

Estimativa de Erro de Classificação

- Principal objetivo de um modelo supervisionado é **prever com sucesso** o valor de saída para objetos ainda não vistos
 - Errar o mínimo possível
- Para **quantificar** o desempenho preditivo (estimado) do modelo criado, existem diversas **medidas** na literatura
 - Cada medida tem um viés... (Teorema do No Free Lunch)
 - Para problemas de **regressão**:
 - Erro quadrático médio (com ou sem raiz)
 - Erro absoluto médio
 - ...
 - Para problemas de **classificação**:
 - Acurácia/Erro
 - Matriz de Confusão
 - Curvas PR e ROC
 - Kappa
 - ...

177

177

Taxa de Classificação Incorreta

- A medida clássica para estimar a taxa de erro de um classificador é denominada de **taxa de classificação incorreta** (*misclassification rate*), ou simplesmente **erro de classificação**
 - Proporção dos objetos de teste que são classificados incorretamente pelo classificador
$$erro = \frac{\#erros}{N_{teste}}$$
- Usualmente é medida de forma indireta através do seu complemento, a **taxa de classificação correta**:
$$acuracia = \frac{\#acertos}{N_{teste}}$$
 - Acurácia
 - $acuracia = (1 - erro)$

178

178

Acurácia

- Do inglês, *Accuracy*
 - Dá **tratamento igual a todas as classes** do problema
 - Não é** uma medida **adequada** para medir problemas com **classes desbalanceadas**
 - A medida privilegia a classe majoritária
 - Na vasta maioria dos problemas desbalanceados, a classe interessante (prioritária) é a classe rara =(
 - Ex: considere um problema de 2 classes
 - Classe 1 = 9990 objetos
 - Classe 2 = 10 objetos
 - Se modelo prevê apenas classe 1, acurácia será de $9990/10000 = 99.9\%$
 - Note que tal modelo não é sequer inteligente!!!

179

179

Tipos de Erros

- Em classificação binária, é comum nomear os objetos da classe de maior interesse de **positivos (+)**
 - Normalmente a classe rara ou minoritária
 - Demais objetos são nomeados **negativos (-)**
- Em alguns casos, os erros têm igual importância
- Em muitos casos, no entanto, **erros têm prioridades distintas (custos!)** considerando as possíveis consequências
 - Ex: diagnóstico negativo para indivíduo doente

180

180

Tipos de Erros

- Existem dois tipos de erro em classificação binária:
 - Classificar objeto negativo como positivo
 - Falso Positivo** (FP), Alarme Falso
 - Ex: paciente diagnosticado como doente, embora esteja saudável
 - Classificar objeto positivo como negativo
 - Falso Negativo** (FN)
 - Ex: paciente diagnosticado como saudável, mas está doente

181

181

Matriz de Confusão

- Também chamada de **Tabela de Contingência**
 - Permite a extração de **diversas medidas** de desempenho preditivo
 - Pode ser utilizada para distinguir os tipos de erros
 - Pode ser utilizada para problemas binários ou multi-classe

Classe Verdadeira	Classe Prevista		
	A	B	C
A	25	0	5
B	10	40	0
C	0	0	20

182

182

Matriz de Confusão

- Também chamada de **Tabela de Contingência**
 - Permite a extração de **diversas medidas** de desempenho preditivo
 - Pode ser utilizada para distinguir os tipos de erros
 - Pode ser utilizada para problemas binários ou multi-classe

Classe Verdadeira	Classe Prevista		
	A	B	C
A	25	0	5
B	10	40	0
C	0	0	20

Diagonal principal: acertos!

183

183

Matriz de Confusão

- Também chamada de **Tabela de Contingência**
 - Permite a extração de **diversas medidas** de desempenho preditivo
 - Pode ser utilizada para distinguir os tipos de erros
 - Pode ser utilizada para problemas binários ou multi-classe

Classe Verdadeira	Classe Prevista		
	A	B	C
A	25	0	5
B	10	40	0
C	0	0	20

Valores fora da diagonal principal: erros!

184

184

Matriz de Confusão

- Também chamada de **Tabela de Contingência**
 - Permite a extração de **diversas medidas** de desempenho preditivo
 - Pode ser utilizada para distinguir os tipos de erros
 - Pode ser utilizada para problemas binários ou multi-classe

Classe Verdadeira	Classe Prevista		
	A	B	C
A	25	0	5
B	10	40	0
C	0	0	20

$$\text{Acurácia: } \frac{25 + 40 + 20}{25 + 40 + 20 + 10 + 5} = \frac{85}{100} = 0.85 \text{ ou } 85\%$$

185

185

Matriz de Confusão Binária

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	70	30
Negativa	40	60

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

186

186

Matriz de Confusão Binária

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	70	30
Negativa	40	60

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

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

187

187

Matriz de Confusão Binária

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	70	30
Negativa	40	60

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

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

$$\text{Erro: } \frac{FP + FN}{VP + VN + FP + FN} = (1 - \text{acurácia})$$

188

188

189

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

189

191

- | Classificador 1 | Classificador 2 | Classificador 3 |
|-----------------|-----------------|-----------------|
| Acurácia = | Acurácia = | Acurácia = |
| Erro = | Erro = | Erro = |
| TFN = | TFN = | TFN = |
| TFP = | TFP = | TFP = |

193

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

193

190

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

190

192

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

192

194

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

194

Precision x Recall

Precisão: (Precision) $\frac{VP}{FP + VP}$

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

O que acontece se um modelo classificar todos exemplos como sendo positivos?

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

Revocação: (Recall) $\frac{VP}{FN + VP}$

195

195

Precision x Recall

Precisão: (Precision) $\frac{VP}{FP + VP}$

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

O que acontece se um modelo classificar todos exemplos como sendo positivos?

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

Revocação: (Recall) $\frac{VP}{FN + VP}$ máximo

196

196

Precision x Recall

Precisão: (Precision) $\frac{VP}{FP + VP}$

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

O que acontece se um modelo classificar todos exemplos como sendo positivos?

Classe Verdadeira	Classe Prevista	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

Revocação: (Recall) $\frac{VP}{FN + VP}$ máximo

197

197

F-Measure

- Média harmônica de *precision* e *recall*
 - Também conhecida como F_1 score ou F-score
 - F_β , onde β é um fator positivo de quão mais importante é *recall* em relação a *precision*

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$

$$F_\beta = (1 + \beta^2) \times \frac{\text{precision} \times \text{recall}}{(\beta^2 * \text{precision}) + \text{recall}}$$

198

198

Resumo das Medidas Apresentadas

$$\begin{aligned} \text{Acurácia} &= \frac{VP + VN}{VP + FP + VN + FN} & \text{Erro} &= \frac{FP + FN}{VP + FP + VN + FN} \\ \text{Especificidade} &= \frac{VN}{FP + VN} & \text{TFP} &= \frac{FP}{FP + VN} \\ \text{(TVN, } 1 - \text{TFP)} & & \text{(Erro tipo I, Custo)} & \\ \text{Recall} &= \frac{VP}{FN + VP} & \text{TFN} &= \frac{FN}{FN + VP} \\ \text{(TVP, Sensibilidade, Benefício)} & & \text{(Erro tipo II, } 1 - \text{Recall)} & \\ \text{Precision} &= \frac{VP}{VP + FP} & F_1 &= \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} & F_\beta &= (1 + \beta^2) \times \frac{\text{precision} \times \text{recall}}{(\beta^2 * \text{precision}) + \text{recall}} \end{aligned}$$

199

199

Estimativas de Erro de Regressão

- Função de perda
- Erro médio absoluto
- Erro médio quadrático
- Raiz do erro médio quadrático
- Erro absoluto relativo
- Erro quadrático relativo
- Raiz do erro quadrático relativo

200

Função de perda

- Erro absoluto:

$$|y_i - y_i'|$$

- Erro quadrático:

$$(y_i - y_i')^2$$

Medem o erro entre o valor real y_i e o valor predito y_i'

Exemplos:

y real	y predito	erro absoluto	erro quadrático
2,5	3,0	0,50	0,25
4,0	2,0	2,00	4,00
1,0	1,5	0,50	0,25
3,7	4,5	0,80	0,64
15,2	17,0	1,80	3,24
3,6	5,0	1,40	1,96

Erro absoluto

- Erro médio absoluto

$$\frac{\sum_{i=1}^d |y_i - y_i'|}{d}$$

- Erro absoluto relativo

$$\frac{\sum_{i=1}^d |y_i - y_i'|}{\sum_{i=1}^d |y_i - \bar{y}|}$$

y real	y predito	Média	Distância à média
2,5	3,0	5,0	3,40
4,0	2,0		
1,0	1,5		
3,7	4,5		
15,2	17,0		
3,6	5,0		

Erro médio absoluto	1,167
Erro absoluto relativo	0,343

201

Erro quadrático

- Erro médio quadrático

$$\frac{\sum_{i=1}^d (y_i - y_i')^2}{d}$$

- Raiz do erro médio quadrático

$$\sqrt{\frac{\sum_{i=1}^d (y_i - y_i')^2}{d}}$$

y real	y predito
2,5	3,0
4,0	2,0
1,0	1,5
3,7	4,5
15,2	17,0
3,6	5,0

Erro médio quadrático	1,723
Raiz do erro médio quadrático	1,313

202

Erro quadrático relativo

- Erro quadrático relativo

$$\frac{\sum_{i=1}^d (y_i - y_i')^2}{\sum_{i=1}^d (y_i - \bar{y})^2}$$

- Raiz do erro quadrático relativo

$$\sqrt{\frac{\sum_{i=1}^d (y_i - y_i')^2}{\sum_{i=1}^d (y_i - \bar{y})^2}}$$

y real	y predito
2,5	3,0
4,0	2,0
1,0	1,5
3,7	4,5
15,2	17,0
3,6	5,0

Erro quadrático relativo	0,079
Raiz do erro quadrático relativo	0,281

203

Sobre erros relativos

- As métricas de erros relativos, Erro absoluto relativo, e Raiz do erro quadrático relativo, tem o objetivo de dar a percepção de quão bom é o regressor em relação a um regressor supersimples, que atribui o valor da média de y reais para qualquer objeto.
- Com isso, quanto menor, melhor. Se é maior que 1, significa que o regressor é PIOR que o regressor mais simples.

Fonte:

<https://www.gepsoft.com/gxpt4kb/Chapter10/Section1/SS07.htm>

Exemplos:

y real	y predito	erro absoluto	erro quadrático
2,5	3,0	0,50	0,25
4,0	2,0	2,00	4,00
1,0	1,5	0,50	0,25
3,7	4,5	0,80	0,64
15,2	17,0	1,80	3,24
3,6	5,0	1,40	1,96
Média	5,00		
Desvio padrão	4,67	5,29	
Erro absoluto	20,40		
Erro quadrático	130,94		

Coefficiente de correlação	0,97797	
Erro médio absoluto	1,167	
Erro absoluto relativo	0,343	
Erro médio quadrático		1,723
Raiz do erro médio quadrático		1,313
Erro quadrático relativo		0,079
Raiz do erro quadrático relativo		0,281

205

206

Exemplos:

	y real	y predito	erro absoluto	erro quadrático
	1,0	1,2	0,20	0,04
	1,0	1,0	0,00	0,00
	1,0	0,8	0,20	0,04
	1,0	1,0	0,00	0,00
	1,0	1,1	0,10	0,01
	1,0	0,9	0,10	0,01
Média	1,00	1,00		
Desvio padrão	0,00	0,13		
Erro absoluto	0,00			
Erro quadrático	0,00			

	Coefficiente de correlação	#DIV/0!
Erro médio absoluto	0,100	
Erro absoluto relativo	#DIV/0!	
Erro médio quadrático		0,017
Raiz do erro médio quadrático		0,129
Erro quadrático relativo		#DIV/0!
Raiz do erro quadrático relativo		#DIV/0!

207

Exemplos:

	y real	y predito	erro absoluto	erro quadrático
	1,2	1,2	0,00	0,00
	1,0	1,0	0,00	0,00
	0,8	0,8	0,00	0,00
	1,0	1,0	0,00	0,00
	1,1	1,1	0,00	0,00
	0,9	0,9	0,00	0,00
Média	1,00	1,00		
Desvio padrão	0,13	0,13		
Erro absoluto	0,60			
Erro quadrático	0,10			

	Coefficiente de correlação	1,00000
Erro médio absoluto	0,000	
Erro absoluto relativo	0,000	
Erro médio quadrático		0,000
Raiz do erro médio quadrático		0,000
Erro quadrático relativo		0,000
Raiz do erro quadrático relativo		0,000

208

Exemplos:

	y real	y predito	erro absoluto	erro quadrático
	1,2	1,0	0,20	0,04
	1,0	1,0	0,00	0,00
	0,8	1,0	0,20	0,04
	1,0	1,0	0,00	0,00
	1,1	1,0	0,10	0,01
	0,9	1,0	0,10	0,01
Média	1,00	1,00		
Desvio padrão	0,13	0,00		
Erro absoluto	0,60			
Erro quadrático	0,10			

	Coefficiente de correlação	#DIV/0!
Erro médio absoluto	0,100	
Erro absoluto relativo	1,000	
Erro médio quadrático		0,017
Raiz do erro médio quadrático		0,129
Erro quadrático relativo		1,000
Raiz do erro quadrático relativo		1,000

209

Exemplos:

	y real	y predito	erro absoluto	erro quadrático
	1,2	0,9	0,30	0,09
	1,0	1,1	0,10	0,01
	0,8	1,0	0,20	0,04
	1,0	0,8	0,20	0,04
	1,1	1,0	0,10	0,01
	0,9	1,2	0,30	0,09
Média	1,00	1,00		
Desvio padrão	0,13	0,13		
Erro absoluto	0,60			
Erro quadrático	0,10			

	Coefficiente de correlação	-0,40000
Erro médio absoluto	0,200	
Erro absoluto relativo	2,000	
Erro médio quadrático		0,047
Raiz do erro médio quadrático		0,216
Erro quadrático relativo		2,800
Raiz do erro quadrático relativo		1,673

210

Gráficos ROC

- Do inglês, *Receiver Operating Characteristics*
- Medida de desempenho originária da área de processamento de sinais
 - Muito utilizada na área médica (e na biologia em geral)
 - Mostra relação entre custo (TFP, Erro do Tipo I) e benefício (TVP, *Recall*)
 - Lembre-se que:
 - TFP é a taxa de alarmes falsos (erros na classe negativa, Erro do Tipo I)
 - TVP é a taxa de acertos na classe positiva (1 - Erro do Tipo II)

Erro do Tipo I: (TFP) $\frac{FP}{FP + VN}$ Sensibilidade: (TVP) $\frac{VP}{FN + VP} = (1 - TFP)$
(Taxa de Alarmes Falsos) (Custo) (Recall, Revocação, Benefício)

211

Gráficos ROC

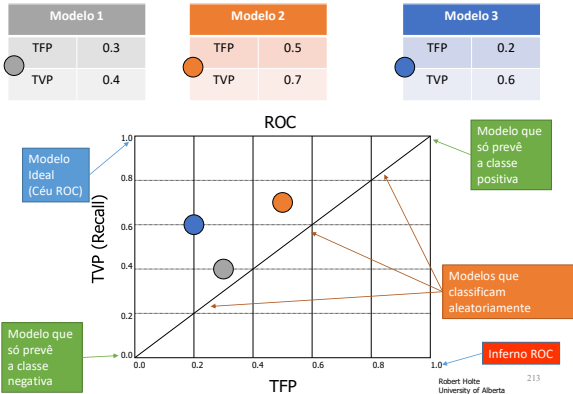
- Medida de desempenho originária da área de processamento de sinais
 - Mostra relação entre custo (TFP, Erro do Tipo I) e benefício (TVP, *Recall*)

Classe Verdadeira	Classe Predita	
	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

Erro do Tipo I: (TFP) $\frac{FP}{FP + VN}$ Sensibilidade: (TVP) $\frac{VP}{FN + VP} = (1 - TFP)$
(Taxa de Alarmes Falsos) (Custo) (Recall, Revocação, Benefício)

212

Imagine a existência de três modelos:



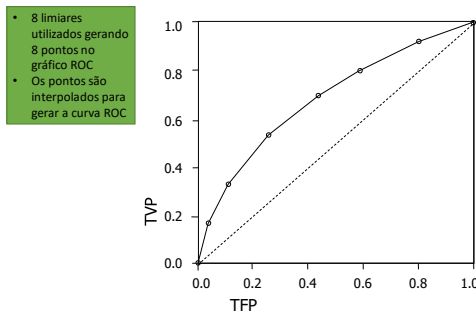
213

Gráficos ROC

- Resumindo:
 - Classificador **ideal** mais a **noroeste**
 - Classificadores próximos ao **canto inferior esquerdo** são **conservadores**
 - Apenas detectam a classe positiva com forte evidência
 - Portanto, cometem poucos FPs
 - Classificadores próximos ao **canto superior direito** são **liberais**
 - Detectam a classe positiva com pouca evidência
 - Correm o risco de alta taxa de alarme falso
 - Classificadores ao **redor da linha central** tem comportamento **similar** ao esperado de **classificação aleatória**

214

Curvas ROC



217

Curvas ROC

Objeto	Classe Real	Escore (Classe +)
$x^{(6)}$	+	0.9
$x^{(3)}$	+	0.8
$x^{(2)}$	-	0.7
$x^{(9)}$	+	0.6
$x^{(5)}$	+	0.6
$x^{(1)}$	-	0.5
$x^{(7)}$	-	0.3
$x^{(8)}$	-	0.2
$x^{(4)}$	-	0.2
$x^{(10)}$	-	0.1

- Ordenar objetos em ordem decrescente de escore para a classe positiva (+)
- Para cada limiar de decisão θ :
 - Classificar todos os objetos
 - Calcular VP, VN, FP, FN
 - Calcular TVP e TFP e plotar ponto no gráfico ROC

$$Classe = \begin{cases} \text{escore} \geq \theta: + \\ \text{escore} < \theta: - \end{cases}$$

218

Curvas ROC

- Algoritmos que geram valores discretos:
 - Podem ser **modificados** para gerar escores
 - Para ADs, pode se utilizar a **fração dos objetos** de treinamento **positivos** do **nó folha** correspondente como **escore**
 - Para k -NN, pode se utilizar a **fração dos k vizinhos** mais próximos que pertencem à classe **positiva** como **escore**
 - Para SVMs, pode se utilizar a **distância normalizada** do objeto ao hiperplano separador como **escore**
 - ...
 - Podem ser **combinados em comitês**
 - Algoritmo é executado sobre **amostragens** do conjunto de **treinamento**, gerando **múltiplos modelos**
 - Cada modelo prevê uma das duas classes (+ ou -)
 - O escore será a **fração dos modelos** que **previram a classe positiva**

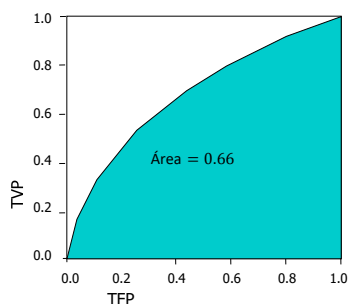
220

Área sob a Curva ROC (AUC)

- Fornecer uma estimativa do desempenho de classificadores
- Valor contínuo no intervalo $[0, 1]$
 - Quanto maior melhor
 - Adição de áreas de sucessivos trapézios
- É possível provar que a AUC equivale à **probabilidade** do modelo **atribuir um escore** $P(+ | x)$ **maior** a um **objeto positivo** escolhido aleatoriamente **do que a um objeto negativo** escolhido aleatoriamente

221

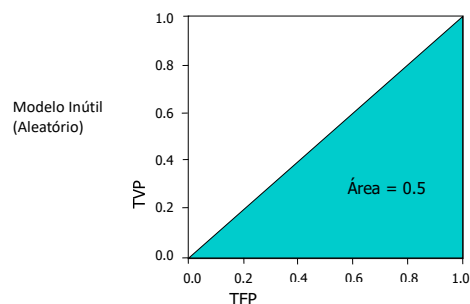
Área sob a Curva ROC (AUC)



222

222

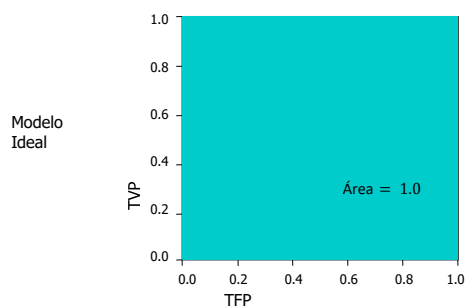
Área Sob Curvas ROC



223

223

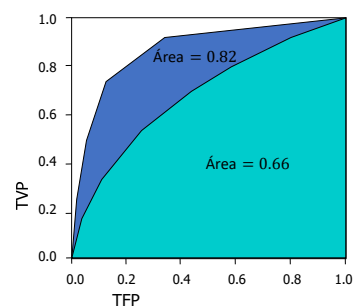
Área sob a Curva ROC (AUC)



224

224

Área sob a Curva ROC (AUC)



225

225

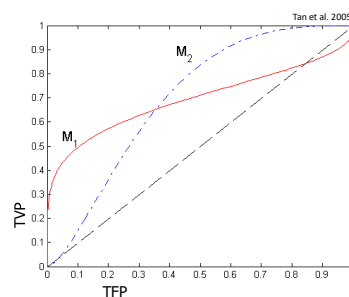
Área sob a Curva ROC (AUC)

- **Nota 1:** um modelo com maior AUC pode apresentar AUC pior em trechos da curva...
 - AUC **não deve ser vista** como **critério absoluto**
 - Deve ser vista como medida de desempenho auxiliar as demais, com suas vantagens e desvantagens

226

226

Área sob a Curva ROC (AUC)



- Modelos similares em desempenho preditivo
 - M_1 é melhor para baixos valores de TFP
 - M_2 é melhor para altos valores de TFP

227

Área sob a Curva ROC (AUC)

- **Nota 2:**

- Para **maior confiabilidade da análise**, calcula-se a AUC utilizando-se algum dos procedimentos de avaliação de desempenho vistos anteriormente (e.g., *cross-validation*) para gerar **múltiplas curvas ROC**
 - AUC mais confiável é tomada a partir de algum tipo de **média** das **AUCs** previamente calculadas, **ou** a partir de uma **curva média**
 - A **variância** das curvas também é um fator a ser analisado

228