

Aprendizado de Máquina

Aula 3: Medidas de desempenho

André C. P. L. F de Carvalho
ICMC/USP

andre@icmc.usp.br



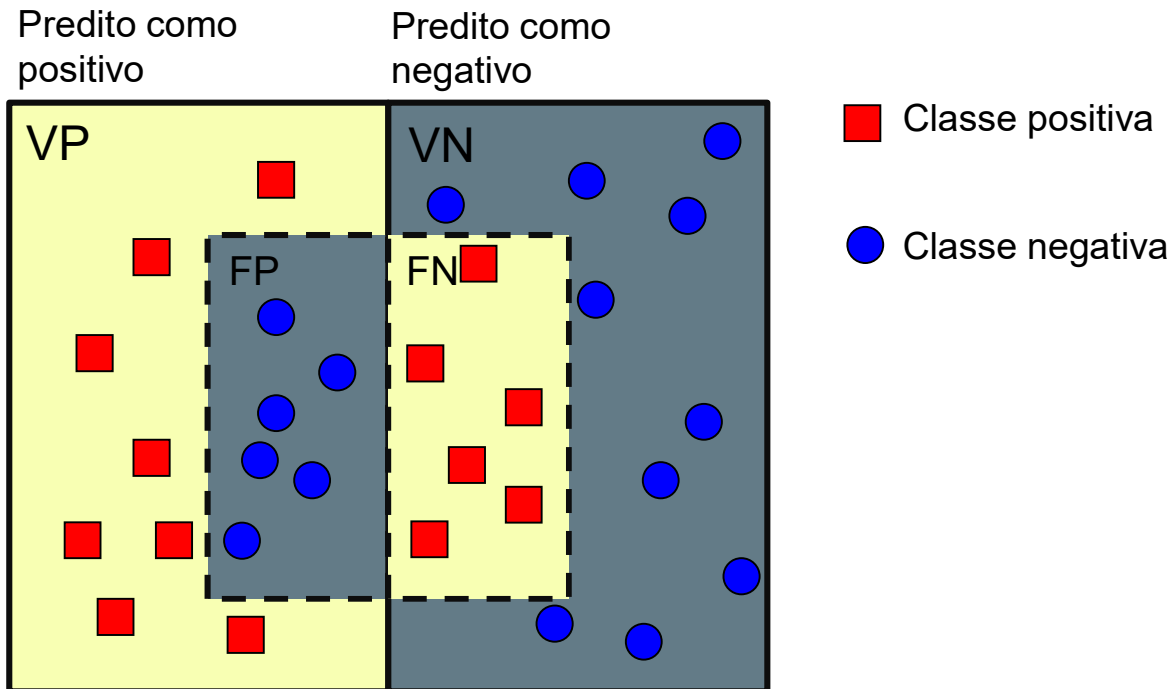
Tópicos

- Tipos de erro
- Avaliação do desempenho preditivo
- Curvas ROC
- Testes de hipóteses

Classificação binária

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

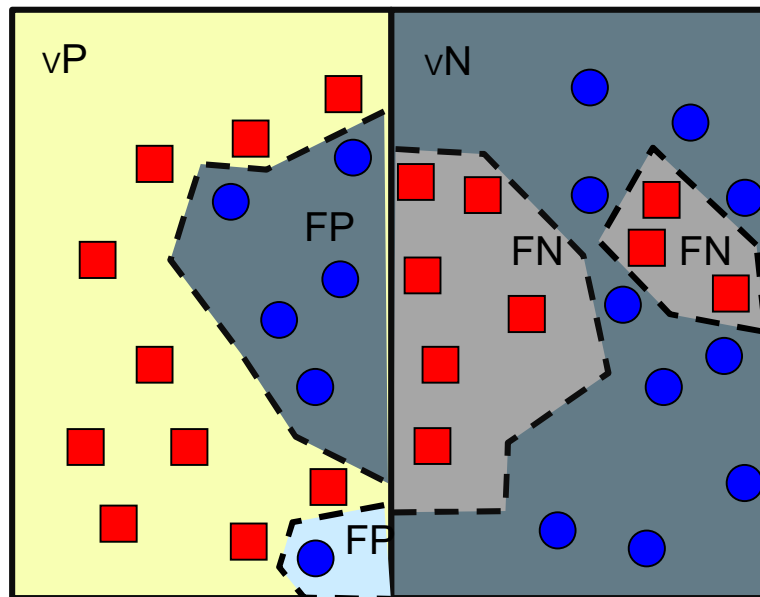
Classificação binária



Classificação binária

Predito como
positivo

Predito como
negativo



■ Classe positiva

● Classe negativa

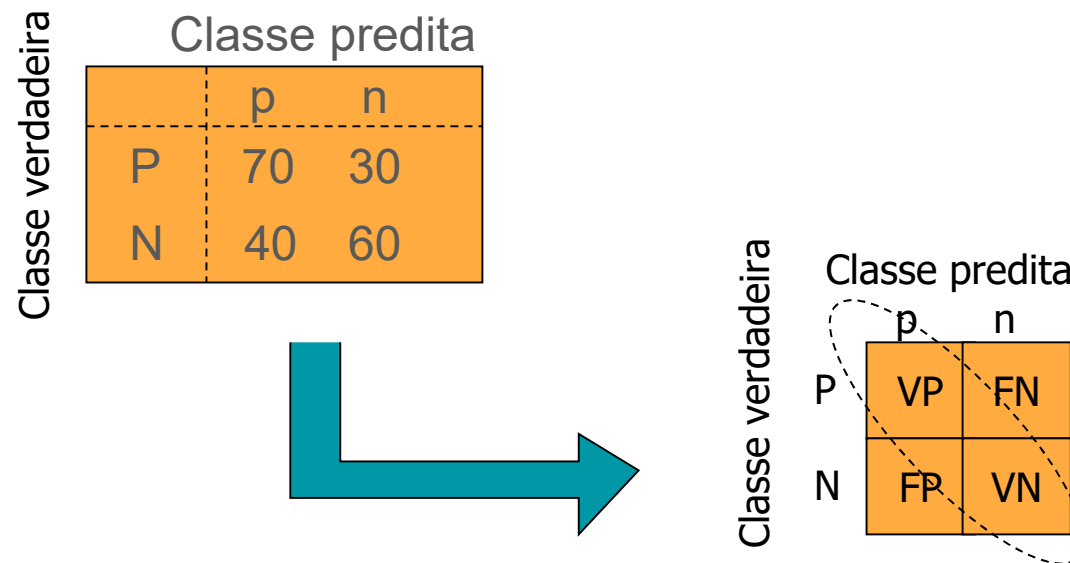
Desempenho preditivo

- Uma matriz de confusão (tabela de contingência) pode ser utilizada para distinguir os erros
 - Base de várias medidas de desempenho preditivo
 - 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



Medidas de avaliação

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

(Alarmes falsos)

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

Erro do tipo I

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

Erro do tipo II

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

Medidas de avaliação

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

Custo

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

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

Benefício

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

Exemplo

- Avaliação de 3 classificadores

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

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

- Avaliação de 3 classificadores

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

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

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

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

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

Taxa de verdadeiro negativo (TVN),
especificidade

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

Valor predito positivo (VPP), precisão

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

Valor predito negativo (VPN)

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

Acurácia

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

Medida-F1

Acurácia

- Uma das mais usadas
- Taxa de objetos corretamente classificados
 - Trata as classes igualmente
 - Não é adequada para dados desbalanceados
 - Pode induzir modelo com baixa taxa de acerto para classe minoritária
 - Geralmente mais interessante que a majoritária
 - Acurácia balanceada

Revocação X Precisão

- Revocação (*recall*)

- Porcentagem de exemplos positivos classificados como positivos

- Nenhum exemplo positivo é deixado de fora

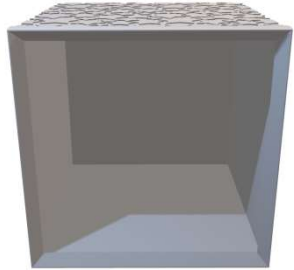
- Todos são lembrados

- Precisão

- Porcentagem de exemplos classificados como positivos que são realmente positivos

- Nenhum exemplo negativo é incluído

- Não tem “intrusos”


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

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

Sensibilidade X Especificidade

- Sensibilidade

- Porcentagem de exemplos positivos classificados como positivos

- Igual a revocação

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

- Especificidade

- Porcentagem de exemplos negativos classificados como negativos

- Nenhum exemplo negativo é deixado de fora

- Todos são lembrados

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

Medida-F

- Média harmônica ponderada da precisão e da revocação

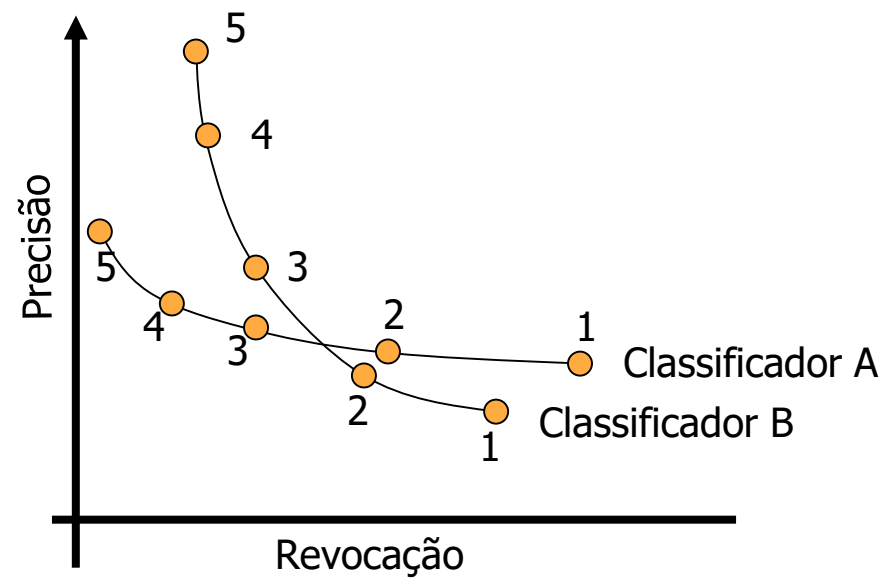
$$\frac{(1 + \alpha) \times (prec \times rev)}{\alpha \times prec + rev}$$

- Medida-F1

- Precisão e revocação têm o mesmo peso

$$\frac{2 \times (prec \times rev)}{prec + rev} = \frac{2}{1/prec + 1/rev}$$

Observação



Outras medidas

- Média geométrica de taxas positivas
 - G-mean $\sqrt{\text{precisão} \times \text{revocação}}$
- Acurácia balanceada
- Kappa
- ...

Gráficos ROC

- Do inglês, *Receiver operating characteristics*
- Medida de desempenho originária da área de processamento de sinais
 - Muito utilizada nas áreas médica e biológica
 - Mostra relação entre custo (TFP) e benefício (TVP)

Exemplo

- Colocar no gráfico ROC os 3 classificadores do exemplo anterior

Classificador 1
TFP = 0.3
TVP = 0.4



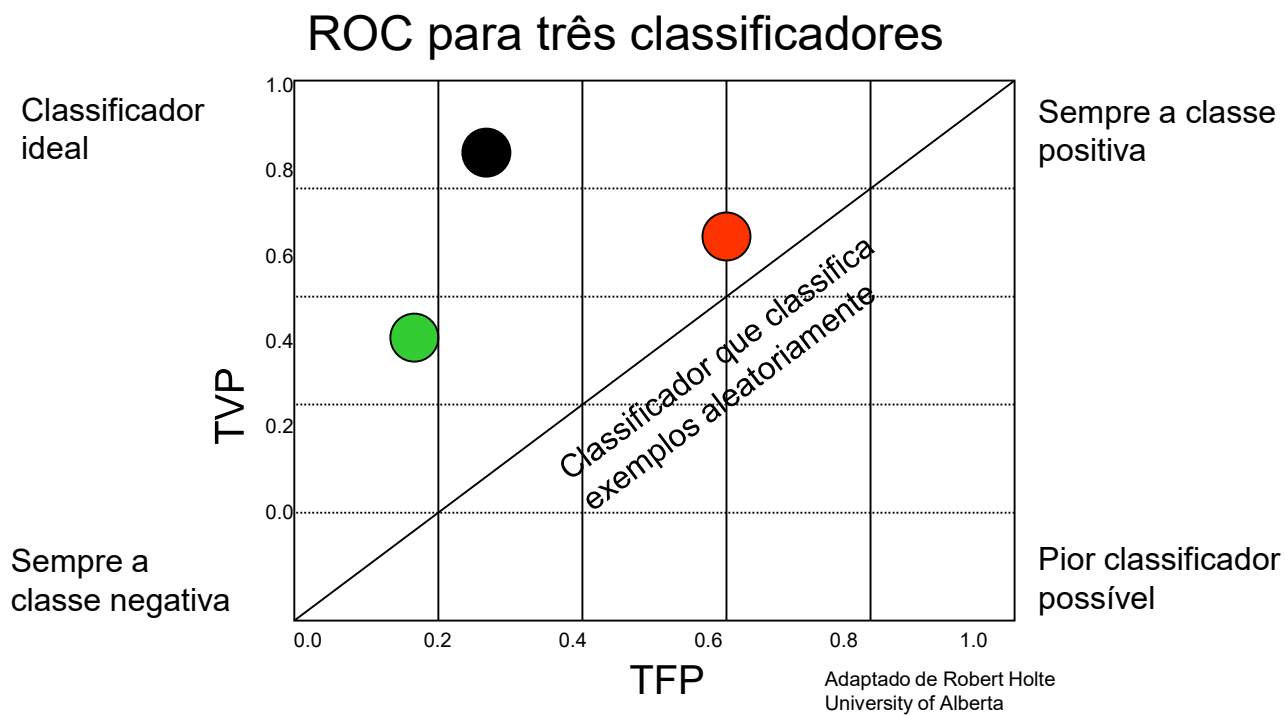
Classificador2
TFP = 0.5
TVP = 0.7



Classificador 3
TFP = 0.2
TVP = 0.6



Gráficos ROC



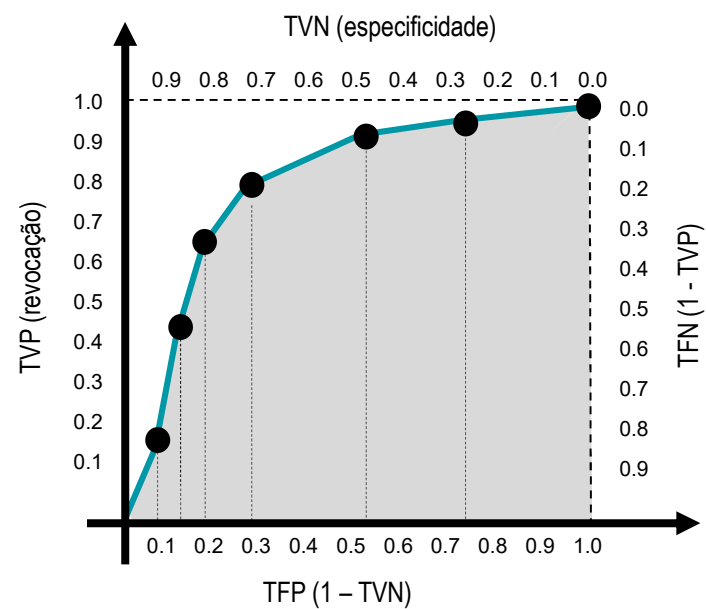
Gráficos ROC

- Classificadores discretos produzem um simples ponto no gráfico ROC
 - Árvores de decisão e conjuntos de regras
- Outros classificadores produzem uma probabilidade ou score
 - Redes neurais e naive bayes
- Curvas ROC permitem uma melhor comparação de classificadores
 - São insensíveis a mudanças na distribuição das classes

Curvas ROC

- Mostram ROC para diferentes variações
- Classificadores que geram valores contínuos (*threshold*, probabilidade)
 - Diferentes valores de *threshold* podem ser utilizados para gerar vários pontos
 - Ligação dos pontos gera uma curva ROC
- Classificadores discretos
 - Convertidos internamente ou usados em comitês

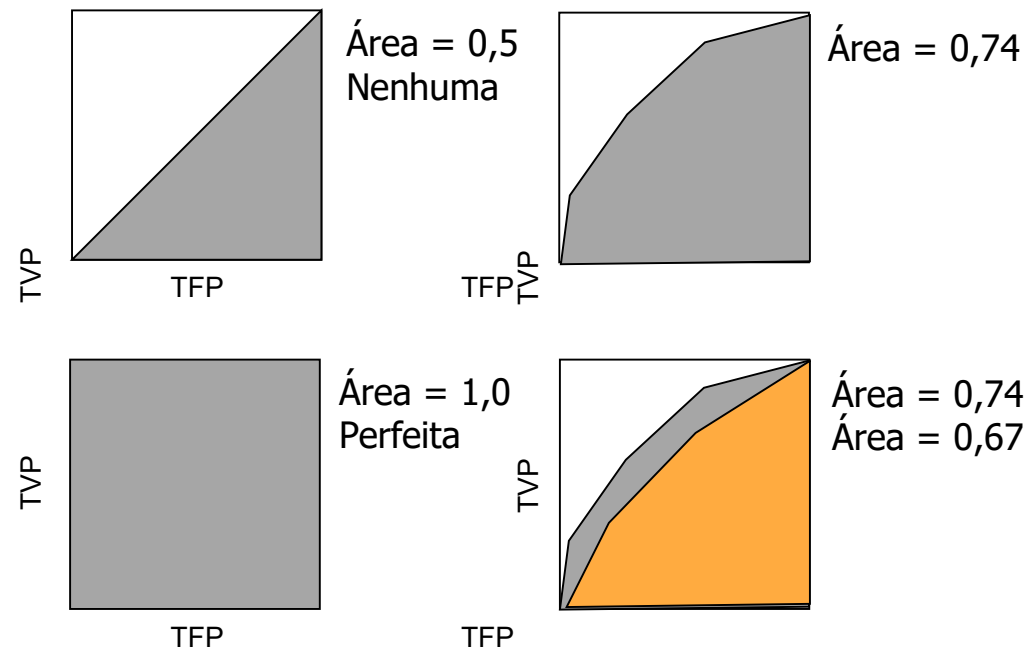
Área Sob a Curva ROC



Área sob a curva ROC (AUC)

- Fornece uma estimativa do desempenho de classificadores
- Gera um valor contínuo no intervalo $[0, 1]$
 - Quanto maior melhor
 - Adição de áreas de sucessivos trapezóides
- Um classificador com maior AUC pode apresentar AUC pior em trechos da curva
- É mais confiável utilizar médias de AUCs

Área Sob Curvas ROC



Teste de hipótese

- Permite afirmar que uma técnica é melhor que outra com X% de confiança
- Podem assumir que os dados seguem uma dada distribuição de probabilidade
 - Paramétricos
 - Não paramétricos
- Número de técnicas comparadas
 - Duas
 - Mais que duas

Teste de hipótese

- Testes usados atualmente são baseados na verificação da hipótese nula
 - Várias deficiências para uso em AM
 - Não geram probabilidades de ocorrência da hipótese nula e da hipótese alternativa
 - E de uma técnica ser melhor que outra
- Alternativa proposta em 2016
 - Teste Bayesiano Hierárquico

Teste bayesiano hierárquico

	A	B			C		
A	----	82%	16%	2%	91%	5%	4%
B			---		75%	10%	15%

Técnica da linha é X% melhor, Y% igual e Z% pior que a técnica da coluna

Probabilidade de ser: Melhor Igual Pior

Considerações Finais

- Desempenho preditivo
- Medidas
- Gráficos e curvas ROC
- Testes de hipóteses

Final da Apresentação

Copyright © 2020. Todos os direitos reservados ao CeMEAI-USP. Proibida a cópia e reprodução sem autorização

