



## Elementos de Aprendizagem Estatística

Lupércio França Bessegato  
Dep. Estatística/UFJF

1



## Modelos de Classificação

Lupércio França Bessegato  
Dep. Estatística/UFJF

2



## Roteiro

- 1. Introdução
- 2. Avaliação de desempenho
- 3. Modelos de classificação linear
- 4. Modelos de classificação não-linear
- 5. Árvores de classificação e modelos baseados em regras
- 6. Desequilíbrio de classes
- 7. Importância e seleção de preditores
- 8. Referências

Elementos de Aprendizagem Estatística – 2024

3



## Medição de Desempenho e Comparação

4



## Roteiro



1. Introdução
2. Predição de classe
3. Calibração de probabilidades
4. Avaliação de probabilidades de classes
5. Critérios não baseados em acurácia
6. Curva *ROC* e *Lift Charts*
7. Referências

Elementos de Aprendizagem Estatística – 2024

5

5



## Introdução



6



## Problema da Classificação



- Como saber se algumas observações pertencem a uma particular população?
  - ✓ Incerteza na classificação

Elementos de Aprendizagem Estatística – 2024

7

7



## Paradoxos da Classificação



- Informação incompleta sobre desempenho futuro:
  - ✓ Classificação de candidato como capaz de concluir ou não um mestrado
- Informação perfeita exige destruição do objeto:
  - ✓ Classificação de itens bons/defeituosos
- Informação cara ou indisponível:
  - ✓ Problemas médicos que só podem ser classificados corretamente com procedimentos complexos

Elementos de Aprendizagem Estatística – 2024

8

2



## Erros de Classificação



- Caso médico:
  - ✓ Em geral, deseja-se diagnosticar um mal a partir de sintomas externos facilmente observáveis
- Erro de classificação:
  - ✓ Pode não ser clara a distinção entre as características medidas das duas populações.

Elementos de Aprendizagem Estatística – 2024

9

9



## Modelos para Respostas Categóricas



- $R^2$  e  $RMSE$  não são adequados no contexto de classificação
- Aspectos das previsões de modelos de classificação e de sua relação com a medição de desempenho.
- Estratégias para avaliar e comparar modelos de classificação usando estatísticas e visualizações.

Elementos de Aprendizagem Estatística – 2024

10

10



## Predição de Classes



## Tipos de Predição



- Em geral, os modelos de classificação geram dois tipos de predição:
  - ✓ Valor contínuo
    - Usualmente na forma de uma probabilidade
  - ✓ Valor discreto
    - Na maioria das aplicações práticas, é necessária uma predição discreta de categoria para tomada de decisão
    - Ex.: Filtro automático de spam

Elementos de Aprendizagem Estatística – 2024

12

12

11

-  • Frequentemente o foco está na predição do valor discreto

- ✓ Mas as estimativas de probabilidade para cada classe podem ser bastante úteis para avaliar a confiança do modelo
  - Ex.:  $P\{\text{spam}\} = 0.51$  vs.  $P\{\text{spam}\} = 0.99$
- ✓ Em algumas aplicações, o resultado desejado são as probabilidades de classe preditas
  - Usadas em outros cálculos
  - Ex.:  $P\{\text{fraude em pagamento de seguros}\}$ , combinada com custo de investigação e de perda para decidir se uma investigação é recomendada.

Elementos de Aprendizagem Estatística – 2024

13

13

## Calibração de Probabilidades

- A probabilidade de classe do modelo deve ser bem calibrada
  - ✓ Deve refletir a probabilidade subjacente da amostra
  - ✓ Exemplo: Filtro de spam
    - Modelo produz uma probabilidade de 20% para a verossimilhança de um particular e-mail ser spam
    - Valor estará bem calibrado se tipos similares de e-mail forem alocados nessa classe, em média, 1 em cada 5 amostras

Elementos de Aprendizagem Estatística – 2024

15

15

-  • Há modelos de predição contínua que não é probabilidade.

- ✓ Redes neurais, modelo de classificação de mínimos quadrados parciais, etc.
- ✓ Deve-se usar alguma transformação para levar a previsões “*probability-like*”
- ✓ Ex.: Transformação *softmax* (Bridle, 1990):

$$\hat{p}_k^* = \frac{e^{\hat{y}_k}}{\sum_{i=1}^C e^{\hat{y}_k}}$$

- $\hat{y}_k$ : previsão do modelo para a k-ésima classe
- $\hat{p}_k^*$ : valor transformado *probability-like* ([0, 1])

Elementos de Aprendizagem Estatística – 2024

14

14

## Gráfico de Calibração

- Probabilidade de um evento
  - ✓ Observada vs. predita.
  - ✓ Pontuar coleção de objetos com classes conhecidas
  - ✓ Agrupar os dados em grupos com base nas probabilidades de classe
    - Ex. intervalos a cada 10%

Elementos de Aprendizagem Estatística – 2024

16

16



- ✓ Determinar a taxa observada de eventos por intervalo

- Ex.: 50 objetos, com alocação de 1 objeto na classe <10%
- Ponto médio do intervalo: 5%, taxa observada: 2%

- ✓ Modelo com probabilidades calibradas:

- Pontos ao longo da reta identidade



Elementos de Aprendizagem Estatística – 2024

17

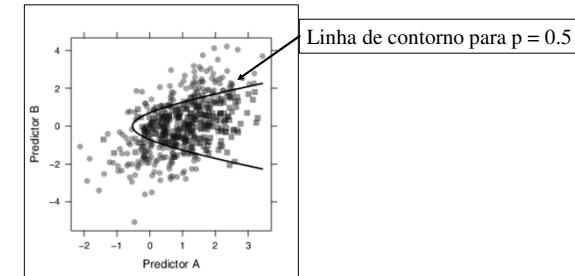
17



- Conjunto de dados simulado:

- ✓ Duas classes e duas variáveis preditoras
- ✓ Probabilidade verdadeira

$$\log\left(\frac{p}{1-p}\right) = -1 - 2A - 0.2A^2 + 2B^2$$



Elementos de Aprendizagem Estatística – 2024

18

18



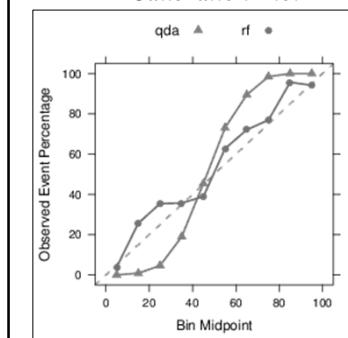
- Ajustados dois modelos de classificação:
  - ✓ Análise discriminante quadrática (QDA)
  - ✓ Random Forest (RF)
  - ✓ Ambos modelos tem acurácia similar (87,1%)
- Conjunto de teste com  $n = 1.000$  objetos



Elementos de Aprendizagem Estatística – 2024

19

19

*Calibration Plot*

Elementos de Aprendizagem Estatística – 2024

20

20

- Modelo QDA tende a ter desempenho pior que do RF
- % observada de eventos no intervalo com probabilidades de classe de 20 a 30%:
  - ✓ QDA (4,6%)
  - ✓ RF (35,4%)
- Modelo QDA mostra padrão sigmoidal, com indicação de probabilidade subestimada
- Modelo indicado para ajuste:
  - ✓ Modelo de regressão logística



- As previsões de classe e os resultados do conjunto de treinamento podem ser usados para pós-processar as estimativas prováveis com a expressão abaixo (Platt 2000):

$$\hat{p}^* = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \hat{p})}$$

✓ Verdadeiras classes preditas com função das probabilidades de classe não calibradas, com os parâmetros  $\beta$

✓ Modelo de classificação QDA do exemplo:

- $-\hat{\beta}_0 = -5.7$
- $-\hat{\beta}_1 = 11.7$

Elementos de Aprendizagem Estatística – 2024



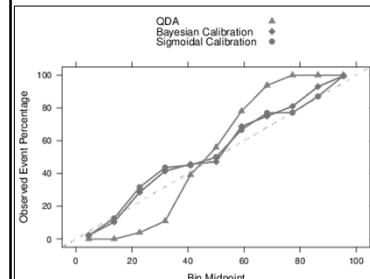
21



### Correção do *Calibration Plot*



- Melhoria na calibração
- Pode-se aplicar a regra de Bayes para recalibrar as previsões.
  - ✓ Abordagem bayesiana também pode melhorar as previsões
- As amostras devem ser reclassificadas após calibração:



Elementos de Aprendizagem Estatística – 2024

22

21

### Visualização das Probabilidades de Classe



- Método efetivo para apresentar os resultados do modelo:
  - ✓ Histograma de classes
  - ✓ *Calibration Plot*
  - ✓ Mapa de calor das probabilidades de classe
  - ✓ Etc.

Elementos de Aprendizagem Estatística – 2024

23



### Classificação binária



- Credit Scoring: Medida da qualidade do crédito
  - Ruim/Bom
- Modelos de classificação usados:
  - Support vector machine – SVM
  - Modelo de regressão logística

Elementos de Aprendizagem Estatística – 2024

24

24

**Histograma das probabilidades -RL**

Kuhn and Johnson (2013) © Springer

- Probabilidade de classificação de crédito ruim para bons clientes é bastante baixa.
  - ✓ Probabilidade da maioria dos clientes é muito baixa.
- A probabilidade de clientes com crédito ruim é plana (uniformemente distribuída)
- Modelo não tem habilidade para distinguir os casos de clientes com crédito ruim

25

Elementos de Aprendizagem Estatística – 2024

25

**Calibration Plot -RL**

- Acurácia da probabilidade de crédito ruim degrada a medida que ela torna-se maior.
  - ✓ Não há observações com crédito ruim acima de 82,7%
- Padrão indicativo de modelo de classificação com calibração e desempenho ruins

26

Elementos de Aprendizagem Estatística – 2024

26

**Mapa de Calor das Probabilidades de Classe**

Kuhn and Johnson (2013) © Springer

- Classificação com três ou mais classes
  - ✓ Mapa de calor das probabilidades de classe pode ajudar a avaliar a confiança nas previsões
  - ✓ Conjunto de teste com:
    - oito classes: nível A a I
    - Objetos: 48

27

Elementos de Aprendizagem Estatística – 2024

27

**Mapa de Calor das Probabilidades de Classe**

Kuhn and Johnson (2013) © Springer

- Classes
  - ✓ Verdadeiras: linhas.
  - ✓ Preditas: probabilidades de classes nas colunas
- Sinal claro da classe predita
  - ✓ Objeto 20: 78,5%
- Casos obscuros:
  - ✓ Objeto 7: 19,6% (B); 19,3% (C); 17,7% (A) e 15% (E)
  - ✓ Embora o modelo alocar o objeto na classe correta (19,6%), há incerteza se ela poderia ser também da classe C, A ou E.

28

Elementos de Aprendizagem Estatística – 2024

7



## Zonas Indeterminadas



- Abordagem para melhorar o desempenho da classificação
  - ✓ Criar zona indeterminada
  - ✓ Classe não é formalmente predita quando a confiança não é alta

29

Elementos de Aprendizagem Estatística – 2024



- Problema de classificação binária:

- ✓ Zona de indeterminação:  $0.5 \pm z$ 
  - $Z = 0.10$ : objetos com probabilidade de predição entre 0.40 e 0.60 considerados indeterminados
- ✓ Desempenho do modelo calculado sem as observações da zona de indeterminação
- ✓ Taxa de indeterminação deve ser relatada
  - Taxa de resultados imprevistos devem ser bem compreendidos



30

Elementos de Aprendizagem Estatística – 2024

29

30



- Classificação com mais de duas classes ( $C > 2$ )



- ✓ Podem ser aplicados limites semelhantes:
  - Maior probabilidade de classe maior que  $\frac{1}{C} + z$
- ✓ Para os dados da figura do mapa de calor:
  - Se  $\frac{1}{C} + z$  for definido como 30%
  - 5 amostras seriam designadas como indeterminadas

31

Elementos de Aprendizagem Estatística – 2024



## Avaliação da Predição das Classes



31

32

**Matriz de Confusão para 2 Classes**

- Tabela de contingência:
  - ✓ Classes observadas e preditas

		Observado	
		Evento	Não evento
Predito	Evento	TP	FP
	Não Evento	FN	TN
– Diagonal: previsões corretas – Fora diagonal: cada um dos erros possíveis • FP: erro tipo I e FN: erro tipo II			
- ✓ Comum para descrever desempenho do modelo de predição

Elementos de Aprendizagem Estatística – 2024

33

33

**Métricas**

- Taxa de acurácia global  
(*overall accuracy rate*)
  - ✓ Reflete a concordância entre observação e predição
$$\text{acurácia} = \frac{TP + TN}{n}$$
- ✓ Probabilidade de classificações corretas:

$P\{(C^+ \cap E^+) \cup (C^- \cap E^-)\}$

Elementos de Aprendizagem Estatística – 2024

34

34

**Desvantagens:**

- ✓ Não leva em consideração o tipo de erro cometido
- ✓ Em situações em que os custos são diferentes, as características importantes do modelo podem não ser medidas pela acurácia
  - Discussão ampla em Provost *et. al* (1998)

Elementos de Aprendizagem Estatística – 2024

35

35

**Desvantagens:**

- ✓ Deve-se considerar as frequências naturais de cada classe
  - Exemplo: prevalência de desordem em recém-nascido é aproximadamente 1 em 800 ≈ 0.1%
  - Um modelo de classificação poderia atingir acurácia quase perfeita ao prever que todos os objetos serão negativos para o desordem

Elementos de Aprendizagem Estatística – 2024

36

**Baseline**

- Qual referência da acurácia deveria ser usada para determinar se um modelo de classificação está adequado?
  - ✓ Taxa de ausência de informação (não-informação):
    - Taxa que pode ser alcançada sem um modelo
    - Para  $C$  classes, definição mais simples baseia-se em aleatoriedade:  $\frac{1}{C}$
    - Não leva em consideração as frequências relativas das classes do conjunto de treinamento

37

Elementos de Aprendizagem Estatística – 2024

37

✓ Exemplo do distúrbio em recém-nascido:

- Em amostra de tamanho 1.000, é pequena a quantidade esperada de bebês com distúrbio para
  - (1 ou 2)
- Taxa de não-informação (50%) é superada, se ele prediz que todos os objetos não tem distúrbio
- Alternativa para definir a taxa de não-informação
  - Porcentagem da maior classe do conjunto de treinamento
  - Modelos com acurácia maior que esta taxa deveriam ser considerados razoáveis

✓ Desequilíbrios graves no tamanho de cada classe impactam a acurácia do modelo. 00

38

Elementos de Aprendizagem Estatística – 2024

38

- Taxa de erro global (overall error rate)
  - ✓ Reflete a discordância entre observação e predição

$$\text{taxa de erro} = 1 - \text{acurácia} = \frac{FP + FN}{n}$$

39

Elementos de Aprendizagem Estatística – 2024

39

- Em vez de calcular a acurácia global e compará-la com a taxa de não-informação podem ser usadas outras métricas que levam em consideração as distribuições de classes dos objetos do conjunto de treinamento.

40

Elementos de Aprendizagem Estatística – 2024

40



- Estatística Kappa (Kappa de Cohen)

✓ Reflete a discordância entre observação e predição

- Desenvolvida originalmente para avaliar a discordância entre dois avaliadores

$$\text{Kappa} = \frac{O - E}{1 - E}$$

- O: acurácia observada (modelo de classificação)  
- E: acurácia esperada (baseada nas marginais)

✓ A estatística considera a acurácia que ocorreria simplesmente por acaso



- A estatística assume valores entre – 1 e 1.

✓ Valor 0:

- não há concordância entre as classes observadas e preditas.

✓ Valor 1:

- concordância perfeita entre as classes observadas e preditas.

✓ Valores negativos:

- Predição está em direção oposta à verdadeira

- Se distribuição das classes é equivalente

✓ Acurácia global e Kappa são proporcionais.



- Dependendo do contexto, Kappa entre 0.30 e 0.50 indicam concordância razoável.

✓ Suponha acurácia do modelo alta (90%), com acurácia esperada também alta (85%).

✓ Kappa mostraria concordância moderada (33%)



- Alguns autores utilizam as seguintes faixas:

✓ Concordância baixa: < 0.20

✓ Concordância satisfatórias: 0.20 a 0.40

✓ Concordância moderada: 0.40 a 0.60

✓ Boa concordância: 0.60 a 0.80

✓ Concordância muito boa: 0.80 a 1.0



- Coeficiente Kappa pode ser estendido para problemas com mais de duas classes
- Quando a variável categórica for ordinal, pode-se ponderar o coeficiente, impondo penalidades maiores sobre erros que estão mais distantes do verdadeiro resultado:
  - ✓ Nível “baixo” predito erroneamente como “alto”, Kappa seria reduzido mais que um nível “baixo” fosse predito como “médio”



Elementos de Aprendizagem Estatística – 2024

45

45

## Classificação Binária



- Em problemas de classificação binária, em que uma classe for interpretada como evento de interesse :
  - ✓ Estatísticas adicionais podem ser relevantes evento de interesse
- Matriz de confusão:

		Observado	
		Evento ( $E^+$ )	Não evento ( $E^-$ )
Predito	Evento ( $C^+$ )	TP	FP
	Não Evento ( $C^-$ )	FN	TN

Elementos de Aprendizagem Estatística – 2024

46

46



## Matriz de Confusão para 2 Classes



- Tabela de contingência:

✓ Classes observadas e preditas

		Observado	
		Evento	Não evento
Predito	Evento	TP	FP
	Não Evento	FN	TN

– Diagonal: previsões corretas

– Fora diagonal: erros de cada um dos casos possíveis

– FP: erro tipo I e FN: erro tipo II

✓ Comum em descrição de desempenho

Elementos de Aprendizagem Estatística – 2024

47

47



### Sensibilidade:



✓ Taxa de predição correta do evento de interesse para todas as observações do evento

$$\text{sensibilidade} = P(C^+|E^+) = \frac{P(C^+ \cap E^+)}{P(E^+)} = \frac{TP}{TP + FN}$$

✓ Algumas vezes a sensibilidade é considerada como taxa de verdadeiros positivos

– Mede a acurácia da predição na população do evento

✓ Algumas autores consideram a taxa de falso-negativos como sendo  $P(C^-|E^+)$  x :

$$1 - \text{sensibilidade}$$

Elementos de Aprendizagem Estatística – 2024

48

48



- Especificidade:

- ✓ Taxa de predição correta de não-eventos na população de não-eventos:
  - Desenvolvida originalmente para avaliar a discordância entre dois avaliadores

$$\text{especificidade} = P(C^- | E^-) = \frac{P(C^- \cap E^-)}{P(E^-)} = \frac{TN}{TN + FP}$$

- ✓ Alguns autores consideram a taxa de falso-positivos como sendo  $P(C^+ | E^-) \times$ :

1 - especificidade



## Sensibilidade e Especificidade Trade-off



- Assumindo um nível fixo de acurácia, há um *trade-off* entre especificidade e sensibilidade
- Intuitivamente:
  - ✓ Aumento da sensibilidade de modelo incorre na perda de especificidade
    - Mais objetos preditos como eventos



- Potenciais *trade-offs* podem ser apropriados

- ✓ Diferentes penalidades associadas a cada tipo de erro.

- ✓ Ex.: Filtragem de spam o foco é especificidade
  - Pessoas podem estar dispostas a receber spams, se e-mails de interesse não forem excluídos.



- Curva ROC

- ✓ *Receiver Operating Characteristic Curve*

- ✓ Técnica para avaliar *trade-off* entre sensibilidade e especificidade.

## Exemplo



- *Credit Scoring*

- ✓ Predição da qualidade de crédito
- ✓ Modelos: SVM e Regressão logística
- ✓ Conjunto de teste: 200 clientes
- ✓ Matriz de confusão – Modelo RL

- Evento: Crédito Ruim

		Observado		
		Ruim	Bom	
Predito	Ruim	24	10	34
	Bom	36	130	166
		60	140	200

		Observado		
		Ruim	Bom	
Predito	Ruim	24	10	34
	Bom	36	130	166
		60	140	200
✓ Acurácia global:	$\frac{24 + 130}{200} = 77\%$			
✓ Taxa de não-informação:	$\frac{140}{200} = 70\%$			
• Acurácia global ligeiramente melhor que taxa de não-informação				

Elementos de Aprendizagem Estatística – 2024

53

53

		Observado		Esperado	
		Ruim	Bom	Ruim	Bom
Predito	Ruim	24	10	34	34
	Bom	36	130	166	166
		60	140	200	200

✓ Acurácia observada:  $O = \frac{24 + 130}{200} = 77\%$

✓ Acurácia esperada:  $E = \frac{10,2 + 116,2}{200} = 63,2\%$

✓ Coeficiente Kappa:

$$\text{Kappa} = \frac{O - E}{1 - E} = \frac{0.77 - 0.632}{1 - 0.632} = 37,5\%$$

- Kappa sugere concordância moderada

Elementos de Aprendizagem Estatística – 2024

54

54

		Observado		
		Ruim(+)	Bom(-)	
Predito	Ruim(+)	24	10	34
	Bom(-)	36	130	166
		60	140	200

✓ Sensibilidade:  $s = \frac{24}{60} = 40\%$

✓ Especificidade:  $e = \frac{130}{140} = 92.9\%$

✓ Dificuldade em predizer crédito RUIM

- Modelo pode estar afetado pelo desequilíbrio das classes e/ou ausência de forte preditor de crédito ruim

Elementos de Aprendizagem Estatística – 2024

55

55

- Índice J de Youden:

✓ Medida para refletir as taxas de falso-positivo e falso-negativo:  
 $J = \text{sensibilidade} + \text{especificidade} - 1$

✓ Utiliza as proporções de previsões corretas para ambas as classes.

✓ Em alguns contextos, pode ser um método apropriado para resumir a magnitude de ambos os tipos de erros

✓ Método mais comum para combinar sensibilidade e especificidade

- Curva ROC

Elementos de Aprendizagem Estatística – 2024

56

56



## Valores Preditivos



- Sensibilidade e especificidade são medidas condicionais
- Suponha paciente não portador de doença, com teste diagnóstico com acurácia de 95%
  - ✓ Qual a probabilidade de paciente portar a doença?
  - ✓ Depende da sensibilidade, da especificidade e da prevalência da doença.

Elementos de Aprendizagem Estatística – 2024

57

57



- Levando em consideração a prevalência:

✓ Valor preditivo positivo (VPP)

-  $P(E^+|C^+)$

- Análogo da sensibilidade

$$VPP = \frac{s \times p}{s \times p + (1 - e)(1 - p)}$$

✓ Valor preditivo negativo (VPN)

-  $P(E^-|C^-)$

- Análogo da especificidade

$$VPN = \frac{e \times (1 - p)}{(1 - s) \times p + e \times (1 - p)}$$

Elementos de Aprendizagem Estatística – 2024

58

58



- Valores preditivos são combinações não-triviais do desempenho do modelo e da taxa de eventos da população.

Elementos de Aprendizagem Estatística – 2024

59

59



- Decisões incorretas:

✓ PFP:  $\overline{PFP} = P(E^-|C^+) = 1 - P(E^+|C^+) = 1 - VPP$

✓ PFN:  $\overline{PFN} = P(E^+|C^-) = 1 - P(E^-|C^-) = 1 - VPN$

Elementos de Aprendizagem Estatística – 2024

60

60

• Efeito da prevalência nos valores preditivos

- Podem ser alcançados grandes VPN com prevalência baixa.
- Com aumento da prevalência VPN torna-se muito pequeno
- Ocorre oposto com VPP
- Variação na sensibilidade (90% vs. 99%) tem pouco efeito sobre VPP.

Elementos de Aprendizagem Estatística – 2024

61

• Nesse contexto é definido *recall* como:

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

✓ A expressão é a mesma da sensibilidade, mas há uma ligeira diferença na interpretação

- Modelo com recall alto tem grande amplitude
  - Captura grande parte dos objetos positivo (evento)

Elementos de Aprendizagem Estatística – 2024

63

• Taxa de eventos balanceada ( $p = 0.50$ )

✓ Neste caso:  $VPP = \frac{s}{s + (1 - e)} = \frac{TP}{TP + FP}$

✓ Denominada **precisão** por alguns autores

- Sensibilidade tem efeito menor que a especificidade.
- Se especificidade for alta (digamos > 90%) VPP grandes podem ser obtidos por meio de ampla gama de sensibilidades.

Elementos de Aprendizagem Estatística – 2024

62

• Medida F (F1-score ou F-escore)

✓ Medida de desempenho do modelo que combina precisão e recall em um único valor.

✓ Usa média harmônica da precisão e do recall

$$\text{F-measure} = \frac{2 \times \text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

✓ Precisão e recuperação estão expressas como proporções entre 0 e 1, podendo ser interpretadas como taxas.

Elementos de Aprendizagem Estatística – 2024

64



- Em geral, os valores preditivos não são usados para caracterizar o modelo:
  - ✓ A prevalência é difícil de ser quantificada
  - ✓ A prevalência é dinâmica
    - Taxa de spams aumenta no início dos esquemas, mas depois cai para níveis básicos
    - Prevalência de doenças pode variar bastante dependendo da localização



## Outros Critérios de Desempenho



## Critérios Não Baseados na Acurácia



- Há situações em que o objetivo principal do modelo preditivo não é a acurácia.
- Aplicações práticas o objetivo de:
  - ✓ Prever oportunidades que maximizem o retorno
  - ✓ Melhorar satisfação do cliente
  - ✓ Aprimorar previsões de demanda para reduzir custos de estoque
  - ✓ Reduzir custo de transações fraudulentas



- Embora importante a acurácia apenas descreve quão bem o modelo prediz dados.
  - ✓ Necessárias outras métricas de desempenho que quantifiquem as consequências de previsões corretas e incorretas
    - Custos e benefícios.



**Exemplo**

- Detecção de fraudes
  - ✓ Evento de interesse: fraude
  - ✓ Predição de fraude (TP ou FP)
    - Custo associado à investigação
    - Benefício de detectar transações ruins (TP)
  - ✓ Não predição de fraude (FN)
    - Perda de renda

69

Elementos de Aprendizagem Estatística – 2024

**Exemplo**

- Promoções de e-commerce
  - ✓ Evento de interesse: responde promoção

		Observado		Observado	
		Resposta	Não resp.	Resposta	Não resp.
Predito	Resposta	TP	FP	\$26.40	-\$2,00
	Não resp.	FN	TN	-\$28,40	

✓ Lucro médio quando há resposta: \$28,40
 

- Custo das despesas de promoção: \$2,00 (FP)
- Lucro líquido quando há resposta: \$26,40 (TP)
- Perda de venda potencial: \$28,40 (FN)

70

Elementos de Aprendizagem Estatística – 2024

**Exemplo**

- Promoções de e-commerce:

		Observado		Observado	
		Resposta	Não resp.	Resposta	Não resp.
Predito	Resposta	TP	FP	\$26.40	-\$2,00
	Não resp.	FN	TN	-\$28,40	

✓ Lucro total para um modelo:

$$\text{Lucro} = \$26.40 \text{ TP} - \$2.00 \text{ FP} + \$28.40 \text{ FN}$$

- Deve-se considerar a prevalência das classes

✓ Sabe-se que taxa de respostas é baixa

- Lucro esperado impulsionado por custos de FN

✓ Lucro líquido quando há resposta: \$26,40

71

Elementos de Aprendizagem Estatística – 2024

**Exemplo**

- Matriz de confusão para modelo preditivo:

- ✓ Clientes potenciais: 20.000
- ✓ Taxa de resposta: 10%
- ✓ Sensibilidade: 75,0%
- ✓ Especificidade: 94,4%

		Observado		Observado	
		Resposta	Não resp.	Resposta	Não resp.
Predito	Resposta	1.500	1.000	\$39.600	-\$2.000
	Não resp.	500	17.000	-\$14.200	
Total			\$23.400 (\$1,17/consumidor)		

72

Elementos de Aprendizagem Estatística – 2024

- Matriz de confusão para modelo preditivo: 

- ✓ Clientes potenciais: 20.000
- ✓ Taxa de resposta: 10%
- ✓ Sensibilidade: 75,0%
- ✓ Especificidade: 100%

		Observado		Observado	
		Resposta	Não resp.	Resposta	Não resp.
Predito	Resposta	1.500		\$39.600	
	Não resp.	500	18.000	-\$14.200	
	Total		\$25.400 (\$1,27/consumidor)		

- ✓ Aumento significativo no desempenho, com ganho marginal
  - Promoção com baixo custo.

Elementos de Aprendizagem Estatística – 2024

73

- Matriz de confusão para modelo preditivo: 

- ✓ Esboço geral para incorporação de custos desiguais como medida de desempenho

- ✓ Drummond e Holt (2001)

- ✓ Função custo-probabilidade

(PCF – Probability Cost Function)

$$\text{PCF} = \frac{p \times C(-|+)}{p \times C(-|+) + (1-p)C(+|-)}$$

- ✓ p: probabilidade *a priori* do evento.

- ✓ C (-|+): custo de predição incorreta de evento (+)

- ✓ C (+|-): custo de predição incorreta de não-evento (-)

Elementos de Aprendizagem Estatística – 2024

75

- Matriz de confusão para modelo preditivo: 

- ✓ Promoção em massa para todos os clientes
- ✓ Sensibilidade perfeita (100%)

		Observado		Observado	
		Resposta	Não resp.	Resposta	Não resp.
Predito	Resposta	2.000	18.000	\$39.600	
	Não resp.			-\$14.200	
	Total			\$16.800 (\$0,84/consumidor)	

- ✓ Deveria ser considerado como *baseline*.

- Lucros dos modelos considerados acima e abaixo da promoção em massa.

Elementos de Aprendizagem Estatística – 2024

74

- Esboço geral para incorporação de custos desiguais como medida de desempenho 

- ✓ Drummond e Holt (2001)

- ✓ Função custo-probabilidade

(PCF – Probability Cost Function)

$$\text{PCF} = \frac{p \times C(-|+)}{p \times C(-|+) + (1-p)C(+|-)}$$

- ✓ p: probabilidade *a priori* do evento.

- ✓ C (-|+): custo de predição incorreta de evento (+)

- ✓ C (+|-): custo de predição incorreta de não-evento (-)

Elementos de Aprendizagem Estatística – 2024

75

- Os autores sugerem usar o custo esperado normalizado 

(NEC – Normalized Expected Cost)

$$\text{NEC} = \text{PCF} \times (1 - TP) + (1 - \text{PCF}) \times FP$$

- ✓ Considera:

- prevalência do evento, desempenho do modelo e os custos

- ✓ Custo total escalonado entre 0 e 1

- ✓ Abordagem atribui custos apenas para erros

- Pode não ser adequada para problemas com outros tipos de custos ou benefícios.

Elementos de Aprendizagem Estatística – 2024

76

75

76



## Probabilidades de Classe



- Probabilidades de classe oferecem mais informação sobre as previsões do modelo que os simples valores de classe.

Elementos de Aprendizagem Estatística – 2024

77

77



## Curva ROC (*Receiver Operating Characteristic*)



## Curva ROC



- Método geral que determina um limite efetivo acima do qual seja indicado um evento específico
  - ✓ Sensibilidade vs.  $(1 - \text{especificidade})$
  - ✓ Pode ser usado no contexto de:
    - Seleção de variáveis
    - Determinação de pontos de corte alternativos para as probabilidades de classe

Elementos de Aprendizagem Estatística – 2024

79

79



- Modelo de predição de crédito ruim
  - ✓ Modelo de regressão logística
    - Sensibilidade: 40,0% (baixa)
    - Especificidade: 92,9% (alta)
    - Classificação calculada com limite default (50%)
  - ✓ Pode-se melhorar a sensibilidade diminuindo o limite?
    - Capturar mais verdadeiros positivos

Elementos de Aprendizagem Estatística – 2024

80

80

**• Curva ROC:**

- ✓ Avalia as probabilidades de corte por meio de contínuos pontos de corte

81

81

**• Modelo de predição de crédito ruim**

- ✓ Análise da curva *ROC*:

82

82

**• Influência do ponto de corte:**

83

83

**• Curva *ROC*:**

- ✓ Gráfico útil para escolher ponto de corte que maximize limite entre sensibilidade e especificidade
- ✓ Entretanto, alterar limite apenas tem efeito de tornar as amostras mais positivas/negativas
  - Alteração de limite não move objetos das células da diagonal secundária da matriz de confusão.
- ✓ Em geral, o aumento da sensibilidade (especificidade) leva à diminuição da outra medida.

84

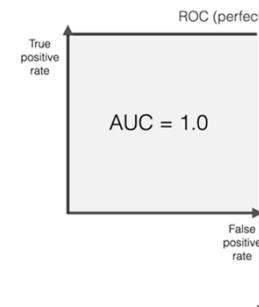
84



## Avaliação Quantitativa do Modelo



- Modelo perfeito  
(Separação completa das duas classes)
  - ✓ Sensibilidade: 100%
  - ✓ Especificidade: 100%
  - ✓ Degrau único:  $(0, 0)$  e  $(0, 1)$   
Constante:  $(0, 1)$  a  $(1, 1)$
  - ✓ Área sob a curva: 1, 0



Elementos de Aprendizagem Estatística – 2024

85



- Modelo não informativo

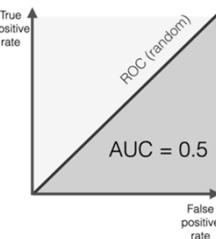
(Modelo sem efetividade)

✓ Sensibilidade = 1 – especificidade

✓ Curva ROC próximo à identidade

- Abaixo método não é melhor que decidir com uma moeda

✓ Área sob a curva: 0, 5



Elementos de Aprendizagem Estatística – 2024

86

85

86



- Comparação de modelos
  - ✓ Sobreposição de curvas no mesmo gráfico
  - ✓ Comparação de:
    - Mesmo modelo com diferentes preditoras
    - Parâmetros de ajuste diferentes
    - Classificadores diferentes
  - ✓ Modelo ótimo deveria estar deslocado para canto superior esquerdo do gráfico.
  - ✓ Modelo mais eficaz
    - Modelo com maior área sob a curva ROC.

Elementos de Aprendizagem Estatística – 2024

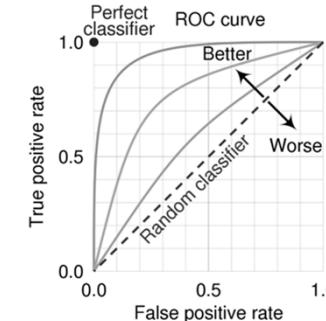
87



- Visualização comparativa de modelos



ROC Visualization:



Elementos de Aprendizagem Estatística – 2024

88

88

**Área sob a Curva - AUC**

- X

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

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

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

$$AUC = \sum_{i=1}^{n-1} \frac{(x_{i+1} - x_i) \times (y_i + y_{i+1})}{2}$$

\* n is the number of threshold points.  
\*  $(x_i, y_i)$  and  $(x_{i+1}, y_{i+1})$  are the FPR and TPR coordinates of two consecutive points on the ROC curve.

89

Elementos de Aprendizagem Estatística – 2024

- Modelo de predição de crédito ruim
  - ✓ Área sob curva *ROC*: 0.78
  - ✓ IC para AUC: (0.70; 0.85)

90

Elementos de Aprendizagem Estatística – 2024

- Escores da área sob a curva ROC

Lantz (2019)

- ✓ Excelente: 0.90 a 1.00
- ✓ Bom: 0.80 a 0.90
- ✓ Aceitável/Regular: 0.70 a 0.80
- ✓ Ruim: 0.60 a 0.70
- ✓ Sem discriminação: 0.50 a 0.60

92

Elementos de Aprendizagem Estatística – 2024

**AUC – Caracterização de Modelos**

- Vantagem:
  - ✓ Insensível a disparidades nas proporções de classe
    - É função apenas da sensibilidade e especificidade
- Desvantagem:
  - ✓ Avaliação por AUC oculta informação.

93

Elementos de Aprendizagem Estatística – 2024



- Exemplos de aplicação
  - ✓ É comum a sobreposição de curvas ROC
    - Nenhum modelo uniformemente melhor que outro
  - ✓ Perda de informação ao resumir a curva
    - Em especial, quando a subárea for de interesse
    - Ex.: Modelo com curva com inclinação acentuada na esquerda e AUC menor que de outro modelo



94

Elementos de Aprendizagem Estatística – 2024



- A curva ROC está definida apenas para problemas de classificação binária
  - ✓ Extensões para uso em situações com três ou mais classes.
    - Hand e Till (2001)
    - Lachice e Flach (2003)
    - Li e Fine (2008)



95

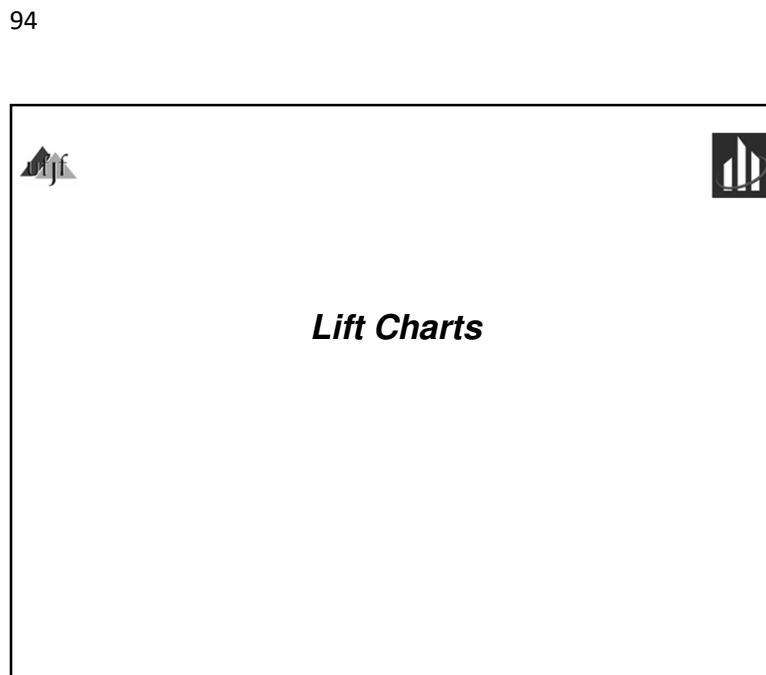
Elementos de Aprendizagem Estatística – 2024

94

95



## **Lift Charts**



96



## **Lift Chart**



- Avaliação da capacidade do modelo em detectar eventos em classificação binária
  - ✓ Objetos pontuados usando a probabilidade de classe de evento
  - ✓ Ordenação por probabilidade de classe
    - Espera-se que os eventos sejam classificados acima dos não-eventos.
  - ✓ *Lift Chart:*
    - Classifica objetos pelo escore e determina taxa acumulada de eventos

Elementos de Aprendizagem Estatística – 2024

97

97



## Procedimento para Construção



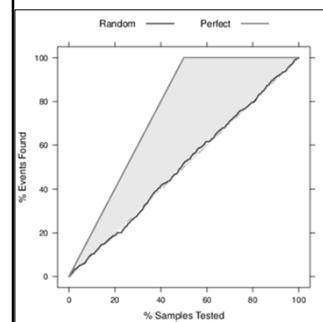
1. Predizer classificação conjunto de teste
  - Não utilizados na construção do modelo
2. Determinar taxa de eventos *baseline*
  - % eventos verdadeiros em todo o conjunto
3. Ordenar dados conforme probabilidade de classificação do evento de interesse.
4. Para cada valor de probabilidade (único)
  - Calcular % eventos verdadeiros abaixo dele
5. Em cada limite de probabilidade
  - Dividir % de eventos verdadeiros pela taxa de eventos baseline.

Elementos de Aprendizagem Estatística – 2024

98

98

✓ Melhor e pior *lift curve*, em conjunto de dados com taxa de eventos de 50%



Elementos de Aprendizagem Estatística – 2024

100

100



- *Lift chart* representa ganho/elevação acumulado em relação à % de objetos que foram classificados

- Modelo não-informativo:

- ✓ X% dos dados com classificação mais alta conteriam X eventos, em média
- ✓ Aumento é quantidade de objetos detectados pelo modelo, acima de uma seleção aleatória dos objetos.

Elementos de Aprendizagem Estatística – 2024

99

99



## Comparação de Modelos



- Comparação de *lift curves* para encontrar modelo mais adequado
  - ✓ Pode-se usar a área sob a curva (AUC)
- Algumas partes da *lift curve* podem ser mais interessantes que outras
  - ✓ Provavelmente a parte mais importante será aquela associada com os objetos com classificação mais alta

Elementos de Aprendizagem Estatística – 2024

101

101



Referências

244



### Bibliografia Recomendada

- JAMES, G. *et al. An introduction to statistical learning with applications in R.* Springer, 2013.
- KUHN, M.; JOHNSON, K. *Applied predictive modeling.* Springer, 2013.
- LESMEISTER, C. *Mastering machine learning with R: Advanced machine learning techniques for Building smart applications with R.* Packet Publishingd. 2019.

Análise Multivariada - 2022

245