

Arquitetura de Dados:

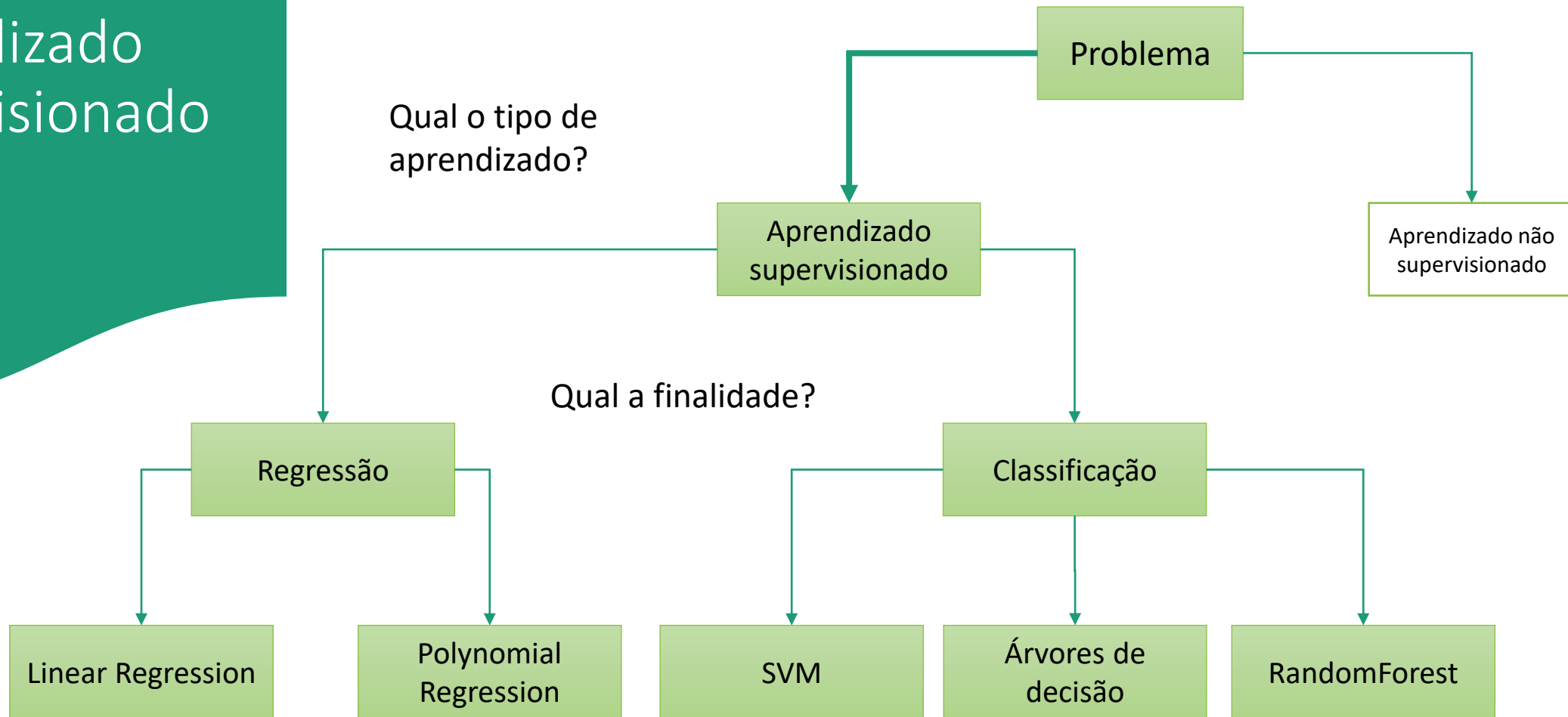
Como saber que as intervenções
nos dados melhoram os resultados?

Matriz de confusão

Prof. Dr. Dieval Guizelini

Aprendizado supervisionado

Qual o algoritmo?



Qual dos diferentes algoritmos irão produzir o “melhor” modelo para o seu problema?

Aprendizado supervisionado

Qual dos diferentes algoritmos irão produzir o “melhor” modelo para o seu problema?

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY	PREDITO	Avaliação
sunny	85	85	false	Don't Play	Don't Play	Concorda
sunny	80	90	true	Don't Play	Play	Discorda
overcast	83	78	false	Play	Play	Concorda
rain	70	96	false	Play	Play	Concorda
rain	68	80	false	Play	Play	Concorda
rain	65	70	true	Don't Play	Don't Play	Concorda
overcast	64	65	true	Play	Play	Concorda
sunny	72	95	false	Don't Play	Play	Discorda
sunny	69	70	false	Play	Play	Concorda
rain	75	80	false	Play	Play	Concorda
sunny	75	70	true	Play	Don't Play	Discorda
overcast	72	90	true	Play	Play	Concorda
overcast	81	75	false	Play	Play	Concorda
rain	71	80	true	Don't Play	Don't Play	Concorda

Fonte: <https://cis.temple.edu/~ingargio/cis587/readings/id3-c45.html>

Aprendizado supervisionado

PLAY	PREDITO	Concorda		Discorda	
		TP	TN	FP	FN
Don't Play	Don't Play		1		
Don't Play	Play			1	
Play	Play	1			
Play	Play	1			
Play	Play	1			
Don't Play	Don't Play		1		
Play	Play	1			
Don't Play	Play			1	
Play	Play	1			
Play	Play	1			
Play	Don't Play				1
Play	Play	1			
Play	Play	1			
Don't Play	Don't Play		1		
		8	3	2	1

		Predito	
		Positivo	Negativo
Real / Conhecido	Verdadeir o	8	2 Erro tipo 2
	Negativo	1 Erro tipo 1	3

Matriz de confusão

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	VP Verdadeiro Positivo	FP Falso Positivo
	negativos	FN Falso Negativo	VN Verdadeiro Negativo

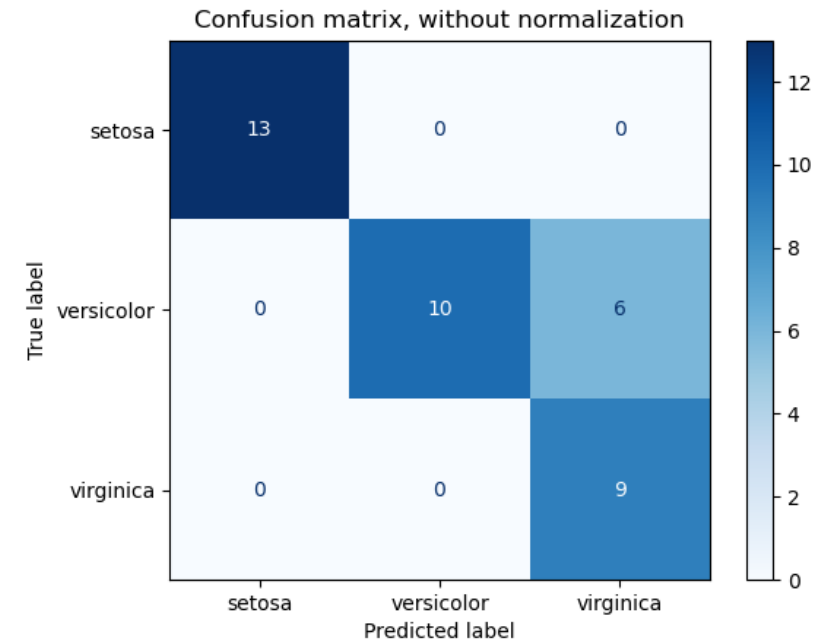
		Predicted	
		+	-
Actual	+	TP	FN Type II error
	-	FP Type I error	TN

Adaptado:

<https://www.nature.com/articles/nmeth.3945>

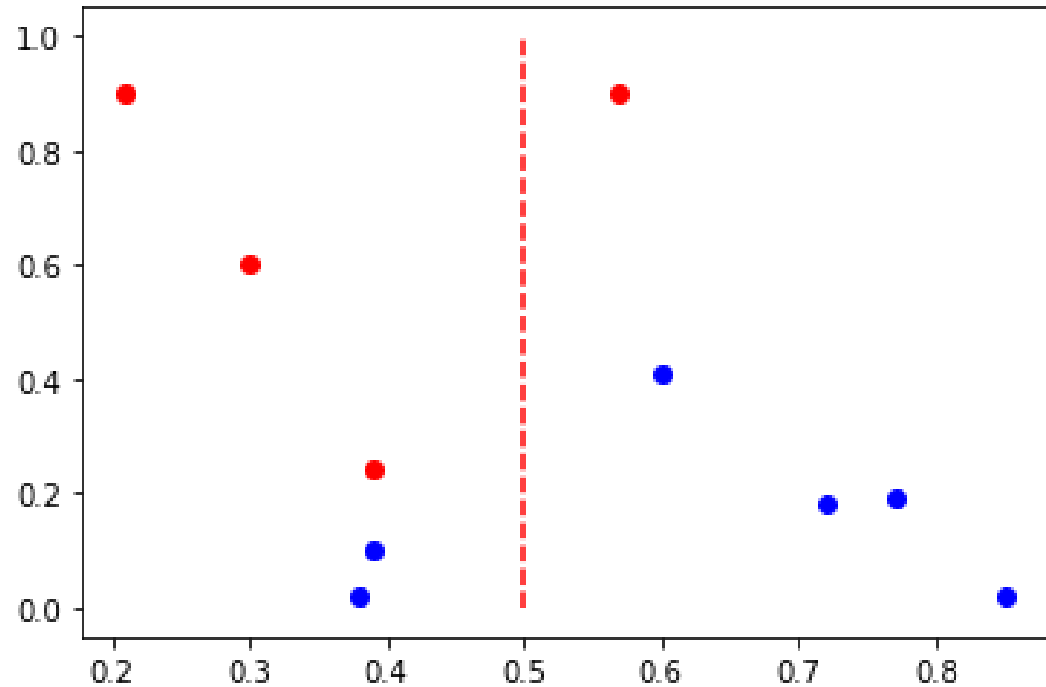
Weka, usa nesse formato

https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html



Mais um exemplo

```
import math
import matplotlib.pyplot as plt
#
# exemplo de um classificado
padroes = [ [0.21, 0.9, 0],
             [.3, 0.6, 0],
             [ 0.39, 0.24, 0],
             [.38, 0.02, 1],
             [.39, 0.1, 1],
             [ 0.57, 0.9, 0],
             [ 0.6, 0.41, 1],
             [ 0.72, 0.18, 1],
             [ 0.77, 0.19, 1],
             [ 0.85, 0.02, 1]
            ]
```

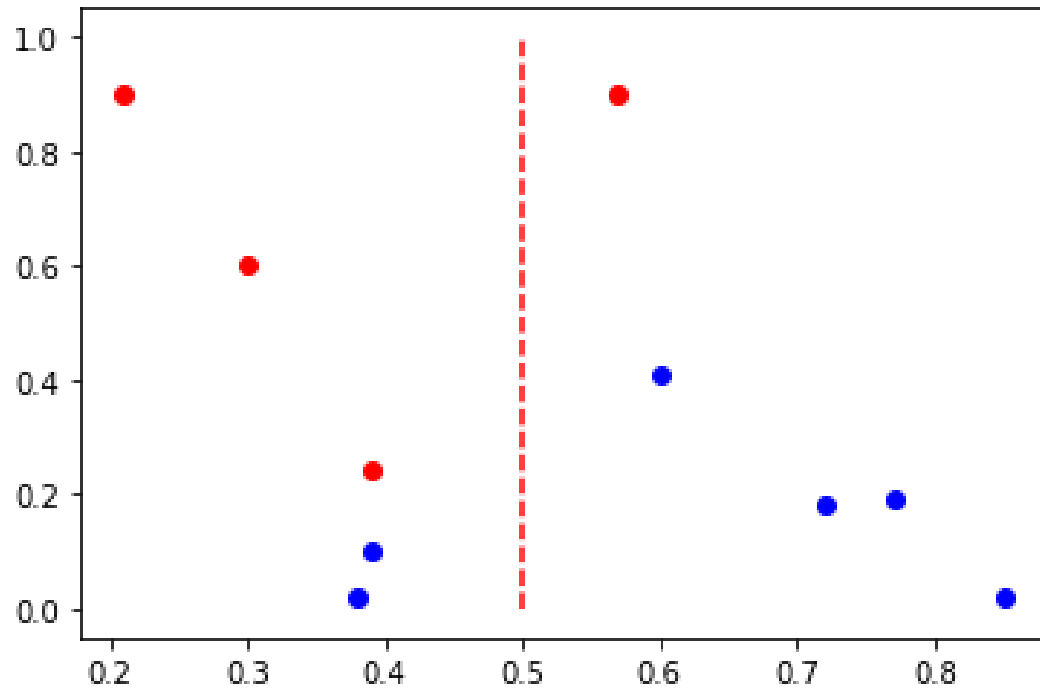


```
padroesCores = ['red', 'blue']

fig, ax = plt.subplots()
for p in padroes:
    ax.scatter(p[0], p[1], c=padroesCores[p[2]])
ax.plot([0.5, 0.5], [0, 1], 'r--')
fig.show()
```

Mais um exemplo

- Classe A
- Classe B



Classe A

Classe B

REAL (Conhecido)

Classe A – Verdadeiro

Classe B – Falso

1) Verdadeiro Positivo (TP):

O modelo previu A e a resposta é A

2) Verdadeiro Negativo (TN):

O modelo previu B e a resposta é B

3) Falso Negativo (FN):

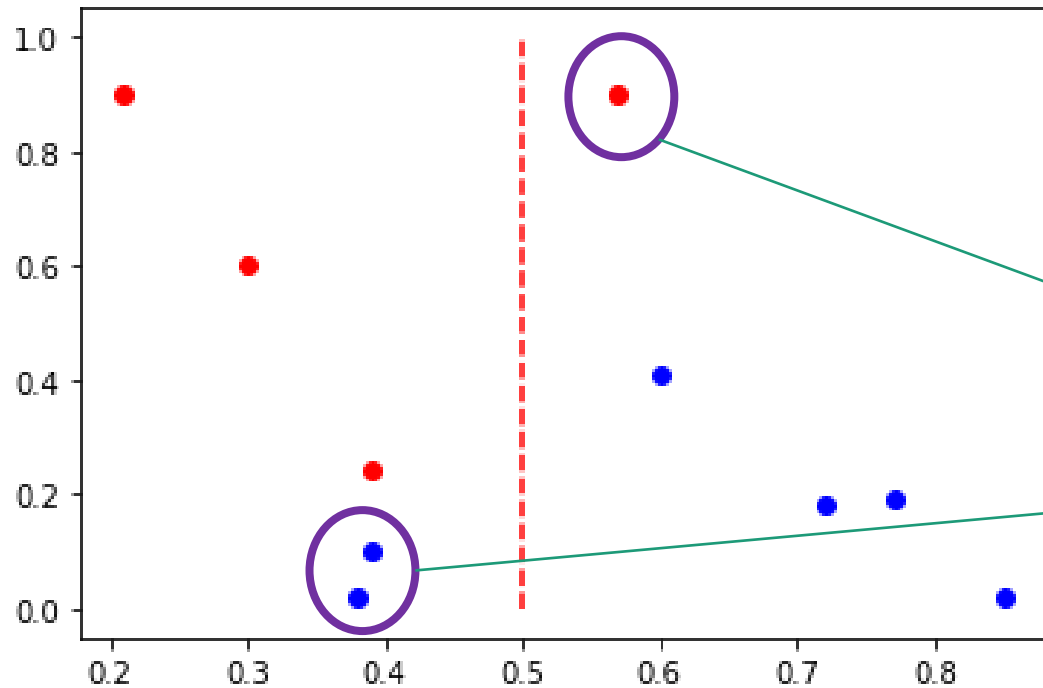
O modelo previu B e a resposta é A

4) Falso Positivo (FP):

O modelo previu A e a resposta é B

Mais um exemplo

- Classe A
- Classe B



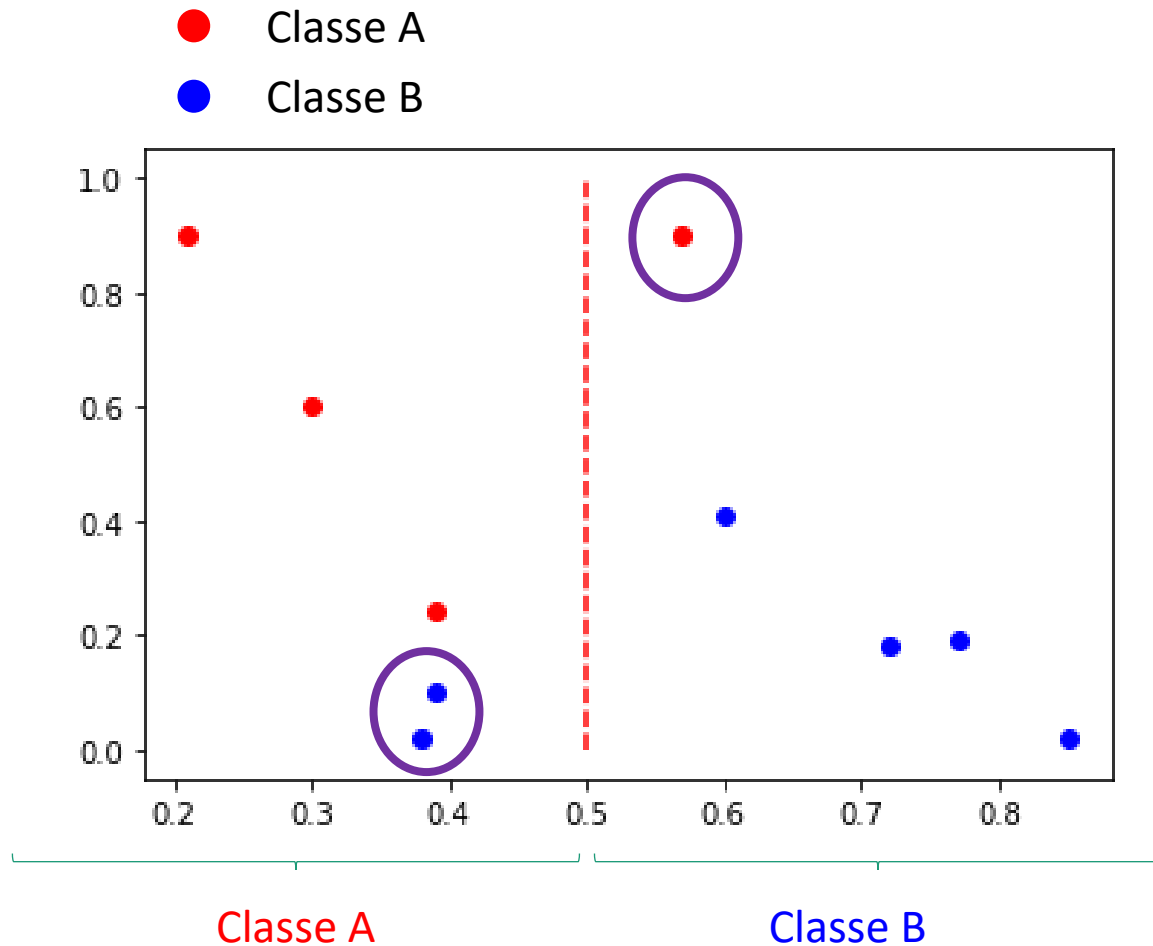
Classe A

Classe B

REAL (Conhecido)
Classe A – Verdadeiro
Classe B – Falso

- 1) Verdadeiro Positivo (TP):
O modelo previu **A** e a resposta é **A**
- 2) Verdadeiro Negativo (TN):
O modelo previu **B** e a resposta é **B**
- 3) Falso Negativo (FN):
O modelo previu **B** e a resposta é **A**
- 4) Falso Positivo (FP):
O modelo previu **A** e a resposta é **B**

Para pensar:



Os dois tipos de erros são iguais?

Existem momentos que um tipo de erro pode ser considerado pior ou melhor que o outro tipo de erro?

Tipo 1 – Falso positivo

Tipo 2 – Falso negativo

Vamos pensar em um modelo que classifique um exame para indicar se a pessoa está ou não com a COVID.

Se o resultado for um falso-positivo, quais as consequências?

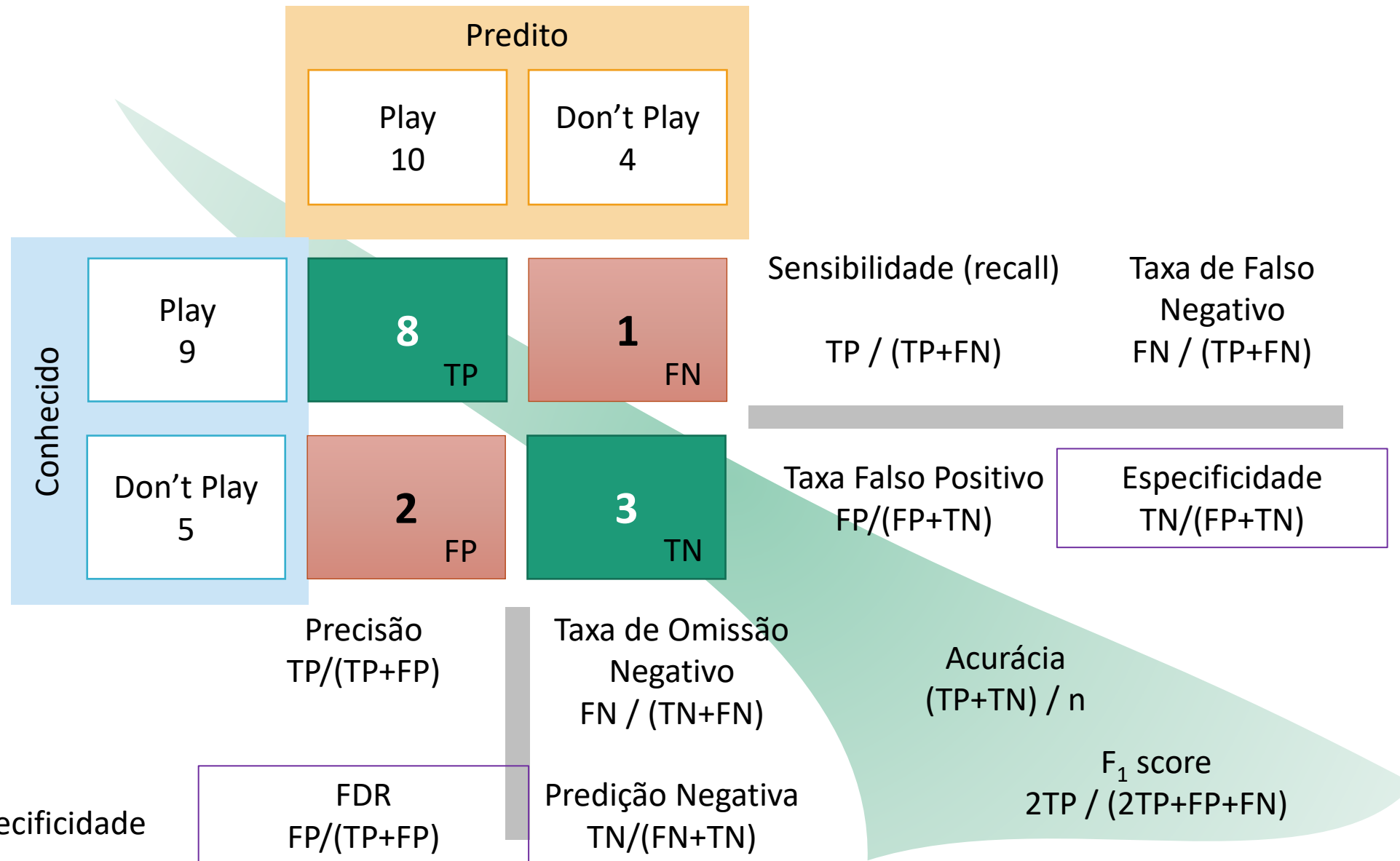
Se o resultado for um falso-negativo, quais as consequências?

Voltando a esse exemplo:

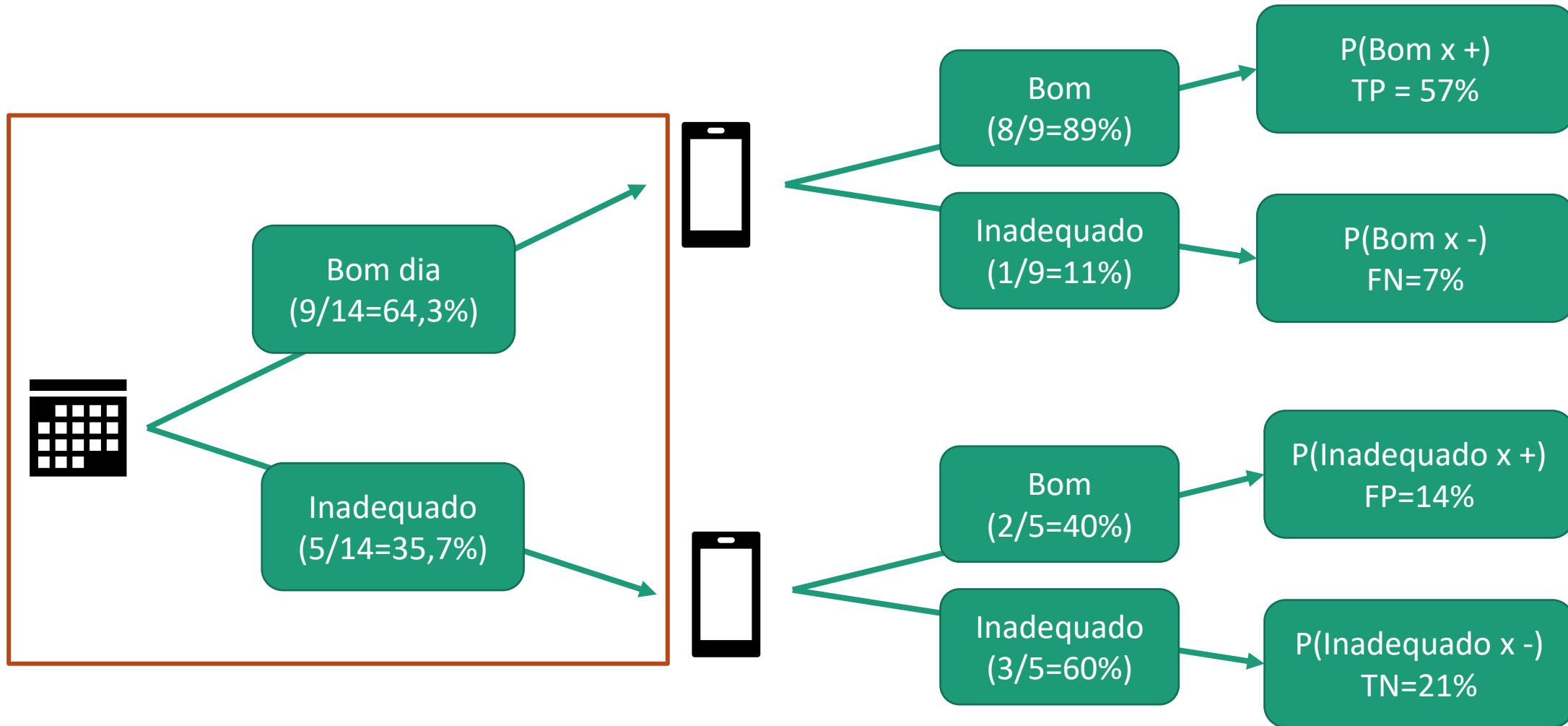
OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY	PREDITO	TP	TN	FP	FN
sunny	85	85	false	Don't Play	Don't Play		1		
sunny	80	90	true	Don't Play	Play			1	
overcast	83	78	false	Play	Play	1			
rain	70	96	false	Play	Play	1			
rain	68	80	false	Play	Play	1			
rain	65	70	true	Don't Play	Don't Play		1		
overcast	64	65	true	Play	Play	1			
sunny	72	95	false	Don't Play	Play			1	
sunny	69	70	false	Play	Play	1			
rain	75	80	false	Play	Play	1			
sunny	75	70	true	Play	Don't Play				1
overcast	72	90	true	Play	Play	1			
overcast	81	75	false	Play	Play	1			
rain	71	80	true	Don't Play	Don't Play		1		
						8	3	2	1

Fonte: <https://cis.temple.edu/~ingargio/cis587/readings/id3-c45.html>

Como construir a matriz de confusão



O dia está bom para o jogo de golf?



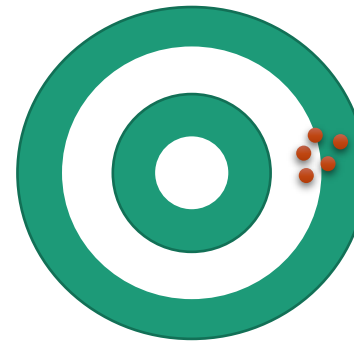
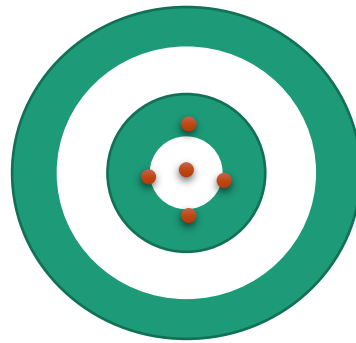
Acurácia

Acurácia significa EXATIDÃO e RIGOR

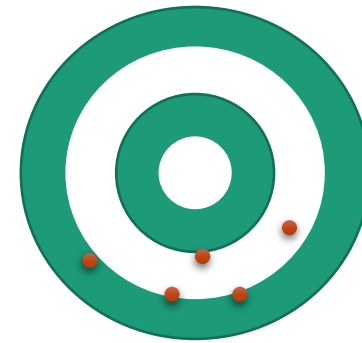
Medidas acuradas são aquelas cujo valor se aproxima do valor correto, ou seja, valor de referência.



+ Exatidão /+ Acurácia



- Exatidão /- Acurácia

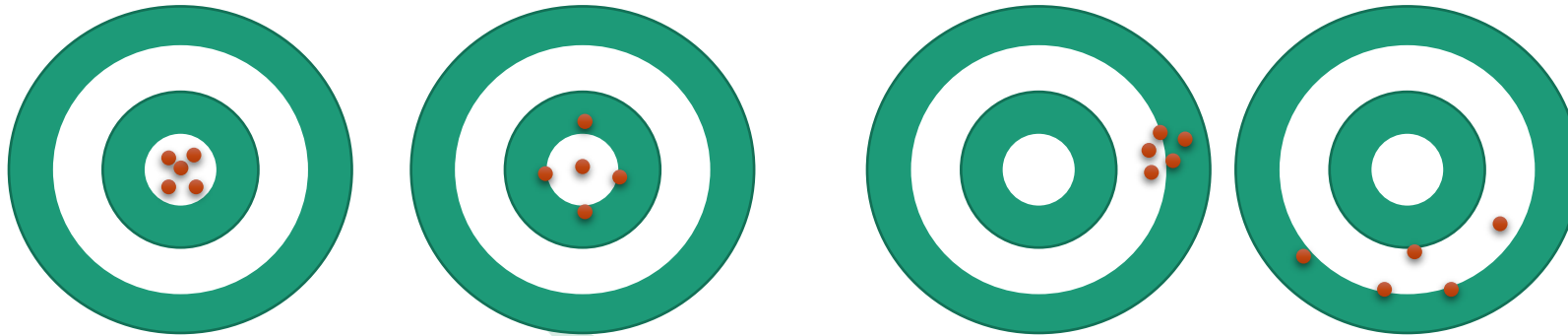


Acurácia: alcance da meta

Acurácia

Acurácia significa EXATIDÃO e RIGOR

Medidas acuradas são aquelas cujo valor se aproxima do valor correto, ou seja, valor de referência.



+ Exatidão /+ Acurácia

- Exatidão /- Acurácia

Acurácia: alcance da meta
Ou quanto o modelo acerta.

Acurácia:
 $11/14=0,785$ ou 78,5%

8	1
2	3

Acurácia
 $(TP+TN) / n$

O problema com a Acurácia:

- Considere um classificador para avaliar os resultados de um exame e indicar se uma pessoa tem ou não uma determinada doença. Considere que a base de dados para treinamento é formada por 90% dos casos de pessoas sem a doença e 10% de pessoas com a doença.
- Agora, se o modelo SEMPRE responder que não tem doença, qual a acurácia do modelo?

O problema com a Acurácia:

- Considere um classificador para avaliar os resultados de um exame e indicar se uma pessoa tem ou não uma determinada doença. Considere que a base de dados para treinamento é formada por 90% dos casos de pessoas sem a doença e 10% de pessoas com a doença.
- Agora, se o modelo SEMPRE responder que não tem doença, qual a acurácia do modelo?
- A acurácia será de 90%, mas totalmente inútil.

O que podemos fazer para reduzir os efeitos do problema da prevalência?

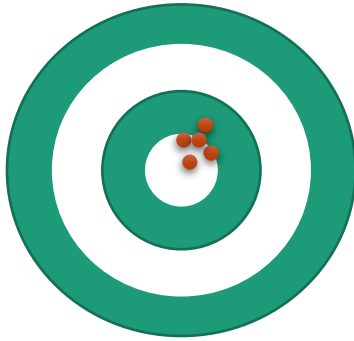
- Prevalência está relacionada a falta de equilíbrio entre as classes.
- Na saúde, usa-se dois conceitos relacionados:
 - Prevalência: a “carga” da doença na população (percentual)
 - Incidência: a “velocidade” em que a população está sendo acometida.

O que podemos fazer para reduzir os efeitos do problema da prevalência?

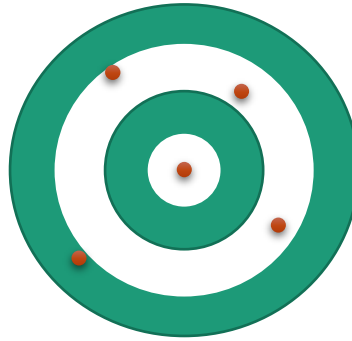
- Prevalência está relacionada a falta de equilíbrio entre as classes.
- Na saúde, usa-se dois conceitos relacionados:
 - Prevalência: a “carga” da doença na população (percentual)
 - Incidência: a “velocidade” em que a população está sendo acometida.
- Soluções possíveis:
 - Amostragem proporcional as classes na etapa de treinamento

Precisão

O conceito de precisão está associado a ideia de medida e de repetição.
“Quanto mais preciso é uma medição, menor será a variação entre os valores medidos.”



Preciso =
Pequena dispersão



Impreciso / pouco preciso =
Grande dispersão

Precisão $TP/(TP+FP)$

$8 / (8+2) = 0,8$ ou 80% preciso

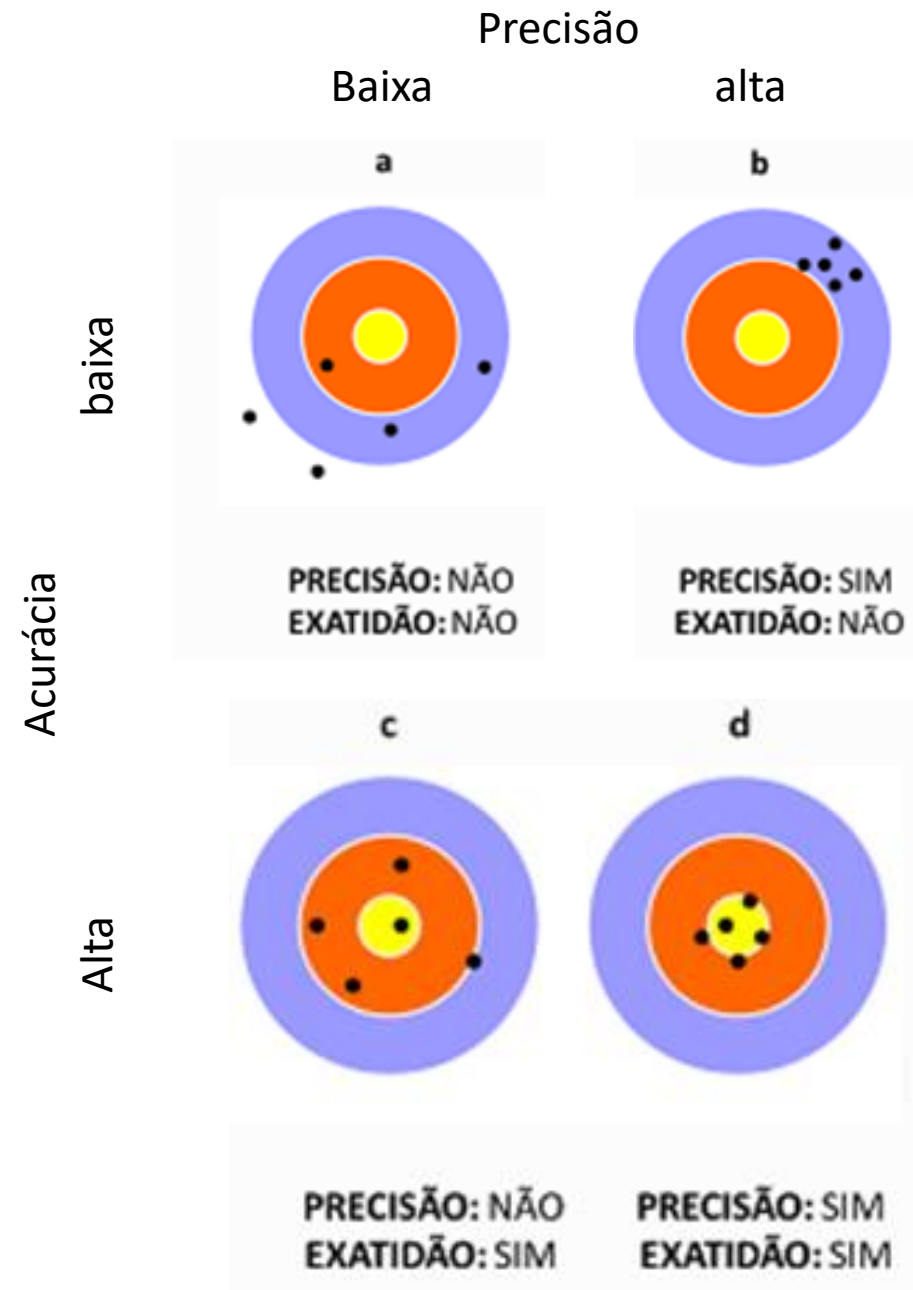
8	1
2	3

Precisão

- A Precisão é utilizada para indicar a relação entre as previsões positivas realizadas corretamente e todas as previsões positivas (incluindo as falsas).
- Em outras palavras, nos dá uma indicação do Falso-positivo.
- Qual o percentual de itens classificados como relevantes são realmente relevantes?

O problema com a precisão:

- Considere uma base com 1.000 padrões, onde apenas 100 destes padrões são positivos.
- Se o modelo indique apenas 1 dos casos positivos, a precisão continuará sendo de 100%
- Isso porque os falsos negativos não são considerados nessa métrica.
- A principal utilização dessa métrica é para modelos onde é preciso minimizar os falsos positivos. Neste caso, quanto mais perto dos 100% chegarmos, melhor.



Acurácia
e
precisão

Recall ou sensibilidade

- A métrica Recall é utilizada para indicar a relação entre as previsões positivas realizadas corretamente e todas as previsões que realmente são positivas (TP e FN)
- De todos os padrões que realmente são positivos, qual percentual é identificado corretamente pelo modelo?

Sensibilidade (recall)

$TP / (TP+FN)$

$8 / (8+1) = 0,88$

8	1
2	3

Recall ou sensibilidade

- A Recall é bastante útil quando precisamos **minimizar os falsos negativos**.
- Isso é especialmente útil para casos de diagnósticos, onde pode haver um dano muito maior em não identificar uma doença, do que identificá-la em pacientes saudáveis.
- Sempre que precisarmos minimizar os falsos negativos devemos buscar uma maior percentual no recall.

Especificidade

$$\text{Especificidade} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

- É a capacidade de acertar o Negativo entre os negativos.
- é a capacidade do teste de detectar pacientes que não possuem câncer, dentre aqueles que de fato não possuem.

Uma discussão baseada na probabilidade bayesiana

$$\textit{precisão} = \frac{P(A) \cap P(B)}{P(B)}$$

$$\textit{precisão} = \frac{TP}{TP + \textcolor{red}{FP}}$$

$$\textit{especificidade} = \frac{TN}{TN + \textcolor{red}{FP}}$$

$$\textit{sensibilidade} = \frac{P(A) \cap P(B)}{P(A)}$$

$$\textit{sensibilidade} = \frac{TP}{TP + \textcolor{red}{FN}}$$

Taxas (proporções)

$$\textit{precisão} = \frac{P(A) \cap P(B)}{P(B)}$$

$$\textit{precisão} = \frac{TP}{TP + \textcolor{red}{FP}}$$

$$\textit{especificidade} = \frac{TN}{TN + \textcolor{red}{FP}}$$

$$\textit{sensibilidade} = \frac{P(A) \cap P(B)}{P(A)}$$

$$\textit{sensibilidade} = \frac{TP}{TP + \textcolor{red}{FN}}$$

Taxa de verdadeiros positivos

$$\text{Taxa falso-positivo} = \frac{FP}{\textcolor{red}{FP} + \textcolor{red}{TN}}$$

Taxa de falso-positivo

$$\text{Taxa falso-negativo} = \frac{FN}{TP + \textcolor{red}{FN}}$$

Taxa de falso-negativo

F1 Score

F1 Score é a média harmônica da precisão e da sensibilidade (recall).

$$\text{precisão} = \frac{TP}{TP + FP}$$

$$\text{sensibilidade} = \frac{TP}{TP + FN}$$

$$M_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

$$F_1\text{score} = \frac{2TP}{2TP + FP + FN}$$

A média harmônica está relacionado a grandezas inversamente proporcionais.

Quando os valores são próximos, o resultado é semelhante a média aritmética.

O F1 Score é uma métrica melhor que a Acurácia, principalmente em casos onde falsos positivos e falsos negativos possuem impactos diferentes para seu modelo. O F1 Score cria um resultado a partir dessas divergências.

Observe que na fórmula não entra o **Verdadeiro-Negativo (TN)**.

Validade

- SENSIBILIDADE: é a capacidade do teste de identificar corretamente os indivíduos que possuem a doença (casos)
- ESPECIFICIDADE: é a capacidade do teste de identificar corretamente os indivíduos que não possuem a doença (não-casos)

Curva ROC (*Receiver Operating Characteristic*)

Sensibilidade é a capacidade de detectar pacientes com câncer, dentre aqueles que de fato possuem câncer, por exemplo

$$\text{Sensibilidade (recall)} \\ \text{TP} / (\text{TP} + \text{FN})$$

Especificidade é a capacidade de detectar pacientes que não possuem câncer, dentre aqueles que de fato não possuem

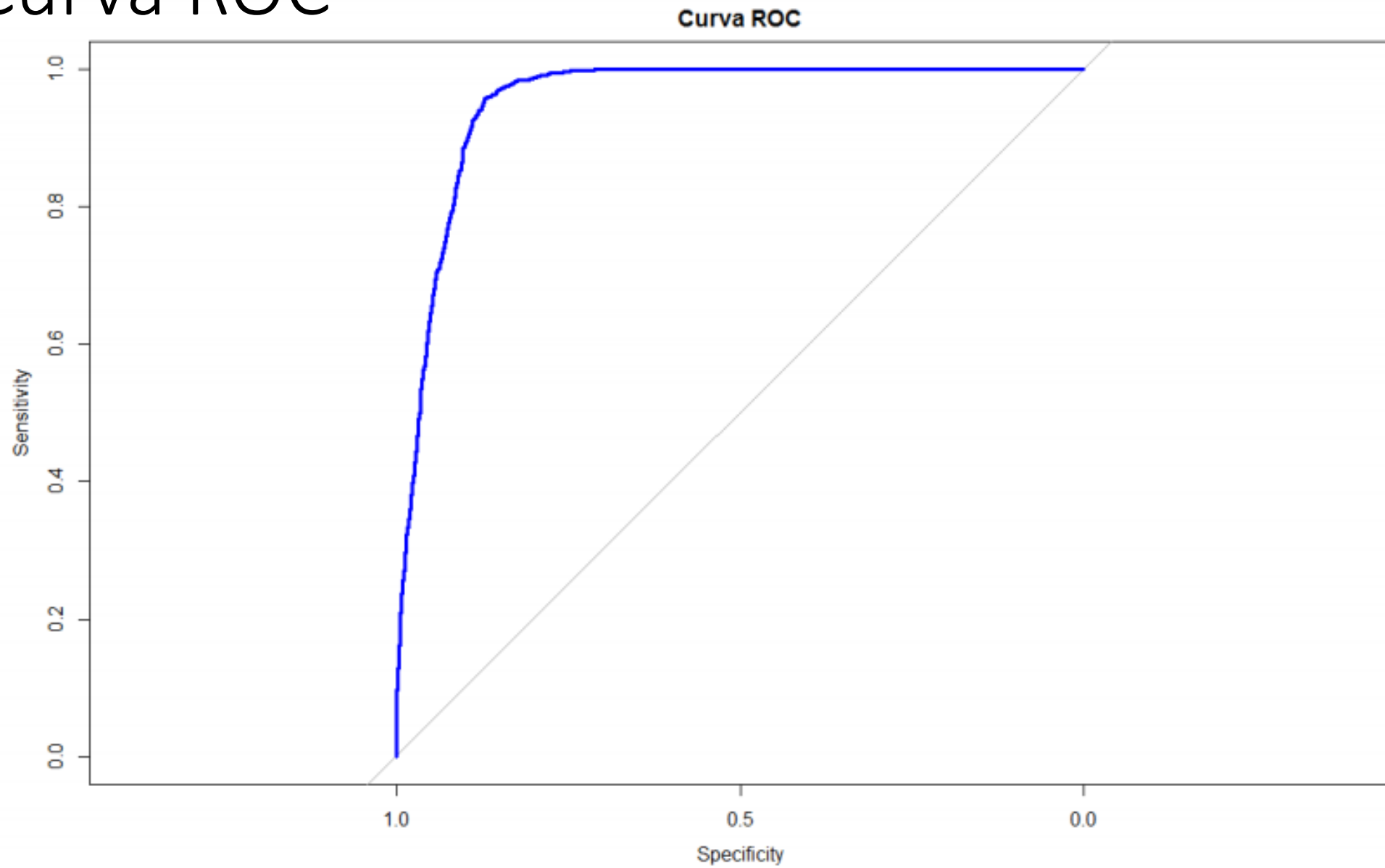
$$\text{Especificidade} \\ \text{TN} / (\text{FP} + \text{TN})$$

Quando aumentamos o valor de corte, aumentamos a **sensibilidade** e reduzimos a **especificidade**

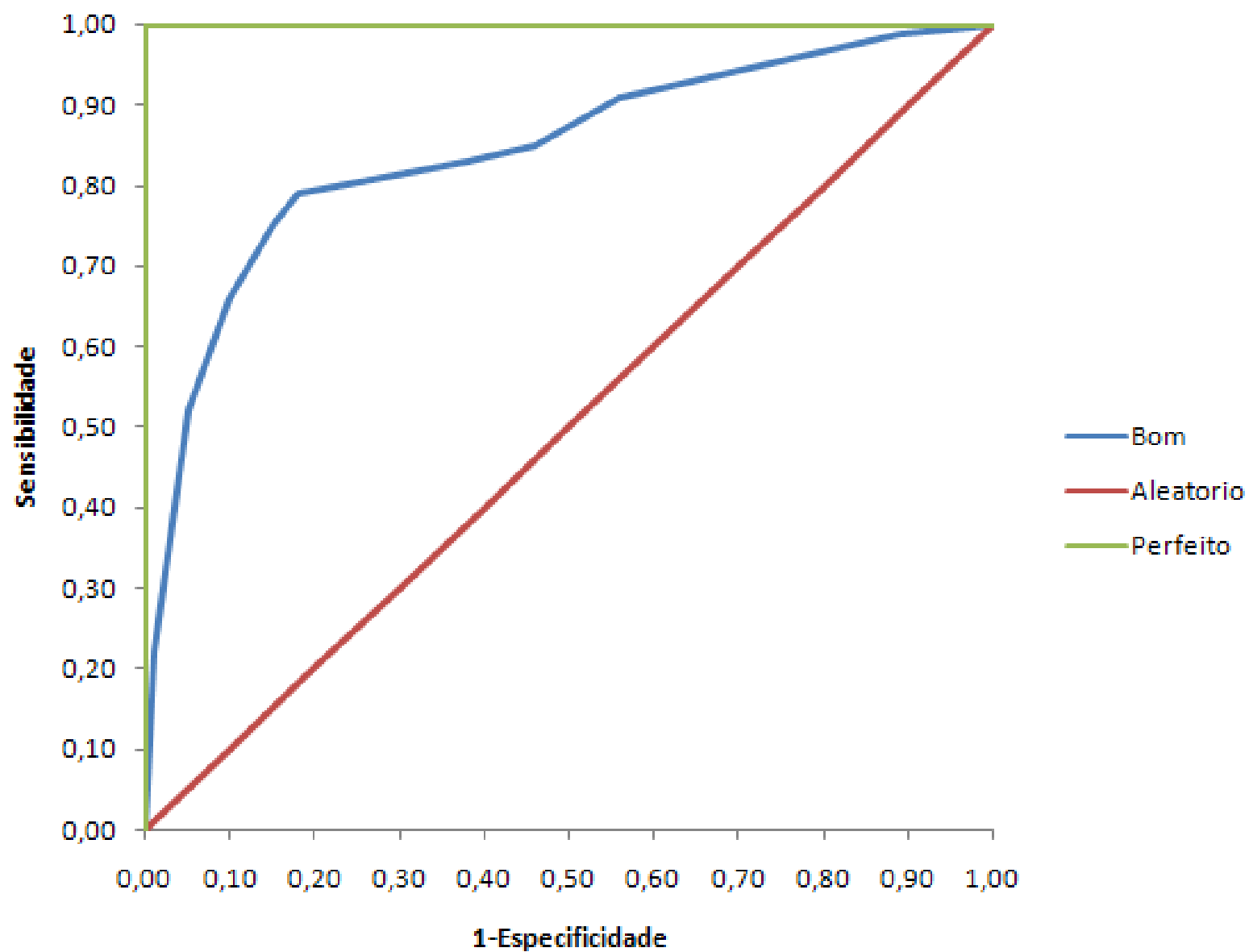
A curva ROC nos fornece um gráfico da sensibilidade versus a especificidade, quando aumentamos o valor de corte

A curva sob a curva, conhecida com AUC (area under the curve), é usada como uma medida de qualidade

Curva ROC



Curvas de Diferentes Classificadores



Tipos de Erros

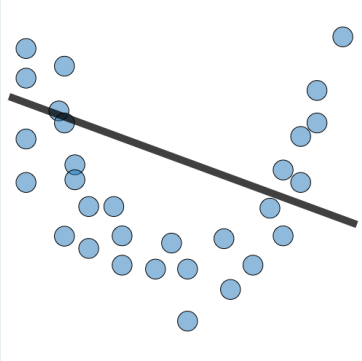
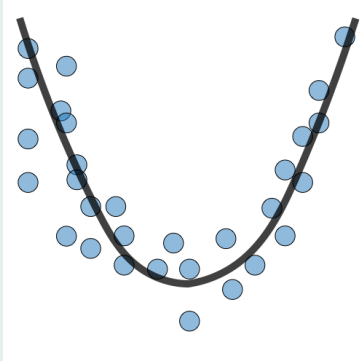
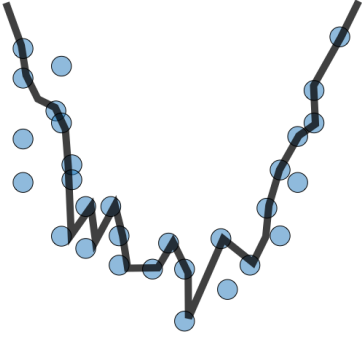
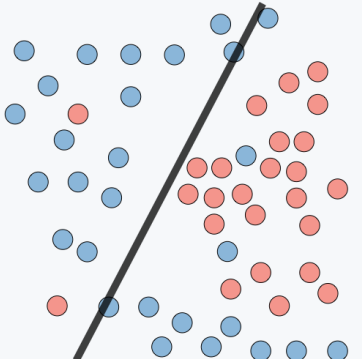
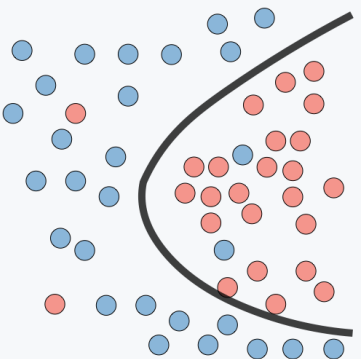
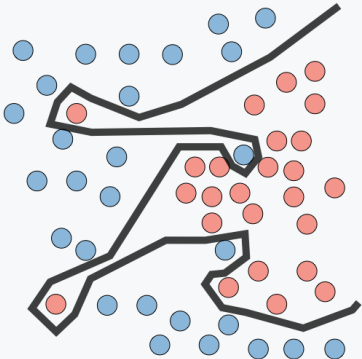
- Precisão = reprodutibilidade (erro aleatório / acaso)
- Validade = acurácia (erro sistemático / viés / bias)
- Falso-Positivo = Erro tipo I
- Falso-Negativo = Erro tipo II

Para mais de uma classe...

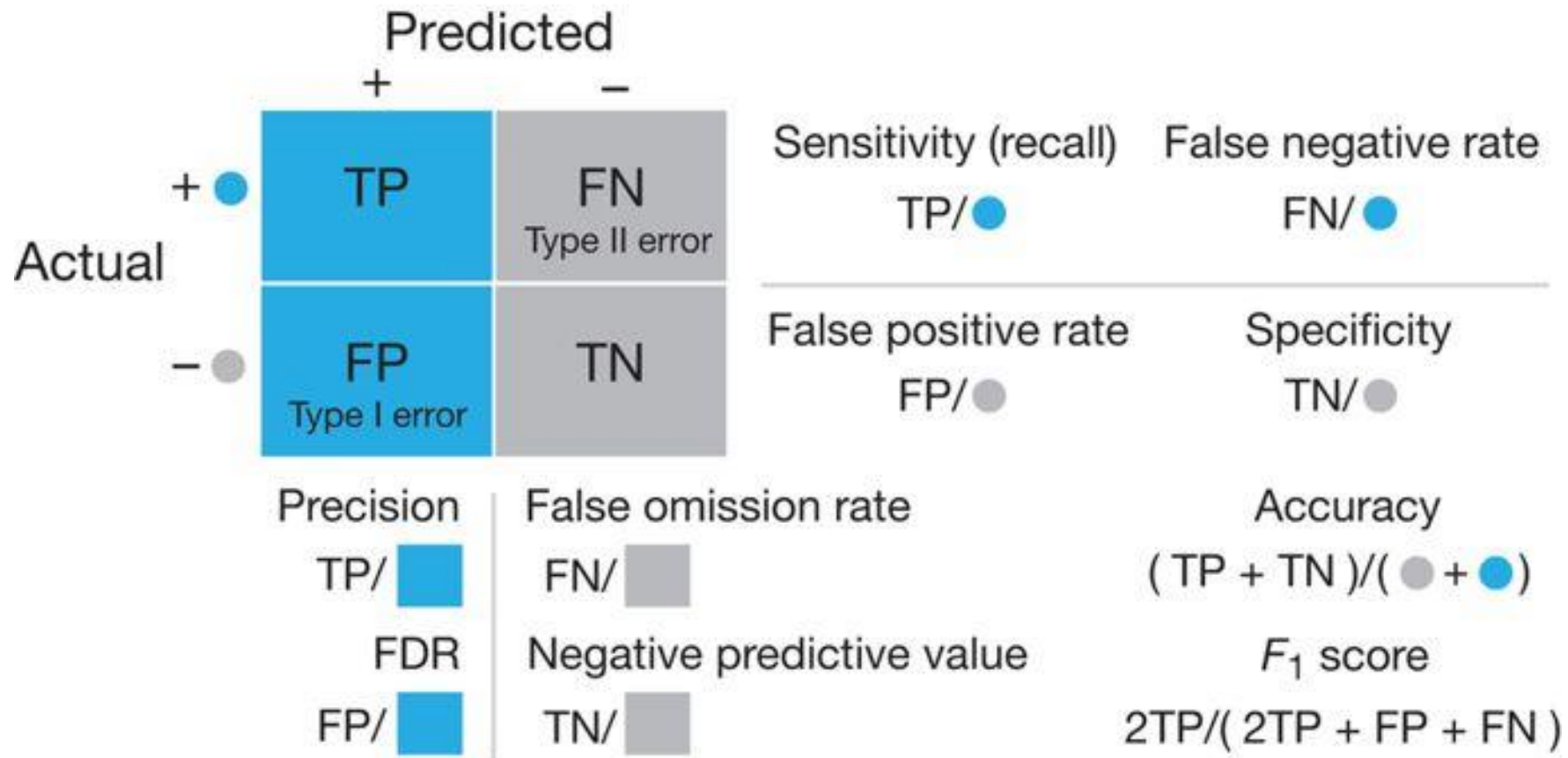
		Previsto		
		realizarPedido	obterStatus	horarioAtendimento
Classe Real	realizarPedido	10	0	0
	obterStatus	0	9	0
	horarioAtendimento	0	0	12

Crie 1 matriz para cada classe (é ou não é a classe)
Calcule a média das taxas para obter o resultado final.

O que precisamos evitar...

	Underfitting	Just right	Overfitting
Sintomas	<ul style="list-style-type: none">• Erro de treinamento elevado• Erro de treinamento próximo ao erro de teste• Viés elevado	<ul style="list-style-type: none">• Erro de treinamento ligeiramente menor que erro de teste	<ul style="list-style-type: none">• Erro de treinamento muito baixo• Erro de treinamento muito menor que erro de teste• Alta variância
Exemplo de regressão			
Exemplo de classificação			

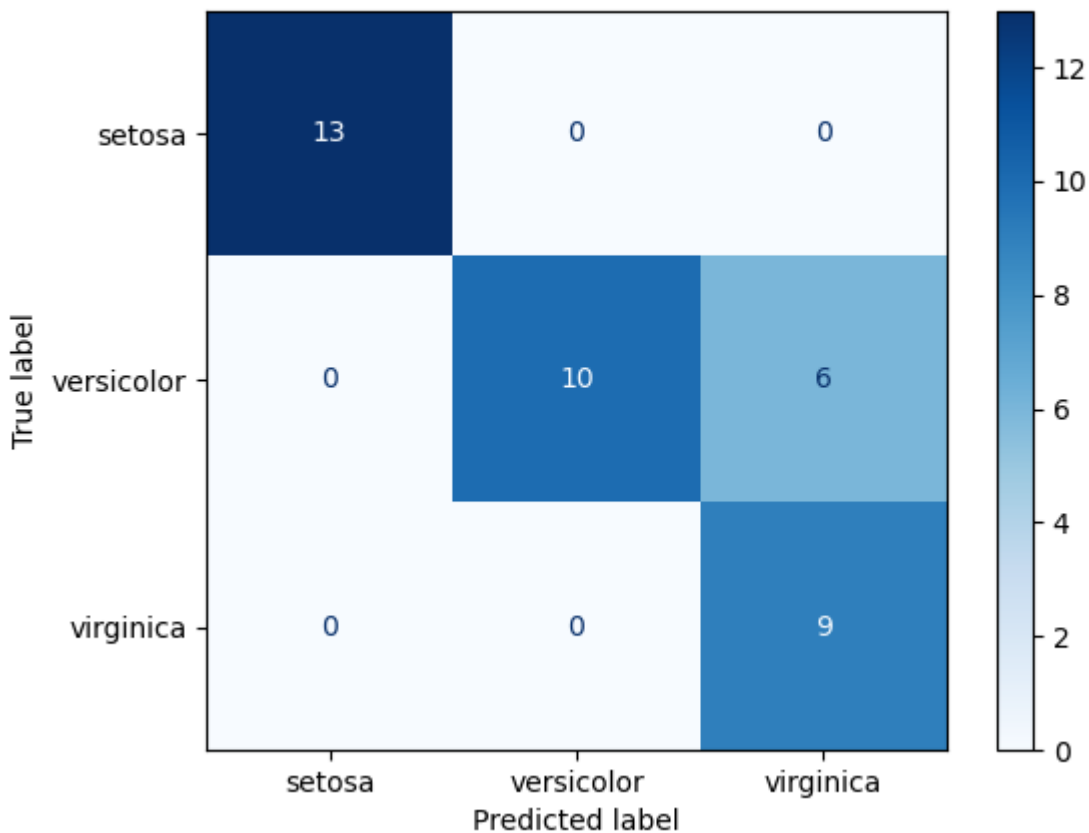
Matriz de confusão



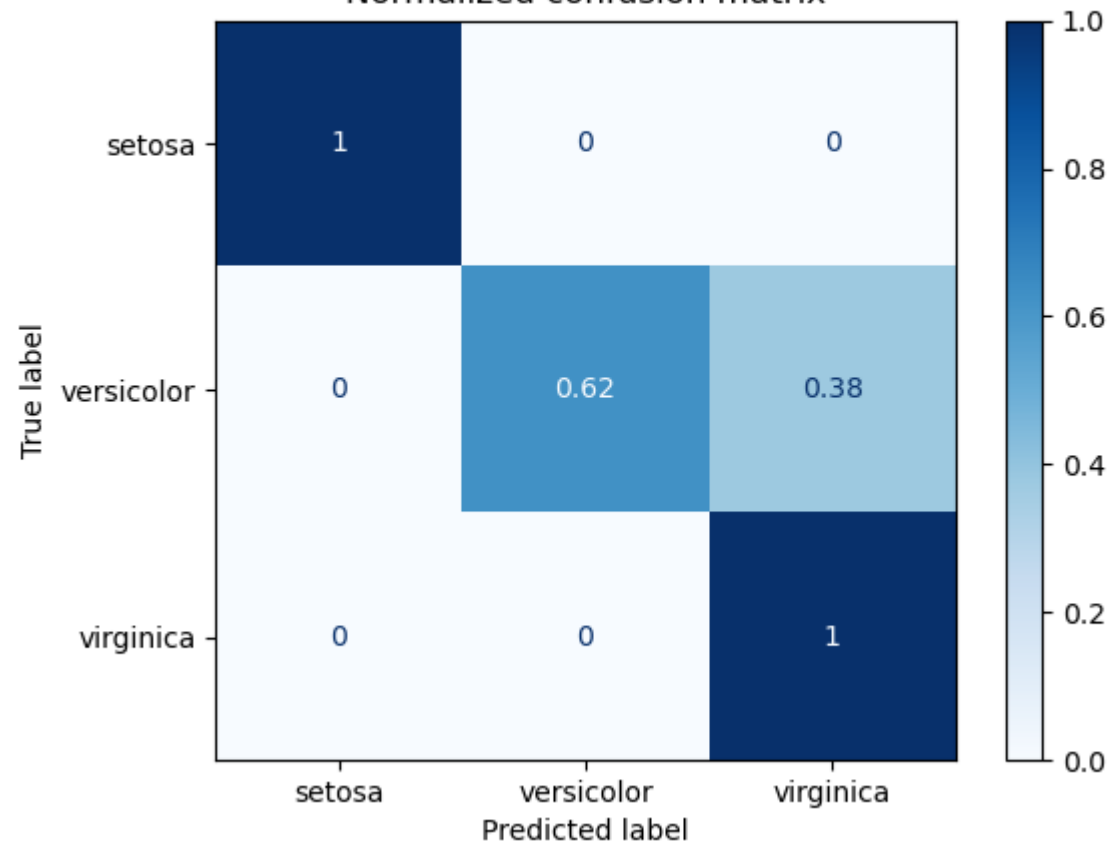


Confusion matrix

Confusion matrix, without normalization



Normalized confusion matrix



Fonte: https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

Voltando ao colab

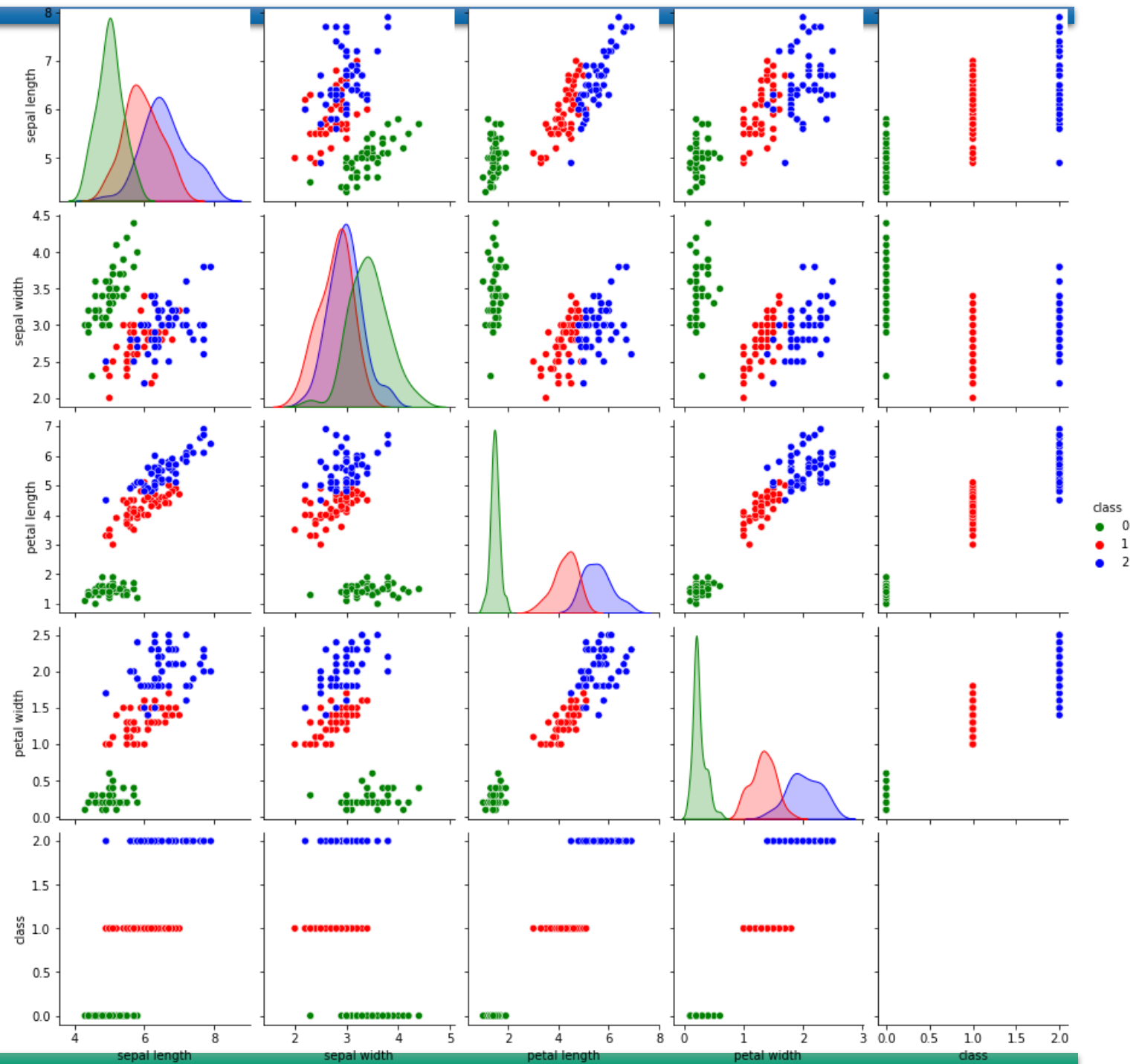
```
from sklearn.datasets import load_iris
import seaborn as sns
import pandas as pd
```

```
dados = load_iris()
#print(dados.DESCR)
nomes = list(dados.target_names)
classes = dados.target
cols=['sepal length','sepal width',
      'petal length','petal width','class']
```

```
#print(dados)
padroes = []
for i in range(len(dados.data)):
    d = dados.data[i]
    p = dados.target[i]
    novo = list(d)
    novo.append(p)
    padroes.append(novo)
```

```
tabela = pd.DataFrame(data=padroes,
                      columns=cols)
cores = {0:'green' ,1:'red',2:'blue' }
sns.pairplot(tabela,vars=cols,
             hue='class',palette=cores)
```

Voltando ao colab



Atividade

- Considerando os critérios abaixo, monte a matriz de confusão indicando o resultado da classificação da bases de dados íris:

petalwidth \leq 0.6: Iris-setosa

petalwidth $>$ 0.6

| petalwidth \leq 1.7

| | petallength \leq 4.9: Iris-versicolor

| | petallength $>$ 4.9

| | | petalwidth \leq 1.5: Iris-virginica

| | | petalwidth $>$ 1.5: Iris-versicolor

| petalwidth $>$ 1.7: Iris-virginica

Bibliografia

- Jake LEVER, Martin KRZYWINSKI, Naomi ALTMAN. Classification evaluation. Nature Methods volume 13, pages 603–604 (2016)
<https://www.nature.com/articles/nmeth.3945>