

Aprendizado de Máquina

Aula 6: Análise de desempenho

André C. P. L. F de Carvalho
ICMC/USP

andre@icmc.usp.br



Tópicos a serem abordados

- Avaliação de desempenho de algoritmos/modelos
- Desempenho preditivo
 - Regressão
 - Classificação
- Medidas de desempenho preditivo
 - Matriz de confusão
 - Duas classes
 - Mais que duas classes

Desempenho preditivo

- Capacidade preditiva
 - Tarefa de classificação
 - Tarefa de regressão
 - Algoritmo e/ou modelo
- Custo
 - Tempo de processamento
 - Espaço de memória
- Interpretabilidade

Medidas de desempenho preditivo em regressão

- Soma dos quadrados dos erros (SSE)

$$SSE = \sum_{i=1}^n (y^i - f(x^i))^2 \text{ ou } SSE = \frac{1}{2} \sum_{i=1}^n (y^i - f(x^i))^2$$

↓ Menor, melhor

- Erro quadrático médio (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y^i - f(x^i))^2 \text{ ou } MSE = \frac{1}{2n} \sum_{i=1}^n (y^i - f(x^i))^2$$

↓ Menor, melhor

- Ao elevar o erro ao quadrado, interpretação do erro se torna mais difícil

Medidas de desempenho preditivo em regressão

- Raiz do erro quadrático médio (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y^i - f(x^i))^2}$$



Menor, melhor

- Por ter a mesma unidade de medida que o valor a ser predito, y , é mais fácil de interpretar que MSE

- Erro absoluto médio (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y^i - f(x^i)|$$



Menor, melhor

- Também tem a mesma unidade de medida que y

Medidas de desempenho preditivo em regressão

- Coeficiente de determinação (R^2)

$$R^2 = 1 - \frac{MSE}{Var(y)}$$

↑ Maior, melhor

- Versão padronizada do MSE, mas fácil de interpretar por ter mesma unidade de valor que y

- Erro percentual absoluto médio (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y^i - f(x^i)|}{y^i} \quad \text{ou} \quad MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y^i - f(x^i)|}{y^i}$$

↓ Menor, melhor

- Mede a acurácia de um regressor em %

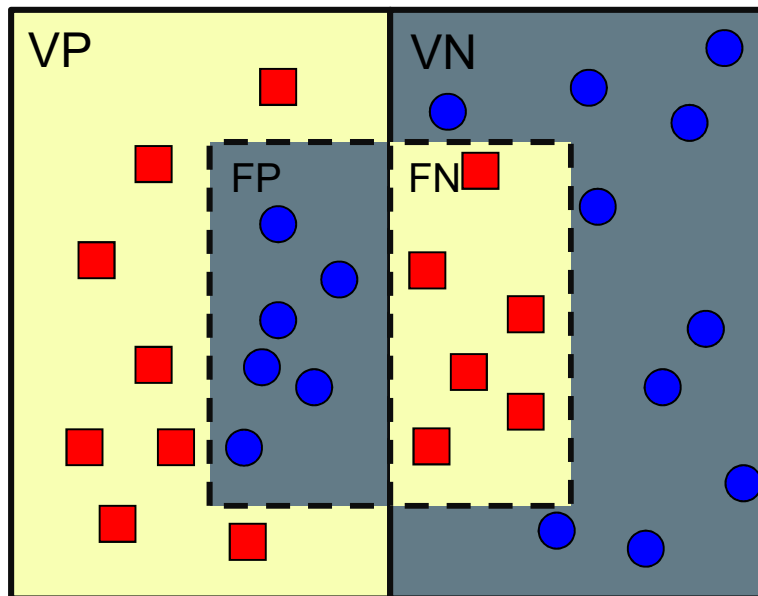
Classificação binária

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

Classificação binária

Predito como
positivo

Predito como
negativo



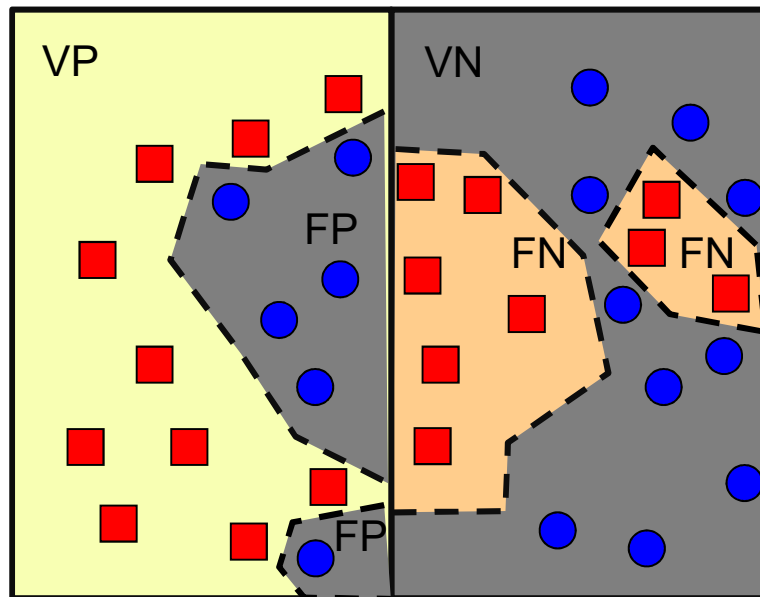
■ Classe positiva

● Classe negativa

Classificação binária

Predito como
positivo

Predito como
negativo



■ Classe positiva

● Classe negativa

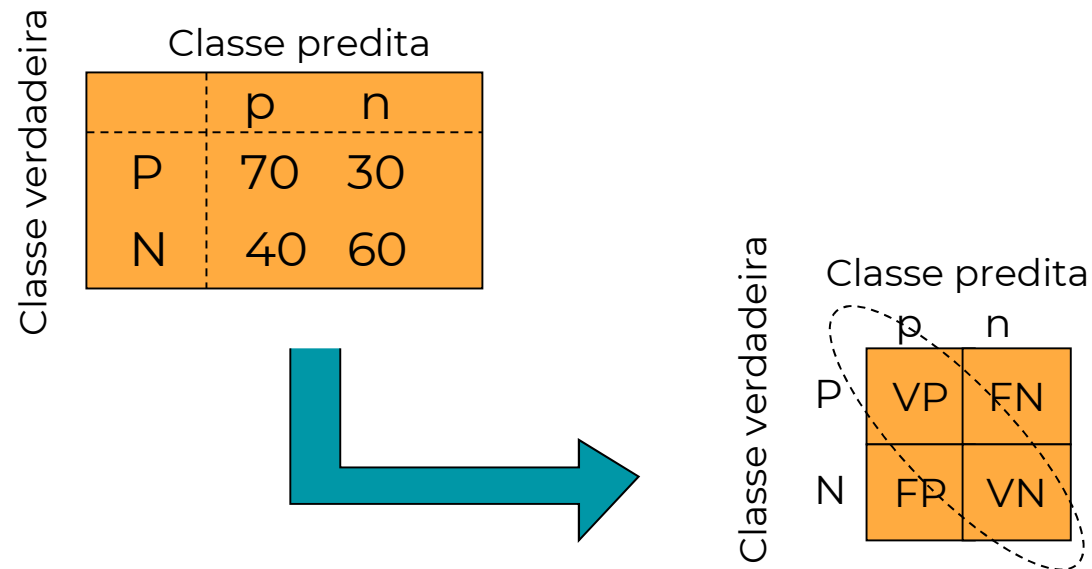
Desempenho preditivo

- 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 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

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

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

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

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

| 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 | N |
| | 20 | 30 |
| | 15 | 35 |

Classificador 1
TVP =
TFP =

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

Classificador 2
TVP =
TFP =

| | | |
|-------------------|----------------|----|
| Classe verdadeira | Classe predita | |
| | p | n |
| | P | N |
| | 60 | 40 |
| | 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 | N |
| | 20 | 30 |
| | 15 | 35 |

Classificador 1
TVP = 0.4
TFP = 0.3

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

Classificador 2
TVP = 0.7
TFP = 0.5

| | | |
|-------------------|----------------|----|
| Classe verdadeira | Classe predita | |
| | p | n |
| | P | N |
| | 60 | 40 |
| | 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

Medida-F

- Média harmônica ponderada da precisão e da revocação

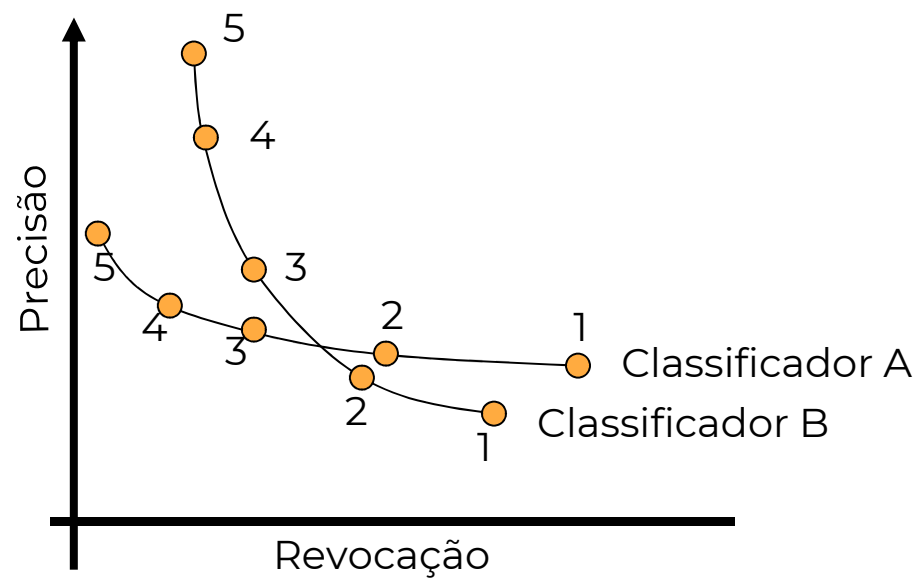
$$\frac{(1 + \alpha) \times (prec \times rev)}{\alpha \times prec + rev}$$

- Medida-F1

- Precisão e revocação têm o mesmo peso

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

Observação



Exemplo

- Seja um classificador com a seguinte matriz de confusão, qual destas medidas apresentará o maior valor?
 - Acurácia
 - Precisão
 - Revocação
 - Especificidade

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

Exemplo

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

$$\text{Precisão} = \frac{VP}{VP + FP}$$

$$\text{Revocação} = \frac{VP}{VP + FN}$$

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

| | | Predito | |
|------------|---|---------|----|
| | | p | n |
| Verdadeiro | P | VP | FN |
| | N | FP | VN |

| | | p | |
|------------|---|----|----|
| | | p | n |
| Verdadeiro | P | 70 | 30 |
| | N | 40 | 60 |

Exemplo

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} = (70 + 60) / (70 + 30 + 40 + 60) = 0.65$$

$$\text{Precisão} = \frac{VP}{VP + FP} = 70 / (70 + 40) = 0.64$$

$$\text{Revocação} = \frac{VP}{VP + FN} = 70 / (70 + 30) = 0.70$$

$$\text{Especificidade} = \frac{VN}{VN + FP} = 60 / (40 + 60) = 0.60$$

| | | Predito | |
|------------|---|---------|----|
| | | p | n |
| Verdadeiro | P | VP | FN |
| | N | FP | VN |

| | | p | |
|------------|---|----|----|
| | | 70 | 30 |
| Verdadeiro | P | 70 | 30 |
| | N | 40 | 60 |

Outras medidas

- Para classificação de dados desbalanceados, a sensibilidade pode ser mais interessante que a especificidade
 - Elas podem ser combinadas em uma medida simples, que busca atender as duas demandas
 - Média geométrica (G-mean)
 - Para duas classes: $G\text{-mean} = \sqrt{\text{revocação} \times \text{especificidade}}$
 - Para mais de duas classes: $G\text{-mean} = (\prod_{i=1}^C \text{Revocação}_i)^{\frac{1}{C}}$
 - Acurácia balanceada
 - Para duas classes: $\text{Acurácia balanceada} = \frac{\text{Especificidade} + \text{Sensibilidade}}{2}$

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



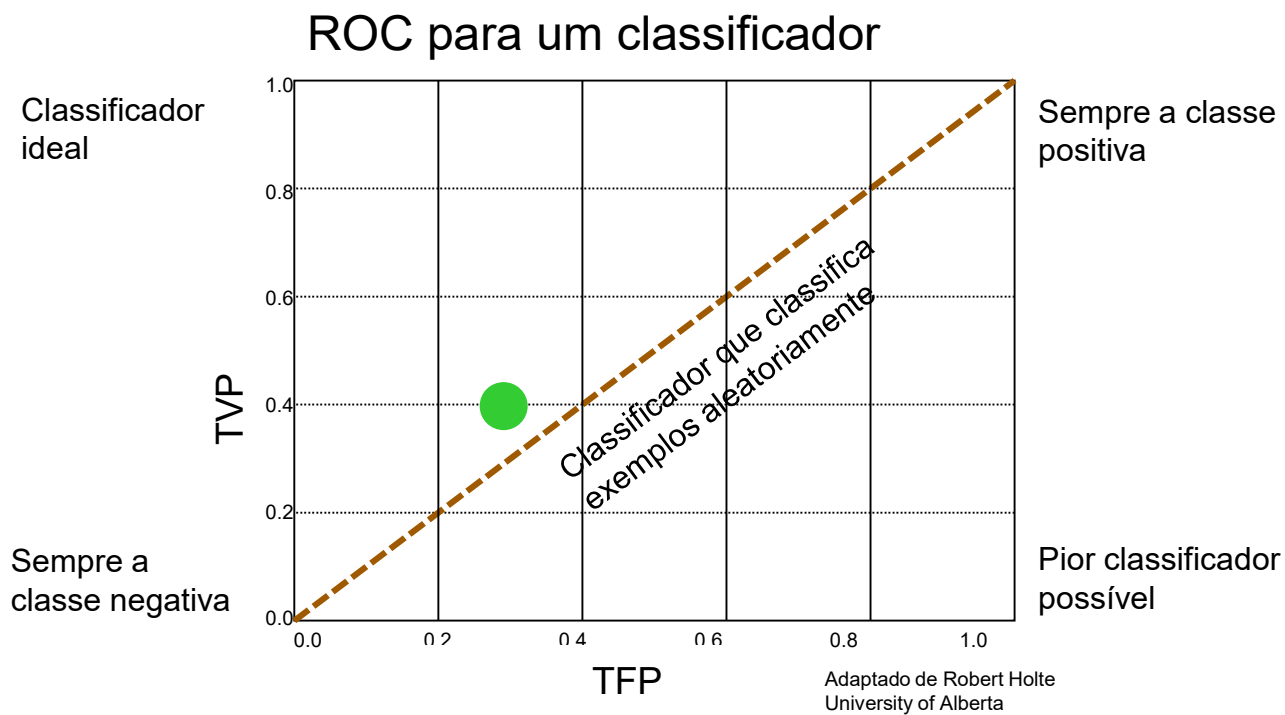
Classificador2
TFP = 0.5
TVP = 0.7



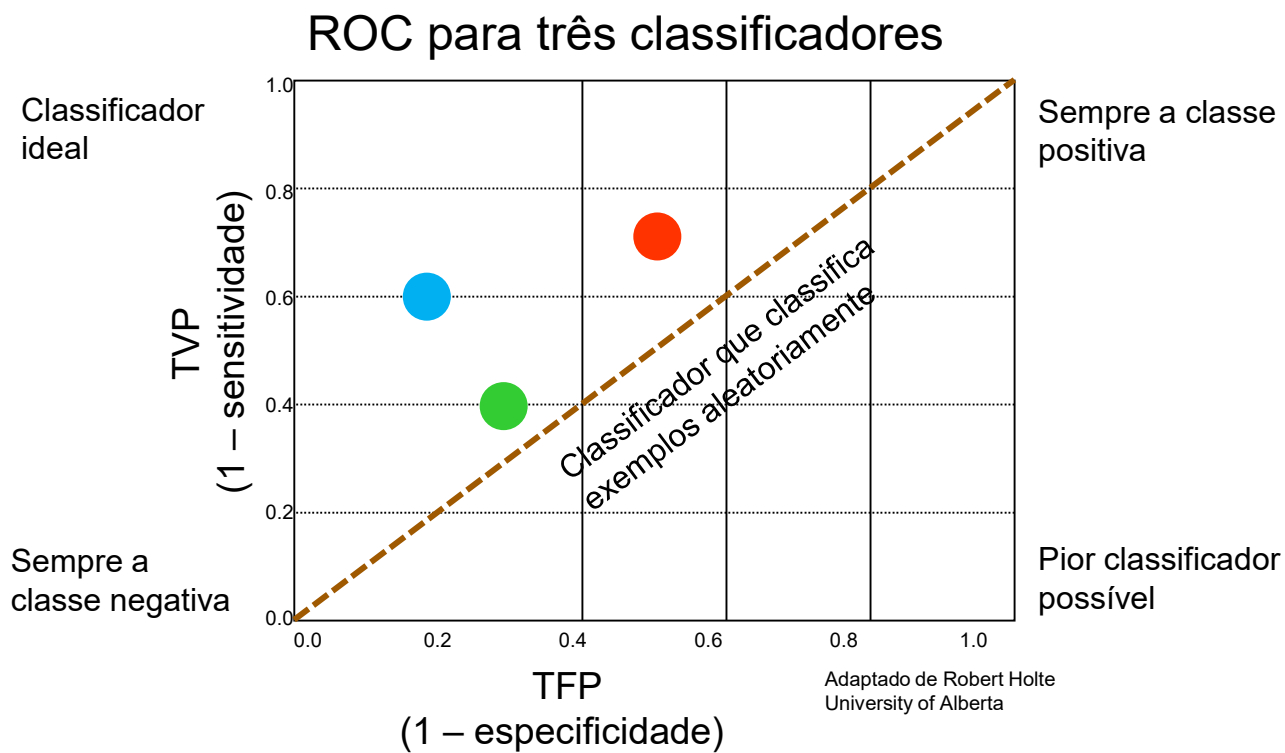
Classificador 3
TFP = 0.2
TVP = 0.6



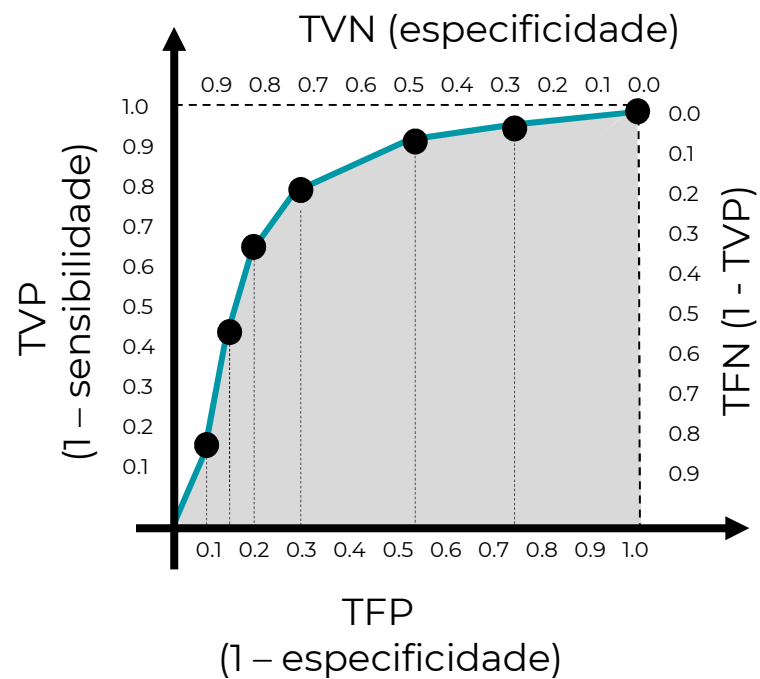
Gráficos ROC



Gráficos ROC



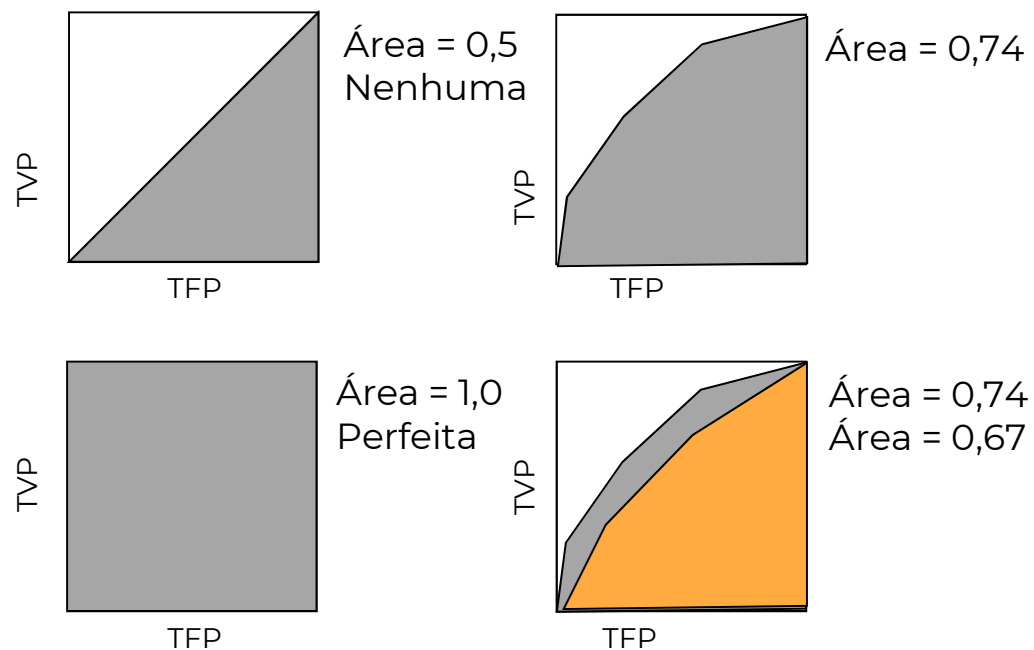
Área sob a Curva ROC



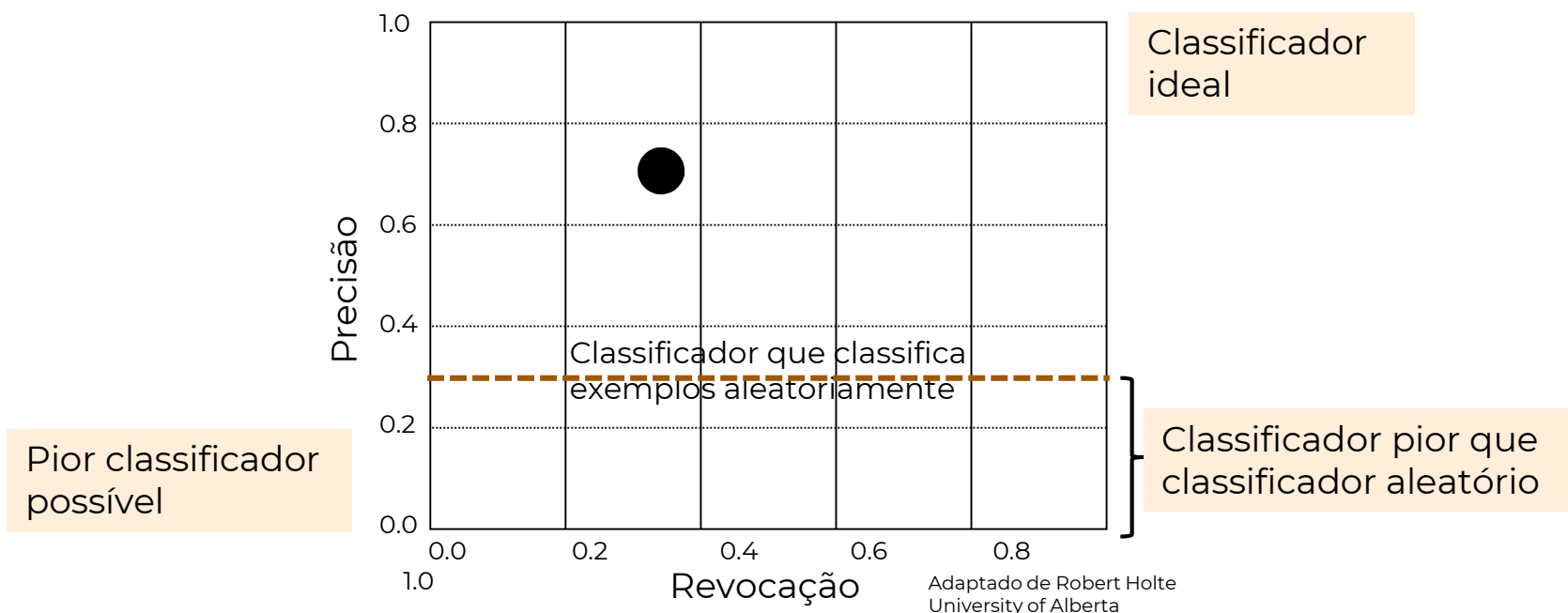
Área sob a curva ROC (AUC)

- 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 (AUC)

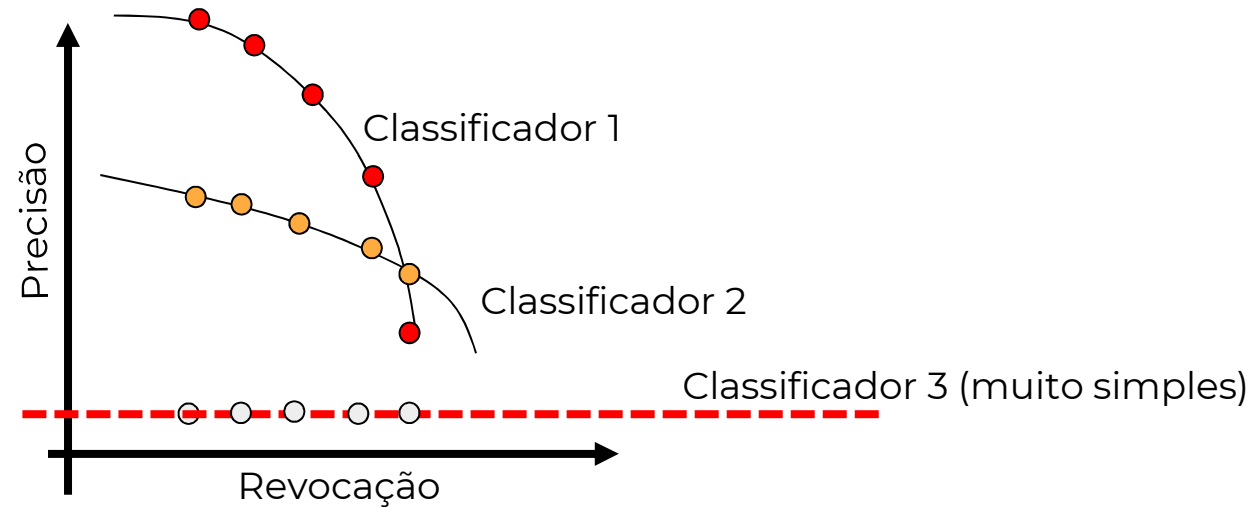


Gráficos precisão-revocação



Curva precisão-revocação

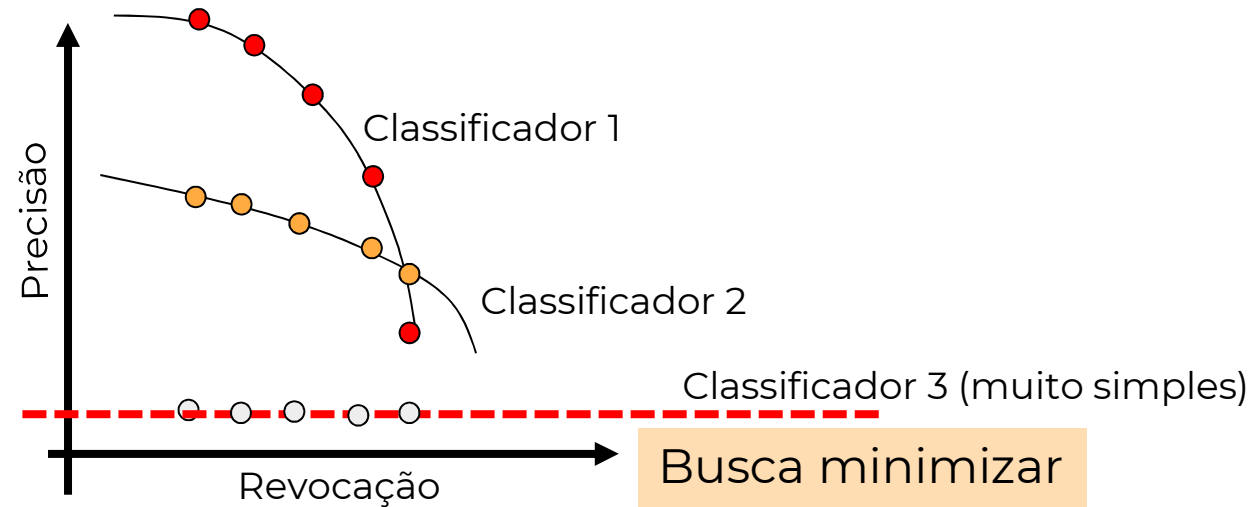
- Mostra relação entre precisão e revocação para diferentes cortes (thresholds)



Curva precisão-revocação

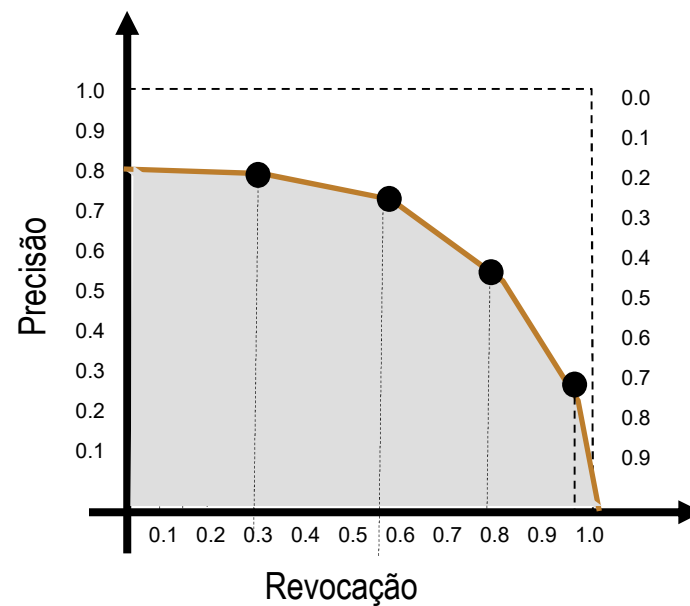
- Mostra relação entre precisão e revocação para diferentes cortes (thresholds)

Busca minimizar
FPs



Busca minimizar
FNs

Área sob a Curva Precisão-Revocação



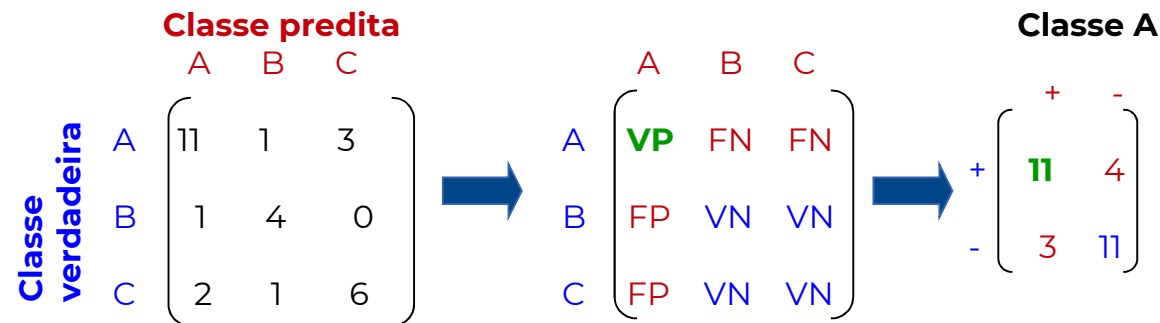
Problemas com mais de 2 classes

- Quando um conjunto de dados tem mais de duas classes:
 - Sejam 3 classes, A, B e C
 - Matriz de confusão:

| | Classe predita | | |
|---|----------------|---|---|
| | A | B | C |
| A | 11 | 1 | 3 |
| B | 1 | 4 | 0 |
| C | 2 | 1 | 6 |

Problemas com mais de 2 classes

- Quando um conjunto de dados tem mais de duas classes:
 - Uma avaliação para cada classe
 - Considera ela a classe positiva (+) e as demais formam a classe negativa (-)
 - Ex. Sejam 3 classes: **A**, B e C



Problemas com mais de 2 classes

- Quando um conjunto de dados tem mais de duas classes:
 - Uma avaliação para cada classe
 - Considera ela a classe positiva (+) e as demais formam a classe negativa (-)
 - Ex. Sejam 3 classes: **A**, B e C

| | | Classe predita | | | | | | | | | | Classe A | |
|-------------------|---|----------------|---|---|---|---|----|----|----|---|---|----------|----|
| | | A | B | C | | | A | B | C | | | + | - |
| Classe verdadeira | A | 11 | 1 | 3 | → | A | VP | FN | FN | → | + | 11 | 4 |
| | B | 1 | 4 | 0 | | B | FP | VN | VN | | - | 3 | 11 |
| | C | 2 | 1 | 6 | | C | FP | VN | VN | | | | |

Precisão (A) = 11/14
Revocação (A) = 11/15

Problemas com mais de 2 classes

- Quando um conjunto de dados tem mais de duas classes:
 - Uma avaliação para cada classe
 - Considera ela a classe positiva (+) e as demais formam a classe negativa (-)
 - Ex. Sejam 3 classes: A, **B** e C

| | | Classe predita | | | | | | | | | | Classe A | |
|-------------------|---|----------------|---|---|---|---|----|-----------|----|---|---|----------|----|
| | | A | B | C | | | A | B | C | | | + | - |
| Classe verdadeira | A | 11 | 1 | 3 | → | A | VN | FP | VN | → | + | 4 | 1 |
| | B | 1 | 4 | 0 | | B | FN | VP | FN | | - | 2 | 22 |
| | C | 2 | 1 | 6 | | C | VN | FP | VN | | | | |

Precisão (A) = 4/6
Revocação (A) = 4/5

Problemas com mais de 2 classes

- Quando um conjunto de dados tem mais de duas classes:
 - Uma avaliação para cada classe
 - Considera ela a classe positiva (+) e as demais formam a classe negativa (-)
 - Ex. Sejam 3 classes: A, B e **C**
 - Exercício: Repetir o exemplo para a classe **C**

Problemas com mais de 2 classes

- Quando um conjunto de dados tem mais de duas classes:
 - Uma avaliação para cada classe
 - Considera ela a classe positiva (+) e as demais formam a classe negativa (-)
 - Ex. Sejam 3 classes: A, B e C

| | | Classe predita | | | | | | | | | | Classe A | |
|-------------------|---|----------------|---|---|---|---|----|----|----|---|---|----------|----|
| | | A | B | C | | | A | B | C | | | + | - |
| Classe verdadeira | A | 11 | 1 | 3 | → | A | VN | VN | FP | → | + | 6 | 3 |
| | B | 1 | 4 | 0 | | B | VN | VN | FP | | - | 3 | 17 |
| | C | 2 | 1 | 6 | | C | FN | FN | VP | | | | |

Precisão (A) = 6/9
Revocação (A) = 6/9

Considerações finais

- Avaliação do desempenho e compreensão
 - Erro
 - Tempo de resposta
 - Memória
 - Interpretabilidade
- Medidas de desempenho preditivo
 - Classificação
 - Duas classes
 - Mais que duas classes

Final da Apresentação

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

