

Ciência de Dados

Planejamento de experimentos

Revisado: Roseli Romero
SCC-ICMC-USP

Prof. Dr. André C. P. L. F. de Carvalho
Dr. Isvani Frias-Blanco
ICMC-USP

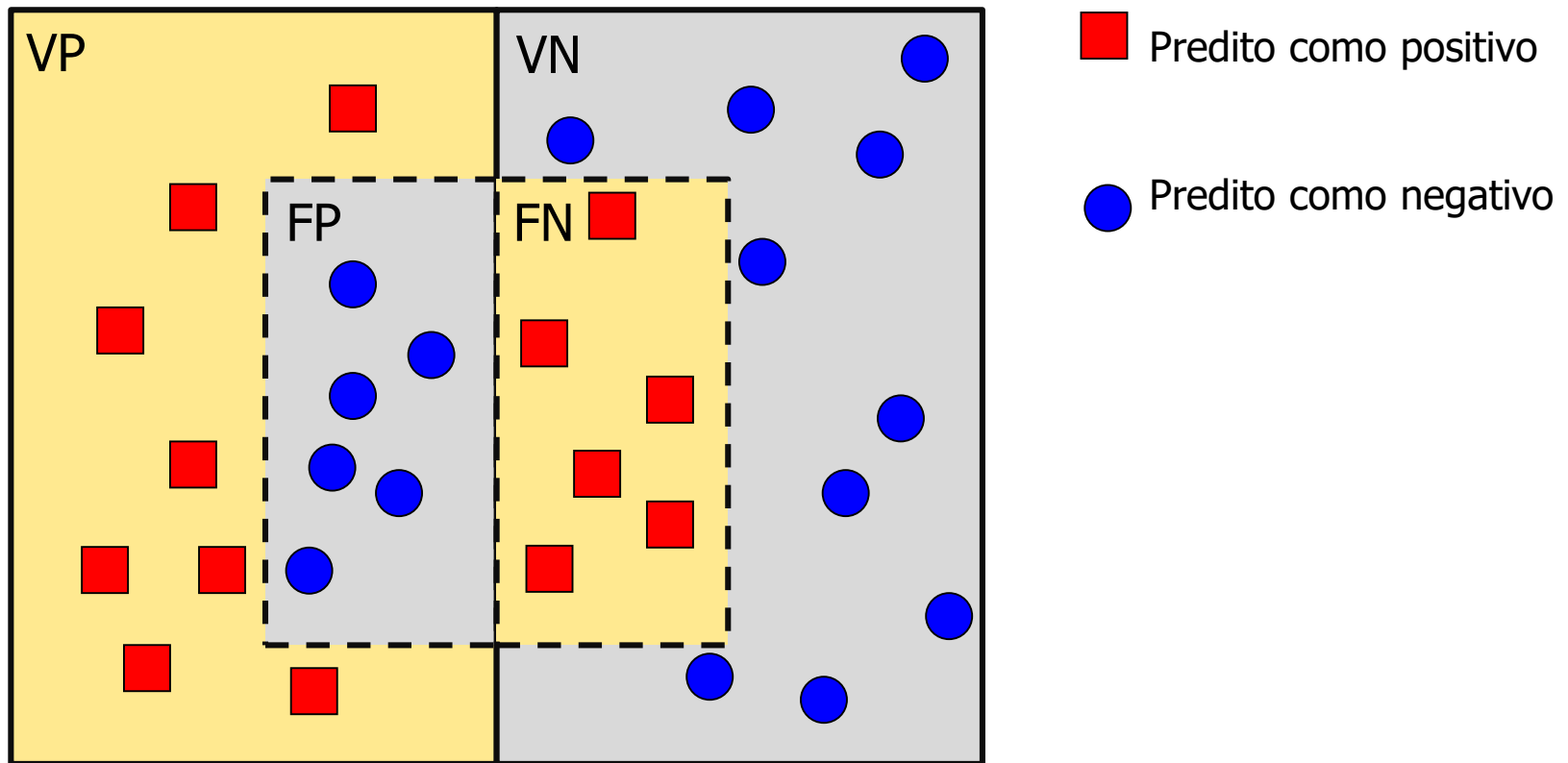




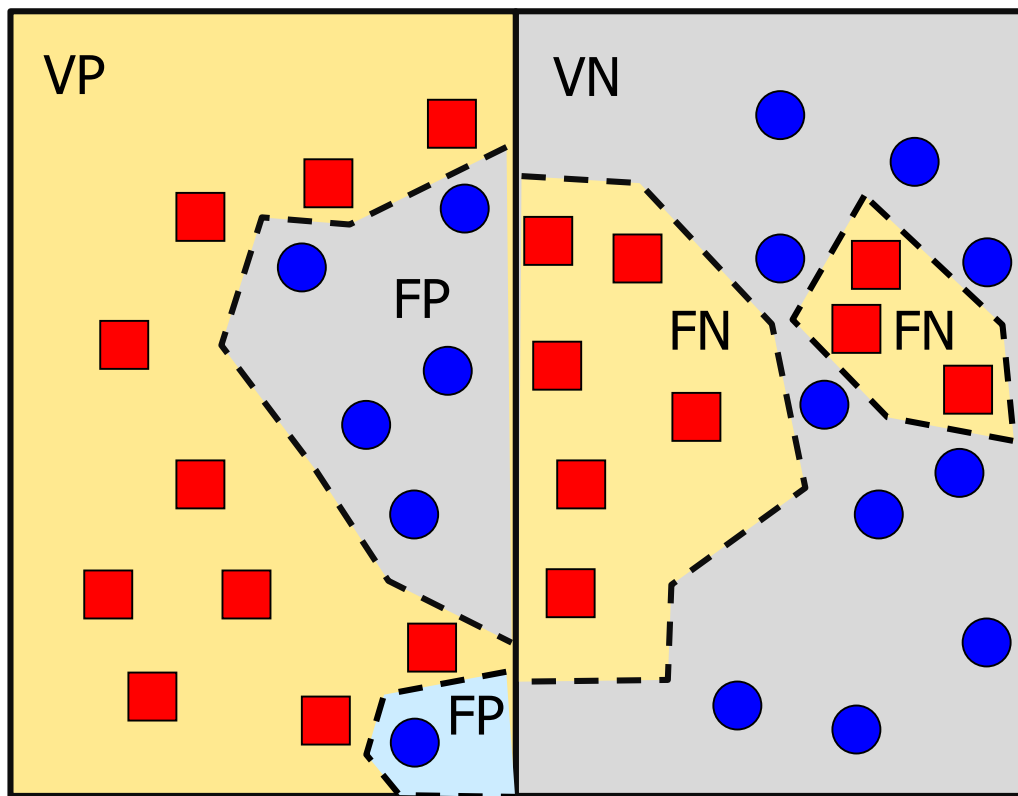
Classificação binária

- Classe de interesse é a classe positiva
- Dois tipos de erro:
 - Classificação de um exemplo N como P
 - Falso positivo (alarme falso)
 - Ex.: Diagnosticado como doente, mas está saudável
 - Classificação de um exemplo P como N
 - Falso negativo
 - Ex.: Diagnosticado como saudável, mas está doente

Classificação binária



Classificação binária



■ Predito como positivo

● Predito como negativo



Desempenho preditivo

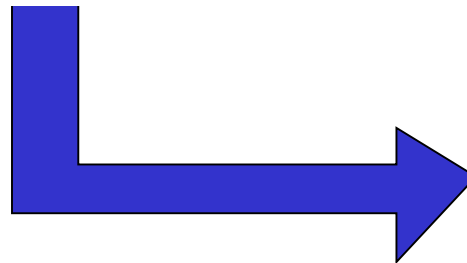
- Matriz de confusão (tabela de contingência) pode ser utilizada para distinguir os erros
 - Base de várias medidas
 - Pode ser utilizada com 2 ou mais classes

Classe verdadeira	Classe predita		
	1	2	3
1	25	0	5
2	10	40	0
3	0	0	20

Exemplo

- Matriz de confusão para 200 exemplos divididos em 2 classes

Classe verdadeira	Classe predita	
	p	n
P	70	30
N	40	60



Classe verdadeira	Classe predita	
	p	n
P	VP	FN
N	FP	VN

Medidas de avaliação

$$\text{Taxa de FP (TFP)} = \frac{FP}{FP + VN}$$

(Alarmes falsos)

Erro do tipo I

		Classe predita	
		p	n
Classe verdadeira	P	VP	FN
	N	FP	VN

$$\text{Taxa de FN (TFN)} = \frac{FN}{VP + FN}$$

Erro do tipo II

		Classe predita	
		p	n
Classe verdadeira	P	VP	FN
	N	FP	VN

Medidas de avaliação

$$\text{Taxa de FP (TFP)} = \frac{FP}{FP + VN}$$

(Alarmes falsos)

Custo

		Classe predita	
		p	n
Classe verdadeira	P	VP	FN
	N	FP	VN

$$\text{Taxa de VP (TVP)} = \frac{VP}{FN + VP}$$

Benefício

		Classe predita	
		p	n
Classe verdadeira	P	VP	FN
	N	FP	VN

Exemplo

$$\frac{VP}{VP + FN} \quad \frac{FP}{FP + VN}$$

■ Avaliação de 3 classificadores

Classe verdadeira	Classe predita	
	p	n
P	20	30
N	15	35

Classificador 1
TVP =
TFP =

Classe verdadeira	Classe predita	
	p	n
P	70	30
N	50	50

Classificador 2
TVP =
TFP =

Classe verdadeira	Classe predita	
	p	n
P	60	40
N	20	80

Classificador 3
TVP =
TFP =



Exemplo

$$\frac{VP}{VP + FN} \quad \frac{FP}{FP + VN}$$

■ Avaliação de 3 classificadores

Classe verdadeira	Classe predita	
	p	n
P	20	30
N	15	35

Classificador 1
TVP = 0.4
TFP = 0.3

Classe verdadeira	Classe predita	
	p	n
P	70	30
N	50	50

Classificador 2
TVP = 0.7
TFP = 0.5

Classe verdadeira	Classe predita	
	p	n
P	60	40
N	20	80

Classificador 3
TVP = 0.6
TFP = 0.2



Medidas de avaliação

$$\frac{FP}{FP + VN}$$

Taxa de falso positivo
(TFP) = 1-TVN

$$\frac{FN}{VP + FN}$$

Taxa de falso negativo
(TFN) = 1-TVP

$$\frac{VP}{VP + FP}$$

Valor predito positivo
(VPP), **precisão**

$$\frac{VN}{VN + FN}$$

Valor predito negativo
(VPN)

$$\frac{VP}{VP + FN}$$

Taxa de verdadeiro
positivo (TVP),
Sensibilidade ou
Revocação (Recall)

$$\frac{VN}{VN + FP}$$

Taxa de verdadeiro
negativo (TVN),
especificidade

$$\frac{VP + VN}{VP + VN + FP + FN}$$

Acurácia

$$\frac{2}{1 / prec. + 1 / revoc.}$$

Medida-F1



Medidas de avaliação

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

False positive rate
(FPR) = 1-TNR

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

False negative rate
(FNR) = 1-TPR

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

True positive rate
(TPR), also known as
recall or sensitivity

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

True negative rate
(TNR), also known as
specificity

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

Positive predictive
value (PPV), also
known as precision

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

Negative predictive
value (NPV)

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

Accuracy

$$\frac{2}{1/precision + 1/recall}$$

F1-measure

Matriz de confusão - 3 classes

		Classe real		
		Gato	Cachorro	Coelho
Classe Predita	Gato	5	2	0
	Cachorro	3	3	2
	Coelho	0	1	11

Matriz de confusão - Classe gato

		Classe real		
		Gato	Cachorro	Coelho
Classe Predita	Gato	5	2	0
	Cachorro	3	3	2
	Coelho	0	1	11

		Classe real	
		Gato	Não gato
Classe Predita	Gato	5 VP	2 FP
	Não gato	3 FN	17 VN

Matriz de confusão - Classe cachorro

		Classe real		
		Gato	Cachorro	Coelho
Classe Predita	Gato	5	2	0
	Cachorro	3	3	2
	Coelho	0	1	11

		Classe real	
		Cachorro	Não Cachorro
Classe Predita	Cachorro	3 VP	5 FP
	Não Cachorro	3 FN	16 VN

Matriz de confusão - Classe coelho

		Classe real		
		Gato	Cachorro	Coelho
Classe Predita	Gato	5	2	0
	Cachorro	3	3	2
	Coelho	0	1	11

		Classe real	
		Coelho	Não coelho
Classe Predita	Coelho	11 VP	1 FP
	Não coelho	2 FN	13 VN