

MATB16

Laboratório de Inteligência Artificial

Validação

Tatiane Nogueira Rios
Ricardo Araújo Rios

LabIA
Instituto de Computação - UFBA



1

Agenda

- Introdução
- Métricas de erro
- Amostragem
- Espaço ROC



2

Introdução

- A avaliação é a chave para fazer um progresso real em AM.
- A fim de determinar quais algoritmos de AM usar em um problema particular, precisamos de formas sistemáticas para avaliar como métodos diferentes trabalham.
- A avaliação não é tão simples quanto parece!



3

Introdução

- Por quê?
 - Precisamos de maneiras de prever limites de desempenho na prática, com base em experimentos com todos os dados que podem ser obtidos.



Introdução

- Comparar o desempenho de diferentes algoritmos de aprendizado de máquina sobre um determinado problema é outro assunto que não é tão fácil quanto parece!
- Para ter certeza que diferenças aparentes não são causados por efeitos aleatórios, os testes estatísticos são também necessários.



Introdução

- Treinamento x Teste
- Podemos assumir que o que está sendo previsto é a capacidade de classificar instâncias de teste com precisão.
- Contudo, algumas situações envolvem a predição de probabilidades das classes ao invés das próprias classes, e outros envolvem predição numérica (regressão), em vez de valores nominais.
- Para cada caso, diferentes métodos são necessários.



Métricas de erro

- Em AM supervisionado, a avaliação é realizada por meio da análise de desempenho do preditor na rotulção de novos objetos.
- Para classificação:
 - Taxa de erro

$$err(\hat{f}) = \frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{f}(x_i))$$

- Taxa de acerto

$$ac(\hat{f}) = 1 - err(\hat{f})$$



Métricas de erro

- Para regressão:
 - O erro da hipótese \hat{f} pode ser calculado pela distância entre o valor y_i conhecido e aquele predito pelo modelo
 - Erro quadrático médio

$$MSE(\hat{f}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$$

- Distância absoluta média

$$MAD(\hat{f}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{f}(x_i)|$$



Amostragem

- Métodos para estimar o erro verdadeiro de um classificador
 - Resubstituição (Resubstitution)
 - Validação simples (Holdout)
 - Amostragem aleatória (Random subsampling)
 - Validação cruzada (r-fold cross-validation)
 - Validação cruzada estratificada (r-fold stratified cross-validation)
 - Leave-one-out
 - Bootstrap



Amostragem

- Métodos para estimar o erro verdadeiro de um classificador
 - Resubstituição (Resubstitution)
 - Validação simples (Holdout)
 - Amostragem aleatória (Random subsampling)
 - Validação cruzada (r-fold cross-validation)
 - Validação cruzada estratificada (r-fold stratified cross-validation)
 - Leave-one-out
 - Bootstrap



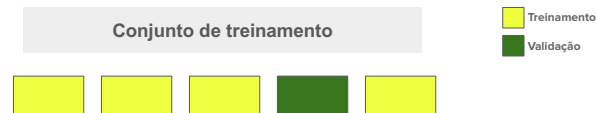
Validação cruzada (R-fold cross-validation)

- Os n exemplos são aleatoriamente divididos em R partições (folds) de tamanho aproximadamente igual (n/R)
- Os exemplos de $(R-1)$ folds são independentemente usados no treinamento e os classificadores obtidos são validados com o fold remanescente



Validação cruzada (R-fold cross-validation)

- O processo é repetido r vezes, e a cada repetição um fold diferente é usado para teste.
- O erro do cross-validation é a média dos erros dos r folds



Validação cruzada (B-fold cross-validation)

- O processo é repetido r vezes, e a cada repetição um fold diferente é usado para teste.
- O erro do cross-validation é a média dos erros dos r folds



Validação cruzada (B-fold cross-validation)

- O processo é repetido r vezes, e a cada repetição um fold diferente é usado para teste.
- O erro do cross-validation é a média dos erros dos r folds



Validação cruzada (B-fold cross-validation)

- O processo é repetido r vezes, e a cada repetição um fold diferente é usado para teste.
- O erro do cross-validation é a média dos erros dos r folds



Validação cruzada (B-fold cross-validation)

- É similar ao cross-validation, mas no processo de geração dos folds a distribuição das classes no conjunto de exemplos é levada em consideração durante a amostragem
- Por exemplo, se o conjunto de exemplos tiver duas classes com uma distribuição de 80% para uma classe e 20% para outra, cada fold também terá essa proporção

Problemas de duas classes

- Em problemas com apenas duas classes denota-se 1 classe como positiva (+) e a outra como negativa (-)
- Matriz de confusão para duas classes:

		Classe Predita	
		+	-
Classe Verdadeira	+	VP	FN
	-	FP	VN

$$n = VP + VN + FP + FN$$

Medidas de desempenho

- Taxa de acerto ou acurácia total

$$ac(\hat{f}) = \frac{VP + VN}{n}$$

Medidas de desempenho

- Taxa de erro na classe positiva
 - Taxa de falsos negativos

$$err_+(\hat{f}) = \frac{FN}{VP + FN}$$

- Taxa de erro na classe neg
 - Taxa de falsos positivos

$$err_-(\hat{f}) = \frac{FP}{FP + VN}$$

Medidas de desempenho

- Sensibilidade ou revocação
 - Taxa de acerto na classe positiva

$$sens(\hat{f}) = rev(\hat{f}) = \frac{VP}{VP + FN}$$

- Especificidade
 - Taxa de acerto na classe negativa

$$esp(\hat{f}) = \frac{VN}{VN + FP}$$

Medidas de desempenho

- Precisão (Medida de exatidão do modelo)
 - Proporção de positivos classificados corretamente entre todos aqueles preditos como positivos por \hat{f}

$$prec(\hat{f}) = \frac{VP}{VP + FP}$$

- Uma precisão de 1,0 para a classe C, significa que cada item rotulado como pertencente à classe C realmente pertence a essa classe, mas não fornece informação a respeito do número de exemplos da classe C que não foram classificadas corretamente



Medidas de desempenho

- Precisão e Revocação não são discutidas isoladamente
- Medida-F: média harmônica ponderada da precisão e revocação

$$F_m(\hat{f}) = \frac{(w + 1) \times rev(\hat{f}) \times prec(\hat{f})}{rev(\hat{f}) + w \times prec(\hat{f})}$$



Medidas de desempenho

- Atribuindo peso igual a 1, que equivale a dar o mesmo grau de importância à precisão e revocação, tem-se medida-F1:

$$F_1(\hat{f}) = \frac{2 \times rev(\hat{f}) \times prec(\hat{f})}{rev(\hat{f}) + prec(\hat{f})}$$



Análise ROC

- ROC (Receiving Operating Characteristics)
- Permite avaliar classificadores em problemas binários: duas classes
- Gráfico ROC: gráfico bidimensional com eixos X (Taxa de Falsos Positivos) e Y (Taxa de Verdadeiro Positivo)
- O desempenho de um dado classificador equivale a um ponto no gráfico ROC



Análise ROC

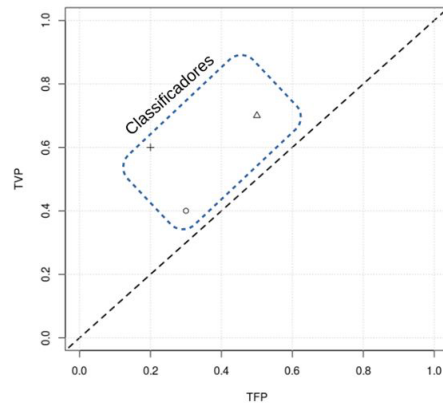
- Alarme falso

$$TFP = err_-(\hat{f}) = \frac{FP}{FP + VN}$$

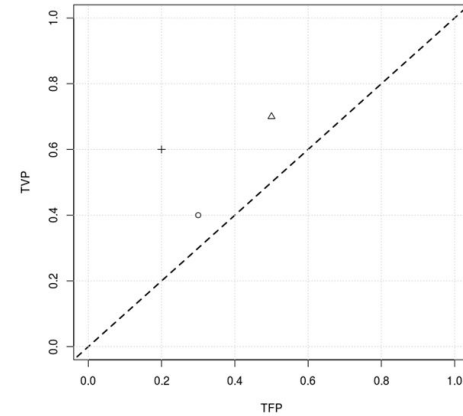
- Revocação

$$TVP = sens(\hat{f}) = rev(\hat{f}) = \frac{VP}{VP + FN}$$

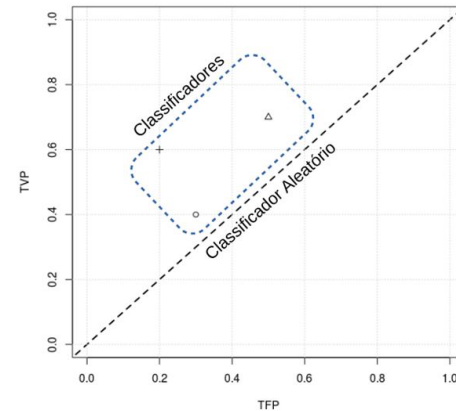
Análise ROC



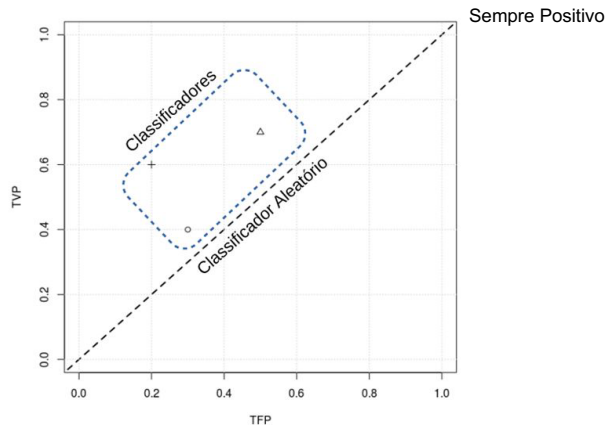
Análise ROC



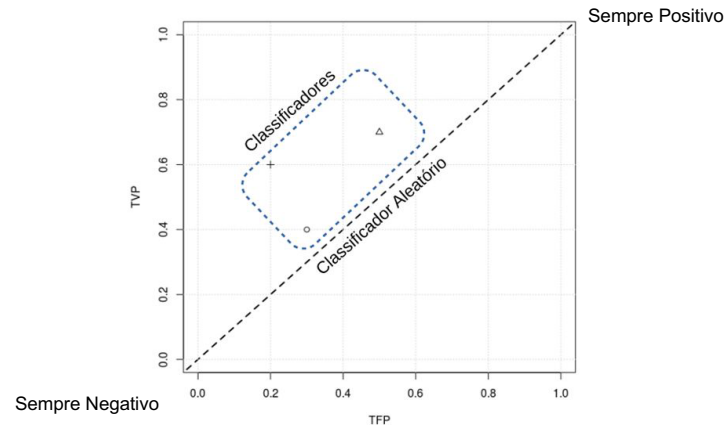
Análise ROC



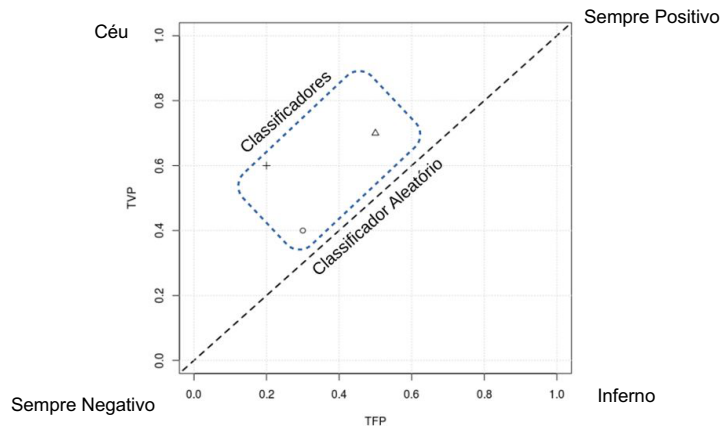
Análise ROC



Análise ROC



Análise ROC



Curva ROC

- Forma mais usual para comparar classificadores
- Vários pontos resultantes do desempenho de classificadores são conectados para formar uma curva
- A curva que mais se aproxima do ponto (0,1) possui o melhor desempenho

AUC

- Quando há interseção entre as curvas, fica difícil estimar quem está mais próximo do ponto (0,1)
- Area Under ROC Curve
- Resultado: [0, 1]
- Quanto mais próximo de 1, melhor
- Recomendação: Validação cruzada → média e desvio padrão



Referências

- Faceli et al., Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina, LTC, 2015.
- Smola, A. and Vishwanathan, S.V.N., Introduction to Machine Learning, Cambridge University Press, 2008.
- Witten et al., Data Mining – Practical Machine Learning Tools and Techniques, 3d edition, Elsevier, 2011.
- J. Han; M. Kamber. Data Mining: Concepts and Techniques, Morgan Kaufmann, 2000.

