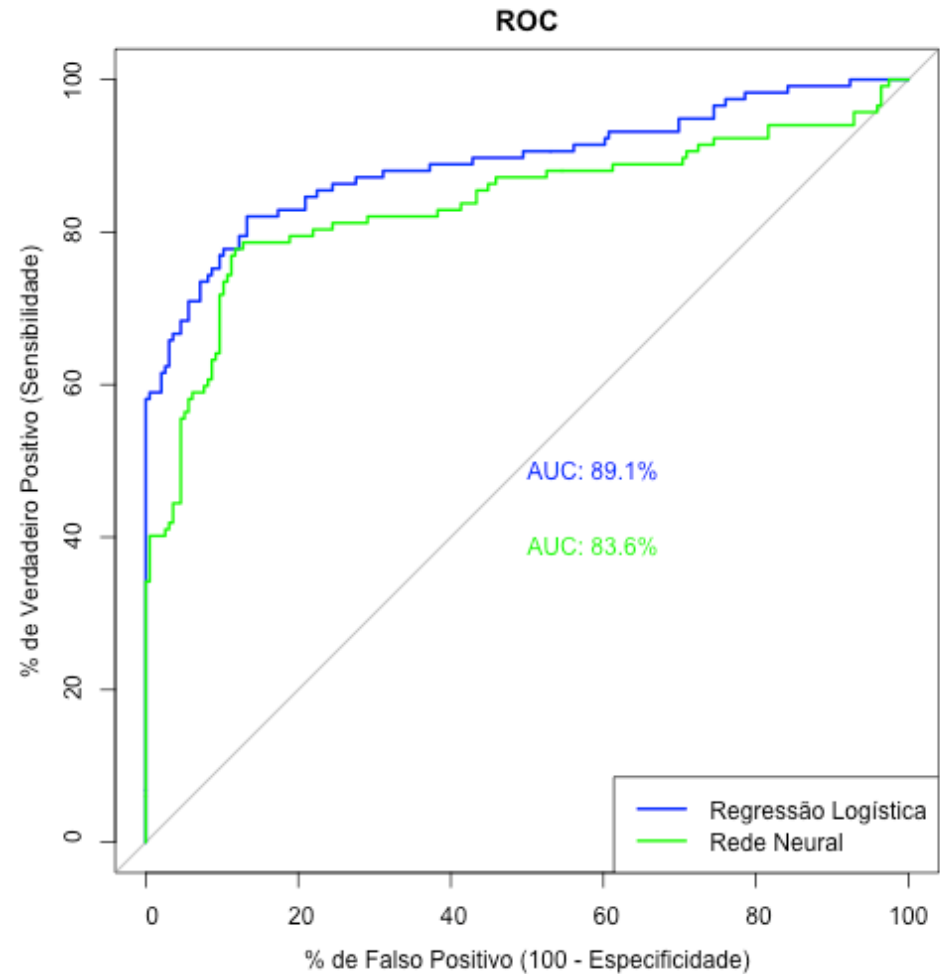
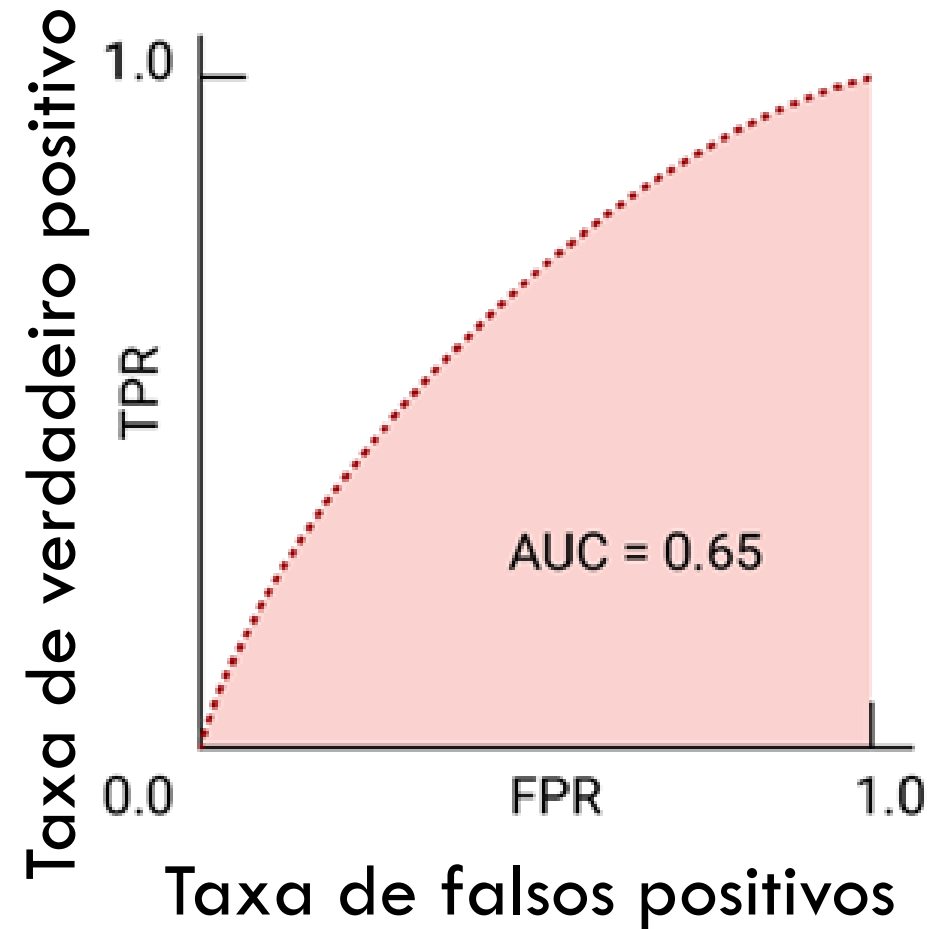
AUC-ROC (Interpretação)

26

- **AUC = 1.0** → **Modelo perfeito** (classifica corretamente todas as instâncias)
- **AUC = 0.5** → **Modelo aleatório** (desempenho equivalente a um classificador que chuta aleatoriamente)
- **AUC < 0.5** → **Modelo ruim** (pior que um chute aleatório, possivelmente invertendo as classes)

AUC-ROC

27



Prevalência de classe

28

- Desbalanceamento de classes em um conjunto de exemplos pode ser um problema

- Suponha um conjunto de exemplos T com a seguinte distribuição de classes
 - ▣ $\text{Dist}(C1, C2, C3) = (99.00\%, 0.25\%, 0.75\%)$
 - ▣ Prevalência da classe $C1$

Prevalência de classe

29

- Um classificador simples que classifique sempre novos exemplos como pertencentes à classe majoritária C1 teria uma precisão de 99,00% ($\text{maj-err}(T) = 1,00\%$)
- Pode ser indesejável quando as classes minoritárias são aquelas que possuem uma informação muito importante
 - ▣ Por exemplo supondo:
 - C1: paciente normal
 - C2: paciente com doença A
 - C3: paciente com doença B

Prevalência de classe

30

- A maioria dos sistemas de aprendizado é projetada para otimizar a precisão
 - ▣ Quando se trabalha com conjuntos de exemplos desbalanceados, é desejável utilizar uma medida de desempenho diferente da precisão
- Normalmente os algoritmos apresentam um desempenho ruim se o conjunto de treinamento encontra-se fortemente desbalanceado

Prevalência de classe

31

- Os classificadores induzidos tendem a ser altamente precisos nos exemplos da classe majoritária
 - ▣ Frequentemente classificam incorretamente exemplos das classes minoritárias

- Algumas técnicas foram desenvolvidas para lidar com esse problema
 - ▣ Introdução de custos de classificação incorreta
 - ▣ Remoção de exemplos redundantes ou prejudiciais
 - ▣ A detecção de exemplos de borda e com ruído

Overfitting (Sobreaajuste)

32

- Ocorre quando o modelo se ajusta muito bem aos dados de treinamento, capturando até mesmo padrões irrelevantes e ruídos.
- Como consequência, o modelo tem um desempenho excelente nos dados de treinamento, mas não generaliza bem para novos dados (dados de teste ou do mundo real).
- Indícios: erro baixo no treino, mas erro alto no teste.
- Soluções:
 - ▣ aumentar os dados de treinamento,
 - ▣ simplificar o modelo
 - ▣ usar técnicas como dropout (em redes neurais).

Underfitting (Subajuste)

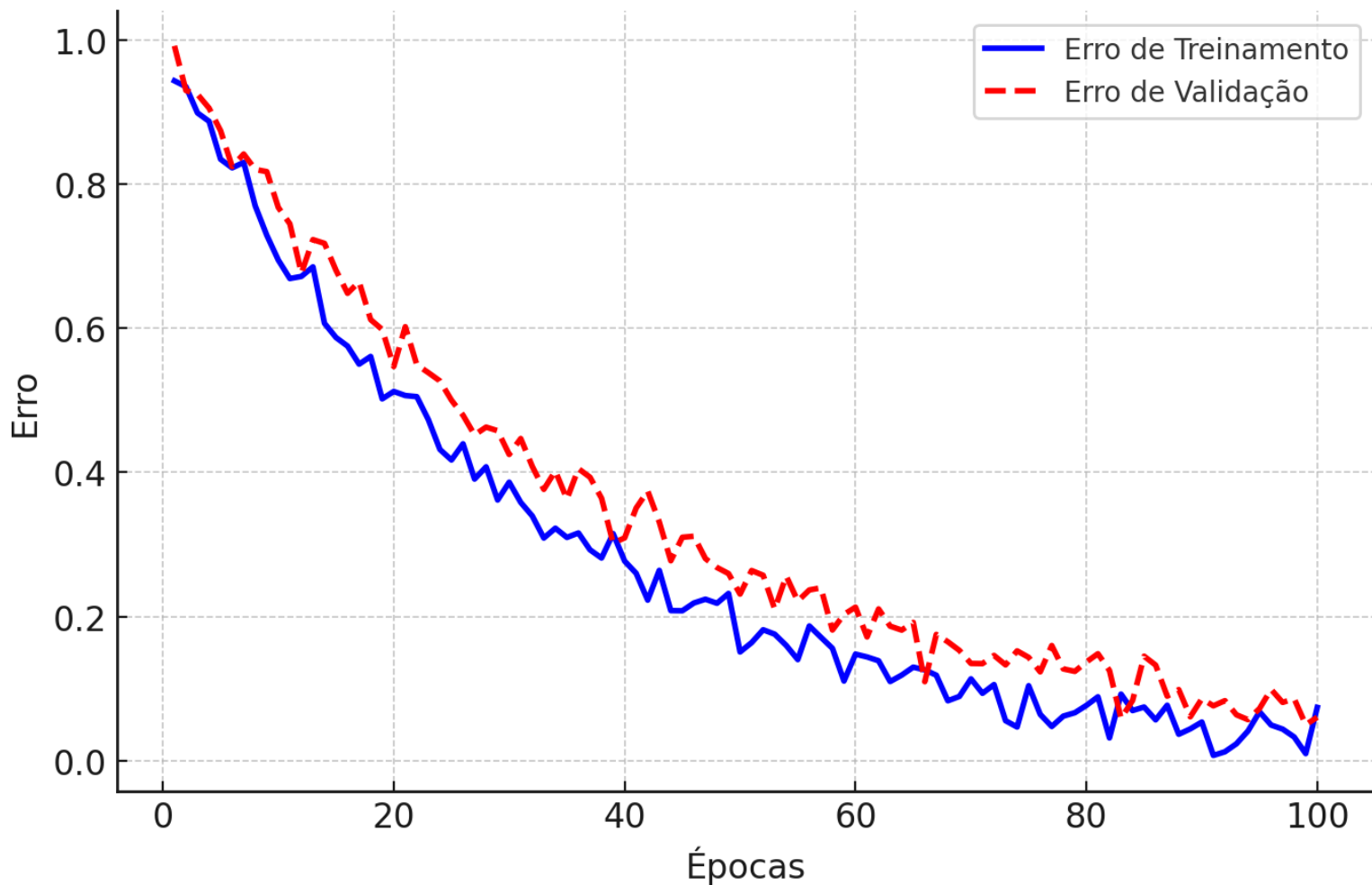
33

- Acontece quando o modelo é muito simples e não consegue capturar os padrões relevantes nos dados.
- O desempenho é ruim tanto nos dados de treinamento quanto nos de teste.
- Indícios: erro alto no treino e no teste.
- Soluções: aumentar a complexidade do modelo, ajustar melhor os hiperparâmetros, treinar por mais tempo ou usar mais dados.

Curva de Erro

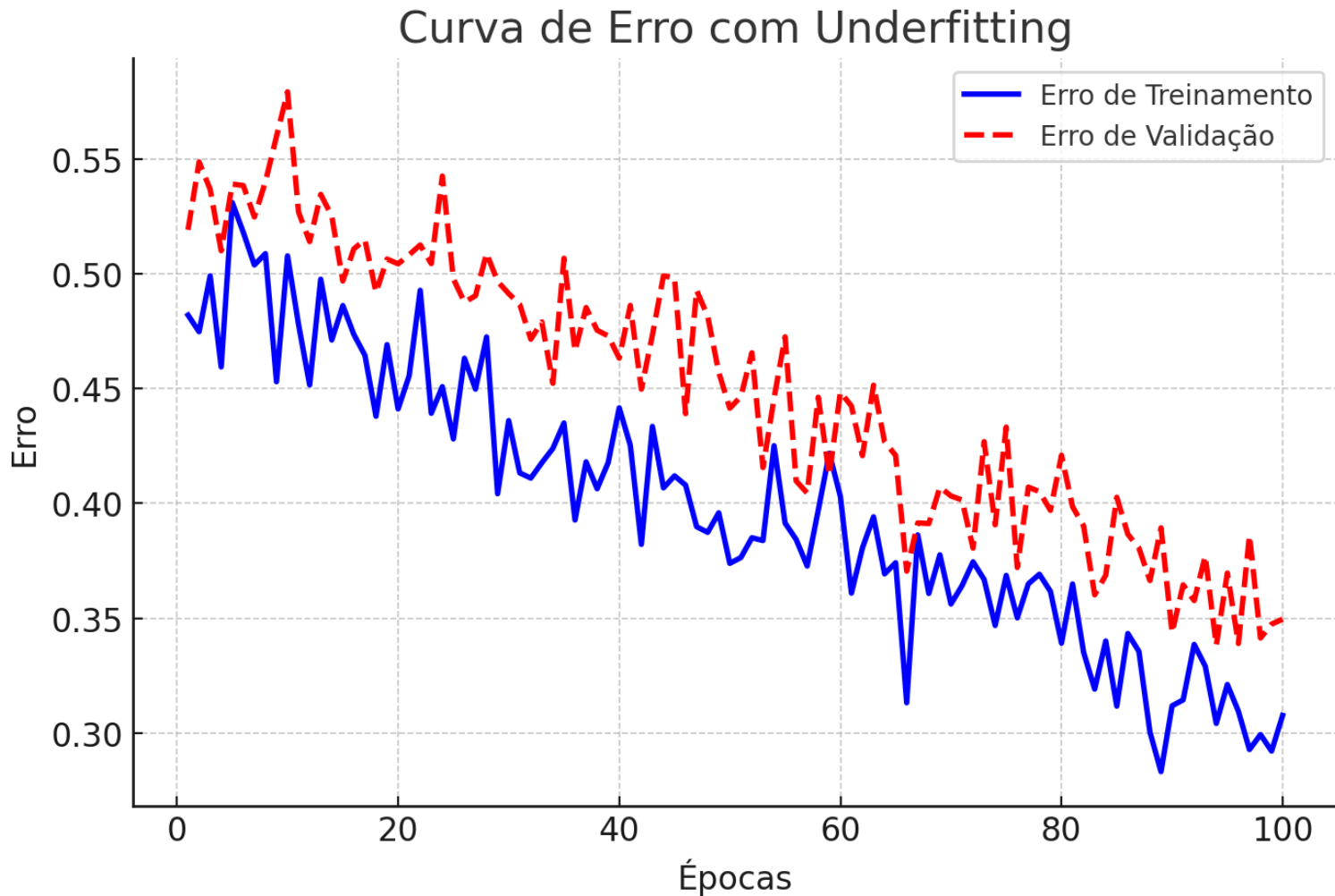
34

Curva de Erro Ideal no Treinamento



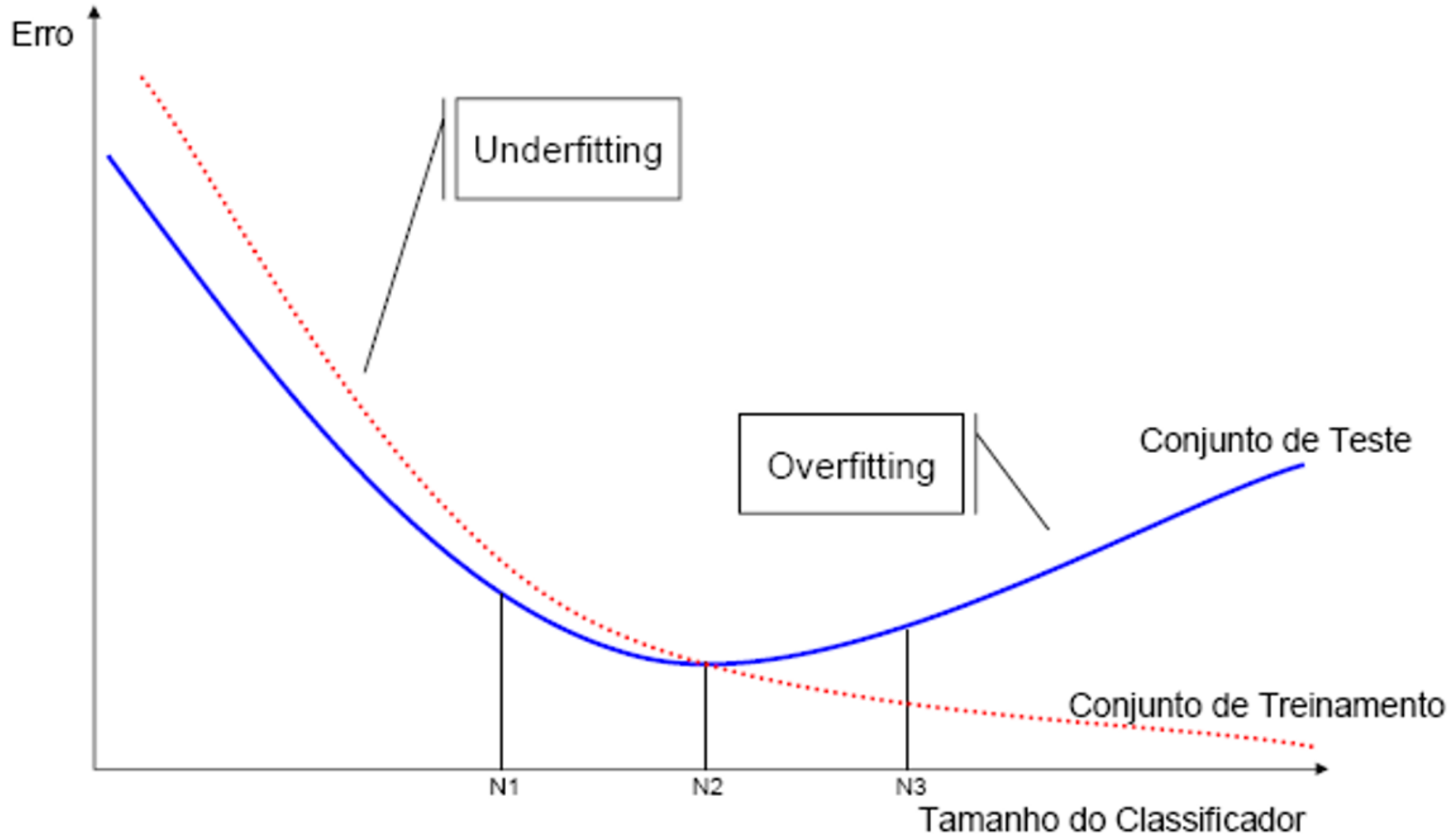
Curva de Erro com overfitting

35



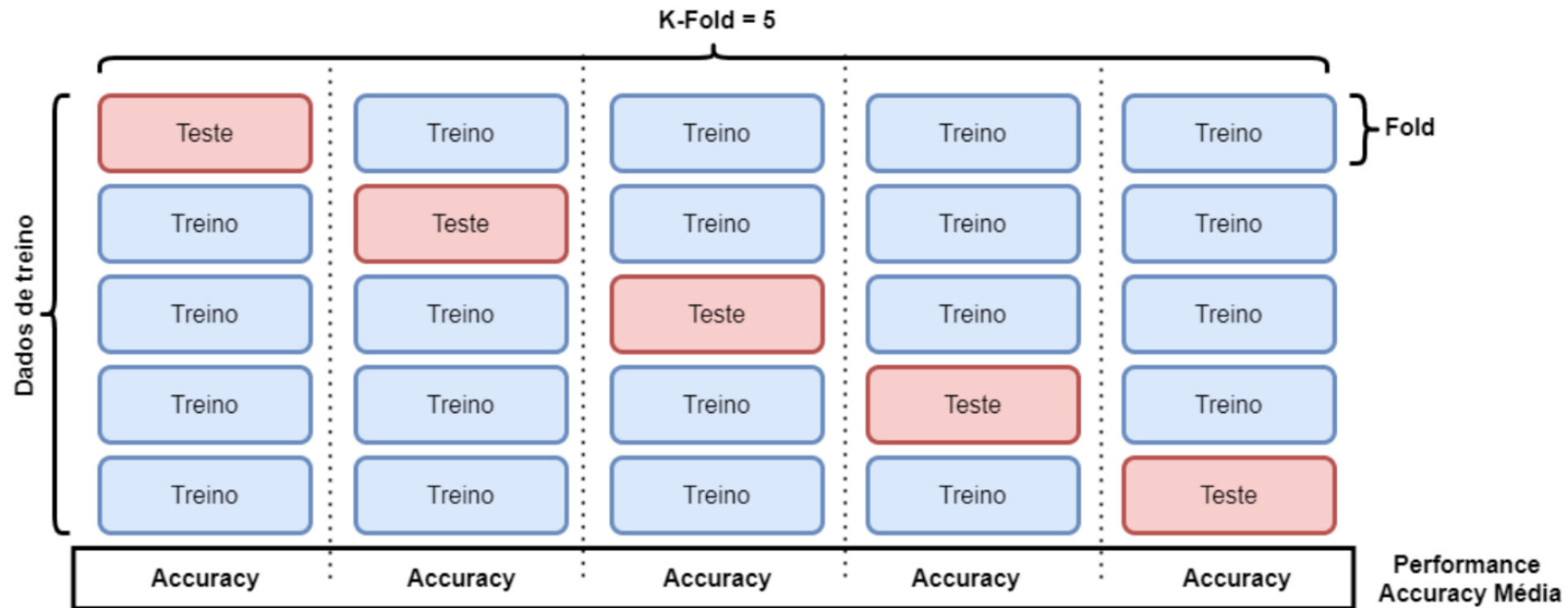
Relação entre o tamanho do classificador e o erro

36



Validação cruzada

37



Síntese da aula

38

□ Medição do desempenho

- Permite quantificar a qualidade do modelo em diferentes cenários, identificando se ele atende aos requisitos do problema.

□ Identificação de Overfitting e Underfitting

- Avaliar as métricas em dados de treinamento e validação ajuda a detectar modelos que memorizam demais (overfitting) ou que generalizam pouco (underfitting).

□ Equilíbrio entre Precisão e Recall

- essencial escolher métricas adequadas, como F1-score, para balancear falsos positivos e falsos negativos

□ Comparação Entre Modelos

- Diferentes classificadores podem ser comparados objetivamente

□ Ajuste Fino e Melhoria Contínua

- Ajustes e melhorias do classificador ao longo do tempo, garantindo eficiência e robustês