


UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO


**Tópicos Especiais em Computação: Aprendizado
de Máquina**
Cap 1: Introdução

Prof. Jefferson Moraes
Email: jmoraes@ufpa.br



Avaliação de Modelos Preditivos

- Não existe técnica de AM universal, que se saia melhor em qualquer tipo de problema
 - Implica na necessidade de **experimentos**
- Características do problema e das técnicas pode auxiliar em alguns casos
 - Ex: modelo deve ser interpretável → técnicas simbólicas, dados possuem alta dimensão → SVM, etc
 - Mesmo assim diversos algoritmos podem ser candidatos



Avaliação Modelos Preditivos

- Mesmo que um único algoritmo seja escolhido
 - Variações de parâmetros produzem diferentes modelos
- → Domínio de AM: necessidade de experimentação
 - Experimentos controlados
 - Procedimentos que garantem a **corretude e reprodutibilidade** dos experimentos

Avaliação Modelos Preditivos

- Diferentes aspectos podem ser considerados
 - Acurácia do modelo nas previsões
 - Compreensibilidade do conhecimento extraído
 - Tempo de aprendizado
 - Requisitos de armazenamento
 - Etc

Concentraremos discussões a medidas de **desempenho preditivo**



Métricas de Erro

- Desempenho na rotulação de objetos
 - Métricas para classificação
 - Taxa de erro
 - Acurácia
 - Métricas para regressão
 - Erro quadrático médio
 - Distância absoluta média

Métricas para classificação

Taxa de erro de um classificador f

- De classificações incorretas

$$\text{err}(f) = (1/n) \sum_{i=1 \dots n} I(y_i \neq f(\mathbf{x}_i))$$

- Proporção de exemplos classificados incorretamente em um conjunto com n objetos
 - Comparação da classe conhecida com a predita
 - I é função identidade
 - $= 1$ se argumento é verdadeiro e 0 em caso contrário
 - Varia entre 0 e 1 e valores próximos de 0 são melhores

Métricas para classificação

Taxa de erro de um classificador f

- De classificações incorretas

$$\text{err}(f) = (1/n) \sum_{i=1 \dots n} I(y_i \neq f(\mathbf{x}_i))$$

- Proporção de exemplos classificados incorretamente em um conjunto com n objetos
 - Comparação da classe conhecida com a predita
 - I é função identidade
 - $= 1$ se argumento é verdadeiro e 0 em caso contrário
 - Varia entre 0 e 1 e valores próximos de 0 são melhores

Métricas para classificação

- **Taxa de acerto** ou **acurácia** de um classificador f
- Complemento da taxa de erro

$$\text{ac}(f) = 1 - \text{err}(f) = (1/n) \sum_{i=1 \dots n} I(y_i = f(x_i))$$

- Proporção de exemplos classificados corretamente em um conjunto com n objetos
 - Varia entre 0 e 1 e valores próximos de 1 são melhores

Métricas para classificação

- Matriz de confusão
 - Alternativa para visualizar desempenho de classificador
 - Predições corretas e incorretas em cada classe

Classe predita							
	c1	c2	c3				
c1	11	1	3				
c2	1	4	0				
c3	2	1	6				
Classe verdadeira							

- Linhas representam **classes verdadeiras**
- Colunas representam **classes preditas**
- Elemento m_j : número de exemplos da classe c_j classificados como pertencentes à classe c_j
- Diagonal da matriz: **acertos** do classificador
- Outros elementos: **erros** cometidos

Métricas para regressão

- **Erro** pode ser calculado pela distância entre o valor conhecido e o valor predito pelo modelo

Erro quadrático médio (*Mean Squared Error – MSE*)

$$\text{MSE}(f) = (1/n) \sum (y_i - f(\mathbf{x}_i))^2$$

Distância absoluta média (*Mean Absolute Distance MDA*)

$$\text{MDA}(f) = (1/n) \sum |y_i - f(\mathbf{x}_i)|$$

MSE e MAD são sempre não negativos;
valores mais baixos correspondem a melhores modelos



Amostragem

- Tem-se usualmente um único conjunto de n objetos
 - Deve ser usado para induzir e avaliar o preditor
 - Desempenho no conjunto de treinamento é otimista
 - Todos os algoritmos tentam de alguma forma melhorar seu desempenho no conjunto de treinamento na fase indutiva
 - Avaliar modelo no conjunto de treinamento é conhecido como resubstituição
 - Produz taxa de acerto/erro aparente

Amostragem

- Métodos de amostragem: obter estimativas de desempenho mais confiáveis
 - Definindo subconjuntos disjuntos

Treinamento

Dados empregados na **indução** e no **ajuste** do modelo

Qualquer ajuste de parâmetros deve ser feito nos **dados de treinamento**

Teste

Simulam a apresentação de **novos exemplos** ao preditor (não vistos em sua indução)

Somente avaliar o modelo obtido

Em algumas situações, dados de treinamento são sub-divididos, gerando conjunto de **validação** dedicado ao ajuste de parâmetros

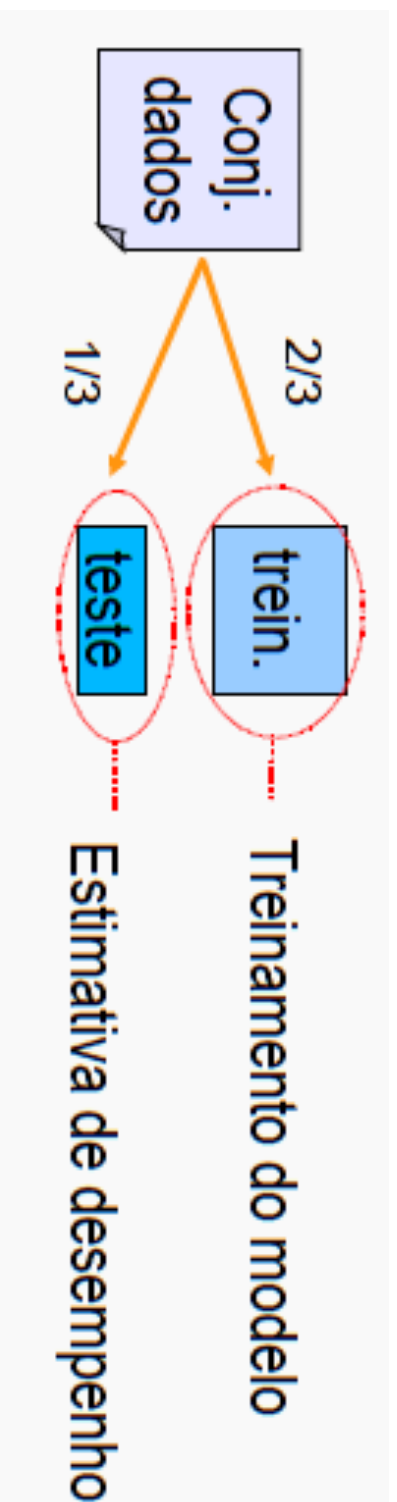


Amostragem

- Principais métodos de amostragem:
 - Holdout
 - Amostragem aleatória
 - Validação cruzada
 - Leave-one-out
 - Bootstrap

Holdout

- Método mais simples:
 - Divide conjunto de dados em proporção p para treinamento e $(p - 1)$ para teste
 - Uma única partição
 - Valores típicos de p : $\frac{1}{2}$, $\frac{2}{3}$ ou $\frac{3}{4}$



Holdout

■ Exemplo

	Objeto	Atributo 1	Atributo 2	Atributo 3	Classe
	1	855	5142	2708	Safra 95
	2	854	23155	2716	Safra 95
	3	885	16586	2670	Safra 95
	4	877	16685	2677	Safra 95
	5	839	5142	2708	Safra 95
	6	854	5005	2685	Safra 95
	7	885	19455	2708	Safra 95
	8	839	5027	2708	Safra 95
	9	877	16823	2677	Safra 95
	10	892	19180	2716	Safra 95
	11	24628	39437	381	Safra 96
	12	43183	39277	328	Safra 96
	13	27871	39712	389	Safra 96
	14	42329	40307	328	Safra 96
	15	41627	40032	335	Safra 96
	16	39399	40322	335	Safra 96
	17	33677	40375	328	Safra 96
	18	33539	40078	335	Safra 96
	19	34150	40353	358	Safra 96
	20	34485	40742	358	Safra 96

Holdout


■ Exemplo

Conjunto de treinamento

Objeto	Atributo 1	Atributo 2	Atributo 3	Classe
4	877	16685	2677	Safra 95
6	854	5005	2685	Safra 95
8	839	5027	2708	Safra 95
2	854	23155	2716	Safra 95
10	892	19180	2716	Safra 95
1	855	5142	2708	Safra 95
6	854	5005	2685	Safra 95
18	33539	40078	335	Safra 96
15	41627	40032	335	Safra 96
19	34150	40353	358	Safra 96
12	43183	39277	328	Safra 96
17	33677	40375	328	Safra 96
20	34485	40742	358	Safra 96
11	24628	39437	381	Safra 96


Conjunto de teste

Objeto	Atributo 1	Atributo 2	Atributo 3	Classe
3	885	16586	2670	Safra 95
5	839	5142	2708	Safra 95
9	877	16823	2677	Safra 95
13	27871	39712	389	Safra 96
14	42329	40307	328	Safra 96
16	39399	40322	335	Safra 96



Holdout

- Indicado para grande quantidade de dados
 - Se pequena quantidade de dados
 - Poucos exemplos são usados no treinamento
 - Modelo pode depender da composição dos conjuntos de treinamento e teste
 - Quanto menor conjunto de treinamento, maior a variância do modelo
 - Quanto menor conjunto de teste, menos confiável a acurácia estimada para ele
- Muito usado para definir subconjuntos de validação

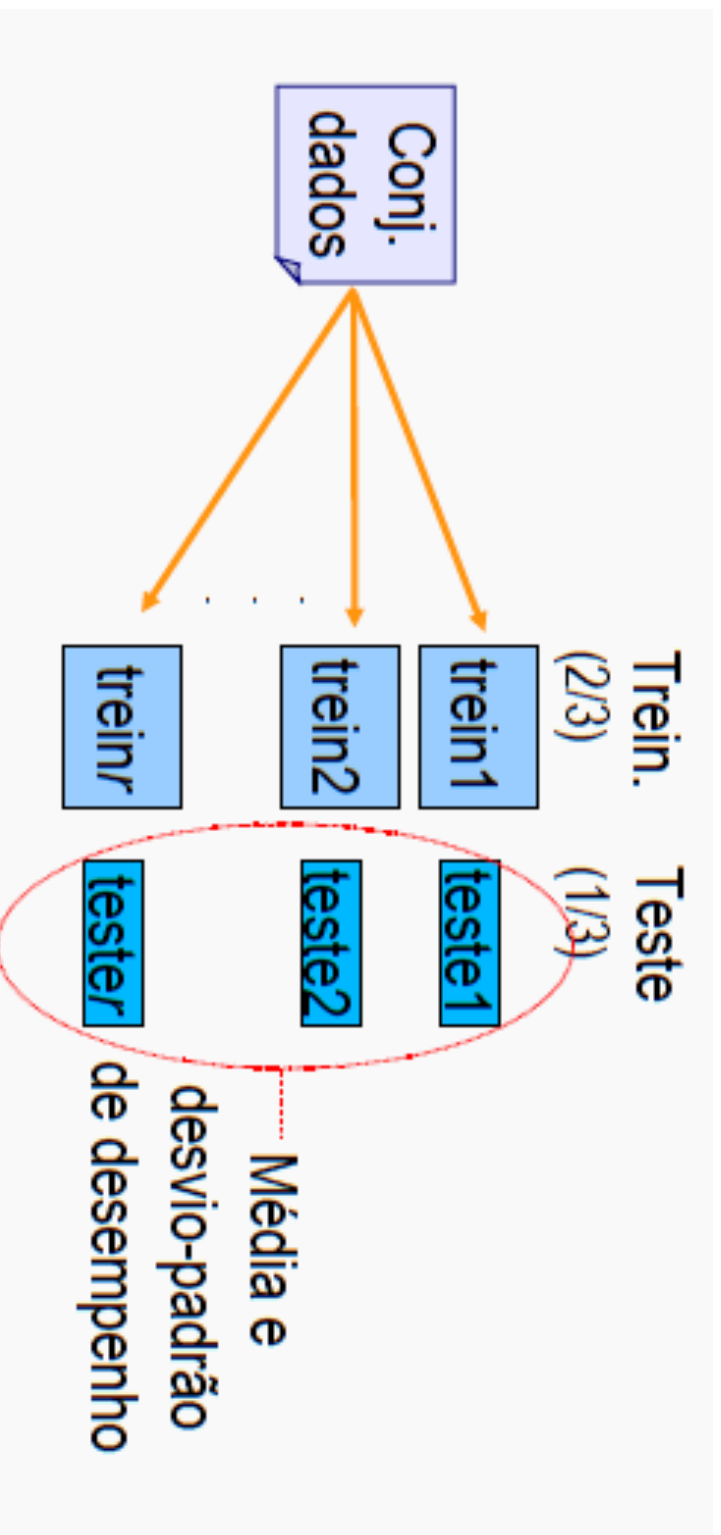


Holdout

- Não avalia o quanto o desempenho de uma técnica varia
 - Quanto a diferentes combinação de exemplos de treinamento
 - É possível que uma divisão deixe no subconjunto de teste exemplos “mais fáceis”
 - Para tornar os resultados menos dependentes da partição feita: vários holdout
 - *Random subsampling* (amostragem aleatória)

Amostragem aleatória

- Repetições de Holdout
 - Há sobreposição entre os conjuntos de teste gerados
 - Fornece uma média de desempenho



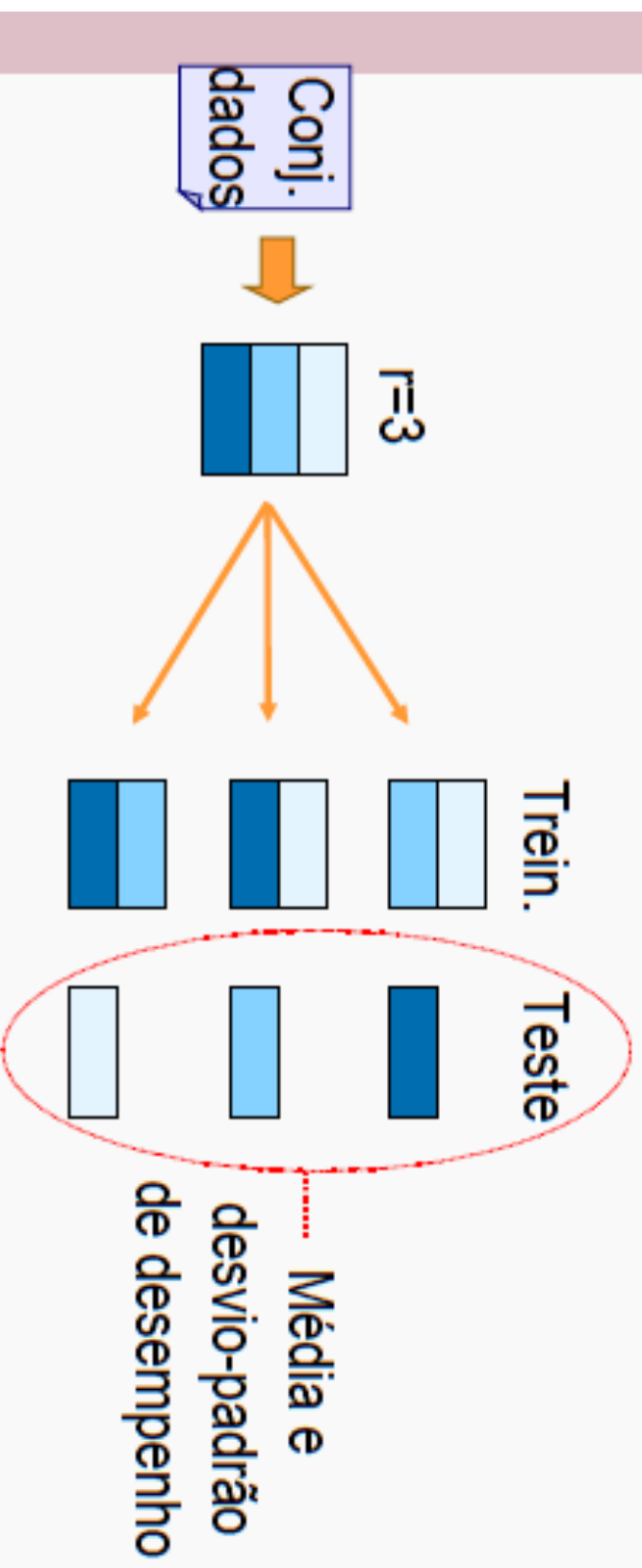
Validação cruzada

Método mais usado: *r-fold cross validation*

- Conjunto é dividido em r partes de tamanho aproximadamente igual
- Objetos de $r-1$ partes são usados no treinamento e a parte restante é usada para teste
- Procedimento é repetido r vezes usando cada partição para teste
 - \Rightarrow subconjuntos de teste são independentes entre si
- Desempenho é dado por média
- Valor típico de r : 10

Validação cruzada

Ilustração para $r = 3$:



Validação cruzada

Variação: *r-fold cross validation* estratificado

- Manter a distribuição de classes em cada partição
 - Ex: se conjunto de dados original tem 20% na classe c_1 e 80% na classe c_2 , cada partição também deve manter essa proporção
- Distribuição de classes: proporção de exemplos em cada classe
 - Para cada classe c_j , $\text{dist}(c_j)$ = número de exemplos que possuem a classe c_j / número total de exemplos

$$\text{dist}(c_j) = \frac{1}{n} \sum_{i=1}^n I(y_i = c_j)$$

Distribuição de classes

Ex.: conjunto de dados com 100 exemplos

- 60 são da classe c_1
- 15 são da classe c_2
- 25 são da classe c_3
- A distribuição de classe é $\text{dist}(c_1, c_2, c_3) = (0,60, 0,15, 0,25) = (60\%, 15\%, 25\%)$
- A classe c_1 é a classe **majoritária** ou prevalente
- A classe c_2 é a classe **minoritária**

Cross-validation estratificado

Exemplo:

- $r = 5$

Objeto	Atributo 1	Atributo 2	Atributo	Classe
4	877	16685	2677	Safra 95
9	877	16823	2677	Safra 95
18	33539	40078	335	Safra 96
11	24628	39437	381	Safra 96
1	855	5142	2708	Safra 95
3	885	16586	2670	Safra 95
14	42329	40307	328	Safra 96
20	34485	40742	358	Safra 96
7	885	19455	2708	Safra 95
10	892	19180	2716	Safra 95
15	41627	40032	335	Safra 96
19	34150	40353	358	Safra 96
6	854	5005	2685	Safra 95
2	854	23155	2716	Safra 95
17	33677	40375	328	Safra 96
12	43183	39277	328	Safra 96
8	839	5027	2708	Safra 95
5	839	5142	2708	Safra 95
15	41627	40032	335	Safra 96
16	39399	40322	335	Safra 96

Cross-validation estratificado

■ Ex.: Iteração 1

Conjunto de treinamento

Objeto	Atributo 1	Atributo 2	Atributo	Classe
1	855	5142	2708	Safra 95
3	885	16586	2670	Safra 95
14	42329	40307	328	Safra 96
20	34485	40742	358	Safra 96
7	885	19455	2708	Safra 95
10	892	19180	2716	Safra 95
15	41627	40032	335	Safra 96
19	34150	40353	358	Safra 96
6	854	5005	2685	Safra 95
2	854	23155	2716	Safra 95
17	33677	40375	328	Safra 96
12	43183	39277	328	Safra 96
8	839	5027	2708	Safra 95
5	839	5142	2708	Safra 95
15	41627	40032	335	Safra 96
16	39399	40322	335	Safra 96

Conjunto de teste

Objeto	Atributo 1	Atributo 2	Atributo	Classe
4	877	16685	2677	Safra 95
9	877	16823	2677	Safra 95
18	33539	40078	335	Safra 96
11	24628	39437	381	Safra 96

Cross-validation estratificado

Ex.: Iteração 2

Conjunto de treinamento

Objeto	Atributo 1	Atributo 2	Atributo	Classe
4	877	16685	2677	Safra 95
9	877	16823	2677	Safra 95
18	33539	40078	335	Safra 96
11	24628	39437	381	Safra 96
7	885	19455	2708	Safra 95
10	892	19180	2716	Safra 95
15	41627	40032	335	Safra 96
19	34150	40353	358	Safra 96
6	854	5005	2685	Safra 95
2	854	23155	2716	Safra 95
17	33677	40375	328	Safra 96
12	43183	39277	328	Safra 96
8	839	5027	2708	Safra 95
5	839	5142	2708	Safra 95
15	41627	40032	335	Safra 96
16	39399	40322	335	Safra 96

Conjunto de teste

Objeto	Atributo 1	Atributo 2	Atributo	Classe
1	855	5142	2708	Safra 95
3	885	16586	2670	Safra 95
14	42329	40307	328	Safra 96
20	34485	40742	358	Safra 96

Cross-validation estratificado

Ex.: Iteração 3

Conjunto de treinamento

Objeto	Atributo 1	Atributo 2	Atributo	Classe
4	877	16685	2677	Safra 95
9	877	16823	2677	Safra 95
18	33539	40078	335	Safra 96
11	24628	39437	381	Safra 96
1	855	5142	2708	Safra 95
3	885	16586	2670	Safra 95
14	42329	40307	328	Safra 96
20	34485	40742	358	Safra 96
6	854	5005	2685	Safra 95
2	854	23155	2716	Safra 95
17	33677	40375	328	Safra 96
12	43183	39277	328	Safra 96
8	839	5027	2708	Safra 95
5	839	5142	2708	Safra 95
15	41627	40032	335	Safra 96
16	39399	40322	335	Safra 96

Conjunto de teste

Objeto	Atributo 1	Atributo 2	Atributo	Classe
7	885	19455	2708	Safra 95
10	892	19180	2716	Safra 95
15	41627	40032	335	Safra 96
19	34150	40353	358	Safra 96

Cross-validation estratificado

Ex.: Iteração 4

Conjunto de treinamento

Objeto	Atributo 1	Atributo 2	Atributo	Classe
4	877	16685	2677	Safra 95
9	877	16823	2677	Safra 95
18	33539	40078	335	Safra 96
11	24628	39437	381	Safra 96
1	855	5142	2708	Safra 95
3	885	16586	2670	Safra 95
14	42329	40307	328	Safra 96
20	34485	40742	358	Safra 96
7	885	19455	2708	Safra 95
10	892	19180	2716	Safra 95
15	41627	40032	335	Safra 96
19	34150	40353	358	Safra 96
8	839	5027	2708	Safra 95
5	839	5142	2708	Safra 95
15	41627	40032	335	Safra 96
16	39399	40322	335	Safra 96

Conjunto de teste

Objeto	Atributo 1	Atributo 2	Atributo	Classe
6	854	5005	2685	Safra 95
2	854	23155	2716	Safra 95
17	33677	40375	328	Safra 96
12	43183	39277	328	Safra 96

Cross-validation estratificado


Ex.: Iteração 5

Conjunto de treinamento

Objeto	Atributo 1	Atributo 2	Atributo	Classe
4	877	16685	2677	Safra 95
9	877	16823	2677	Safra 95
18	33539	40078	335	Safra 96
11	24628	39437	381	Safra 96
1	855	5142	2708	Safra 95
3	885	16586	2670	Safra 95
14	42329	40307	328	Safra 96
20	34485	40742	358	Safra 96
7	885	19455	2708	Safra 95
10	892	19180	2716	Safra 95
15	41627	40032	335	Safra 96
19	34150	40353	358	Safra 96
6	854	5005	2685	Safra 95
2	854	23155	2716	Safra 95
17	33677	40375	328	Safra 96
12	43183	39277	328	Safra 96

Conjunto de teste

Objeto	Atributo 1	Atributo 2	Atributo	Classe
8	839	5027	2708	Safra 95
5	839	5142	2708	Safra 95
15	41627	40032	335	Safra 96
16	39399	40322	335	Safra 96



Leave-one-out

- Caso extremo de *cross-validation* com $r=n$
 - A cada ciclo exatamente um exemplo é separado para testes
 - Os $n-1$ restantes são usados no treinamento
 - Desempenho: soma dos desempenhos calculados para cada exemplo
 - Produz estimativa mais fiel do desempenho preditivo
 - Mas é computacionalmente caro
 - Usado para conjuntos de dados pequenos

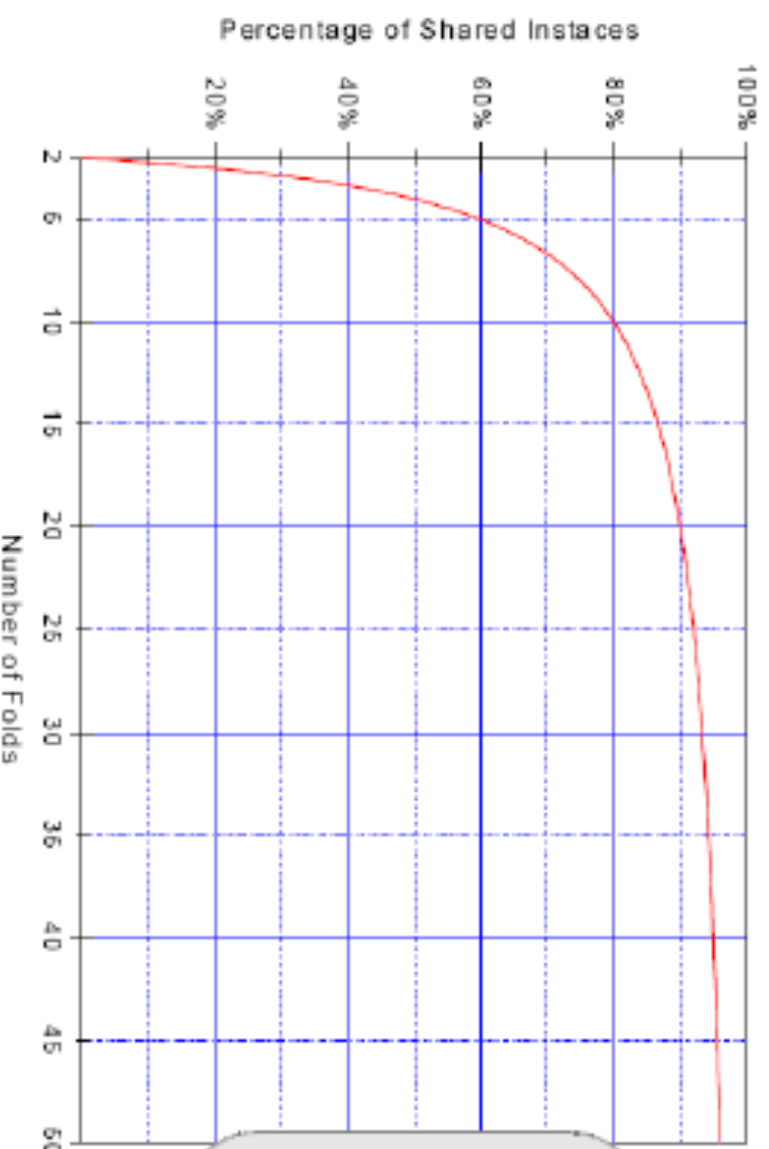


Leave-one-out

- Ex.: no caso do nosso exemplo a amostra tem 20 exemplos, então $r=20$
 - Ou seja, cada conjunto de treinamento será formado por 19 exemplos e o conjunto de teste por um único exemplo
 - Assim, o processo todo será repetido 20 vezes
 - E se o conjunto de dados tivesse 500 exemplos?

Validação cruzada

- Crítica: uma parte dos dados é partilhada entre os subconjuntos de treinamento



Para $r \geq 2$, uma proporção de $(1 - 2/r)$ dos objetos é compartilhada
Ex. $r = 10 \Rightarrow 80\%$ dos objetos são compartilhados

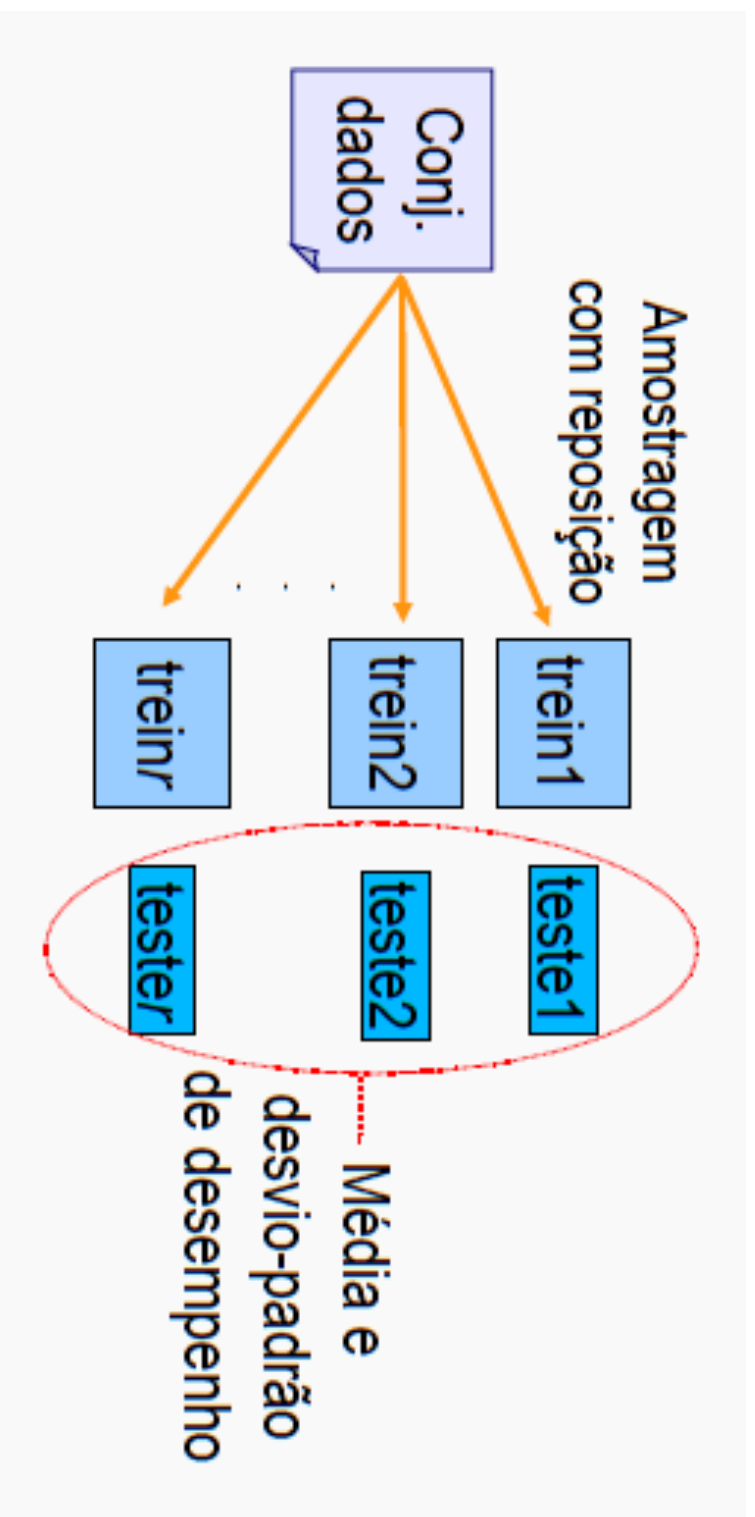


Bootstrap

- Baseado em amostragem com reposição
 - r subconjuntos de treinamento s.o amostrados, com reposição
 - Um exemplo pode estar presente mais de uma vez em um conjunto de treinamento
 - Exemplos não selecionados compõem conjuntos de teste
 - Desempenho: média dos desempenhos nos testes
 - Valor típico para r : 100 ou mais
- É um procedimento custoso aplicado em conjuntos pequenos

Bootstrap

- Ilustração:



Bootstrap

- Há vários estimadores *bootstrap*, mais comum: e_0
 - Cada subconjunto de treinamento tem n exemplos
 - Cada exemplo tem probabilidade $1 - (1 - 1/n)^n$ de ser selecionado ao menos uma vez
 - Para n grande, tende a $1 - 1/e = 0,632$
 - Fração de exemplos não repetidos é de 63,2%
 - Exemplos remanescentes formam subconjunto de teste
 - Desempenho: média das iterações
 - Estimativa estatisticamente equivalente a LOO, com menor variância



Amostragem

- Observações:
 - Para médias de desempenho, é importante reportar também os valores de desvio-padrão
 - Alto desvio padrão → alta variabilidade dos resultados
 - Indicativo de sensibilidade a variações nos dados de treinamento
- Estimativas mais precisas também podem ser obtidas usando intervalos de confiança

Amostragem

Ex. Seja um dos métodos de amostragem

- r-fold CV, por ser o mais utilizado
- Um indutor A gerará r hipóteses h_1, h_2, \dots, h_r
 - E cada hipótese terá uma taxa de erro, medida em no i -ésimo fold
- A média e desvio-padrão do desempenho de A são então dados por (usando medida de erro):

$$med(A) = \frac{1}{r} \sum_{i=1}^r er(h_i)$$

$$dp(A) = \sqrt{\left[\frac{1}{r-1} \sum_{i=1}^r (er(h_i) - med(A))^2 \right]}$$

Amostragem

Exemplo:

- Em 10-fold CV, algoritmo A obteve os erros:
 - (5,5; 11,40; 12,70; 5,20; 5,90; 11,30; 10,90; 11,20; 4,90; 11,00)
 - Temos então:

$$med(A) = \frac{90}{10} = 9 \quad dp(A) = \sqrt{\frac{1}{9} 90,30} = 3,17$$

Classificação binária

- Seja um problema com duas classes: + e -
 - Matriz de confusão:

Classe predita			
Classe verdadeira	+	VP	FN
	-	FP	VN

	+	-
+	VP	FN
-	FP	VN

VP: verdadeiros positivos

Número de exemplos da classe + classificados corretamente

VN: verdadeiros negativos

Número de exemplos da classe - classificados corretamente

FP: falsos positivos

Número de exemplos da classe - classificados incorretamente como +

FN: falsos negativos

Número de exemplos da classe + classificados incorretamente como -

Medidas de desempenho

- Outras medidas calculadas por matriz de confusão:

Taxa de erro na classe + (taxa de falsos negativos):

Proporção de exemplos da classe + incorretamente classificados

$$\text{err}+(f) = \frac{FN}{VP+FN}$$

VP	FN
FP	VN

Taxa de erro na classe - (taxa de falsos positivos):

Proporção de exemplos da classe - incorretamente classificados

$$\text{err}-(f) = \frac{FP}{VN+FP}$$

VP	FN
FP	VN

Medidas de desempenho

- Outras medidas calculadas por matriz de confusão:

Taxa de erro total:

Soma da diagonal secundária da matriz / n

$$\text{err}(f) = \frac{\text{FP} + \text{FN}}{n}$$



VP	FN
FP	VN

Taxa de acerto ou acurácia total:

Soma da diagonal principal / n

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



VP	FN
FP	VN

Medidas de desempenho

Precisão:

Proporção de exemplos + classificados corretamente entre os preditos como +

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

VP	FN
FP	VN

Sensibilidade ou revocação:

Taxa de acerto na classe positiva (taxa de verdadeiros positivos)

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

VP	FN
FP	VN

Medidas de desempenho

Especificidade:

Taxa de acerto na classe -. Seu complemento é a taxa de falsos positivos.

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

VP	FN
FP	VN

Medidas de desempenho

Ex.: avaliação de três classificadores

Classe verdadeira

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

Classificador 1
TVP = 0.4
TFP = 0.3

Classe verdadeira

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

Classificador 2
TVP = 0.7
TFP = 0.5

Classe verdadeira

Classe predita			
		p	n
		P	N
Classe verdadeira	P	60	40
	N	20	80

Classificador 3
TVP = 0.6
TFP = 0.2

Medidas de desempenho

Seja um classificador com a seguinte matriz de confusão, definir:

- Acurácia
- Precisão
- Revocação
- Especificidade

		Classe predita	
		p	n
Classe verdadeira	P	70	30
	n	40	60

Exemplo

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

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

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

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

		Verdadeiro	
		p	n
Predito	p	VP	FN
	n	FP	VN
	p	70	30
	n	40	60

Exemplo

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

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

$$\text{Revocação} = \frac{VP}{VP + FN} = 7 / (70 + 30) = 0.70$$

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

		Verdadeiro	
		P	n
Predito	P	VP	FN
	n	FP	VN
	P	70	30
	n	40	60

Generalizando para mais classes

- Para mais que duas classes:
 - Considera cada uma + e as demais -
 - Ex. C1:

	C1	C2	C3
C1	TP	FN	FN
C2	FP	TN	TN
C3	FP	TN	TN

	C1	C2	C3
C1	49	1	0
C2	0	47	3
C3	0	2	48

C1		
	+	-
+	TP	FN
-	FP	TN

C1		
	+	-
+	49	1
-	0	100

$$\text{prec}(C1) = 1$$

$$\text{rev}(C1) = 0,98$$

Generalizando para mais classes

- Para mais que duas classes:
- Ex. C2:

	C1	C2	C3
C1		TN	FP
C2	FN		TP
C3	TN	FP	

	C1	C2	C3
C1		49	1
C2	0		47
C3	0	2	

	C2	
	+	-
+	TP	FN
-	FP	TN

	C2	
	+	-
+	47	3
-	3	97

$$\text{prec}(C2) = 0,94$$

$$\text{rev}(C2) = 0,94$$

Generalizando para mais classes

- Para mais que duas classes:
- Ex. C2:

	C1	C2	C3
C1	TN	FP	TN
C2	FN	TP	FN
C3	TN	FP	TN

	C1	C2	C3
C1	49	1	0
C2	0	47	3
C3	0	2	48

	C2	
+	+	-
+	TP	FN
-	FP	TN

	C2	
+	+	-
+	47	3
-	3	97

$$\text{prec}(C2) = 0,94$$

$$\text{rev}(C2) = 0,94$$

Generalizando para mais classes

Para mais que duas classes:

- Ex. C3:

	c1	c2	c3
c1	TN	TN	FP
c2	TN	TN	FP
c3	FN	FN	TP

	c1	c2	c3
c1	49	1	0
c2	0	47	3
c3	0	2	48

c3		
	+	-
+	TP	FN
-	FP	TN

c3		
	+	-
+	48	2
-	3	97

$$\begin{aligned}\text{prec}(C3) &= 0,94 \\ \text{rev}(C3) &= 0,96\end{aligned}$$

Precisão vs revocação

Precisão e revocação costumam ser discutidas em conjunto, combinadas em uma medida F:

$$F(f) = \frac{(w + 1) \text{rev}(f) \text{prec}(f)}{\text{rev}(f) + w \text{prec}(f)}$$

- Média harmônica da previsão e revocação
- Usando $w = 1 \Rightarrow$ mesmo grau de importância para duas medidas $\Rightarrow F1$

$$F1(f) = \frac{2 \text{rev}(f) \text{prec}(f)}{\text{rev}(f) + \text{prec}(f)}$$

Exemplos de Resultado no Weka

■ Data set: Soybean

=== Detailed Accuracy By Class ===							
TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class	
0.6	0.012	0.6	0.6	0.6	0.992	diaporthe-stem-canker	
1	0	1	1	1	1	charcoal-rot	
1	0	1	1	1	1	rhizoctonia-root-rot	
1	0.007	0.957	1	0.978	0.995	phytophthora-rot	
1	0	1	1	1	1	brown-stem-rot	
1	0	1	1	1	1	powdery-mildew	
1	0	1	1	1	1	downy-mildew	
0.913	0.007	0.955	0.913	0.933	0.999	brown-spot	
1	0	1	1	1	1	bacterial-blight	
1	0	1	1	1	1	bacterial-pustule	
1	0	1	1	1	1	purple-seed-stain	
0.727	0.013	0.8	0.727	0.762	0.861	anthracnose	
1	0.012	0.714	1	0.833	0.999	phylosticta-leaf-spot	
0.739	0.02	0.85	0.739	0.791	0.991	alternaria-leaf-spot	
0.826	0.041	0.76	0.826	0.792	0.988	frog-eye-leaf-spot	
1	0	1	1	1	1	diaporthe-pod-&-stem-blight	
1	0	1	1	1	1	cyst-nematode	
0.25	0	1	0.25	0.4	0.996	2-4-d-injury	
1	0.018	0.4	1	0.571	1	herbicide-injury	
0.883	0.012	0.896	0.883	0.881	0.987	Weighted Avg.	



Análise ROC

- Receiver Operating Characteristic
- É uma técnica para visualizar, avaliar, organizar e selecionar classificadores baseado em suas performances
- Para realizar estas análises, gráficos ROC podem mostrar o limiar entre taxas de acertos e alarmes falsos (taxa de erros) dos classificadores
- Gráfico bidimensional plotado em espaço ROC
 - Eixo x = taxa de falsos positivos (TFP)
 - Eixo y = taxa de verdadeiros positivos (TVP)
- Desempenho de um classificador pode ser plotado nessa curva
 - Equivale a um ponto no espaço ROC

Análise ROC

Ex.: colocar no gráfico ROC os 3 classificadores do exemplo anterior

Classificador 1
TFP = 0.3
TVP = 0.4

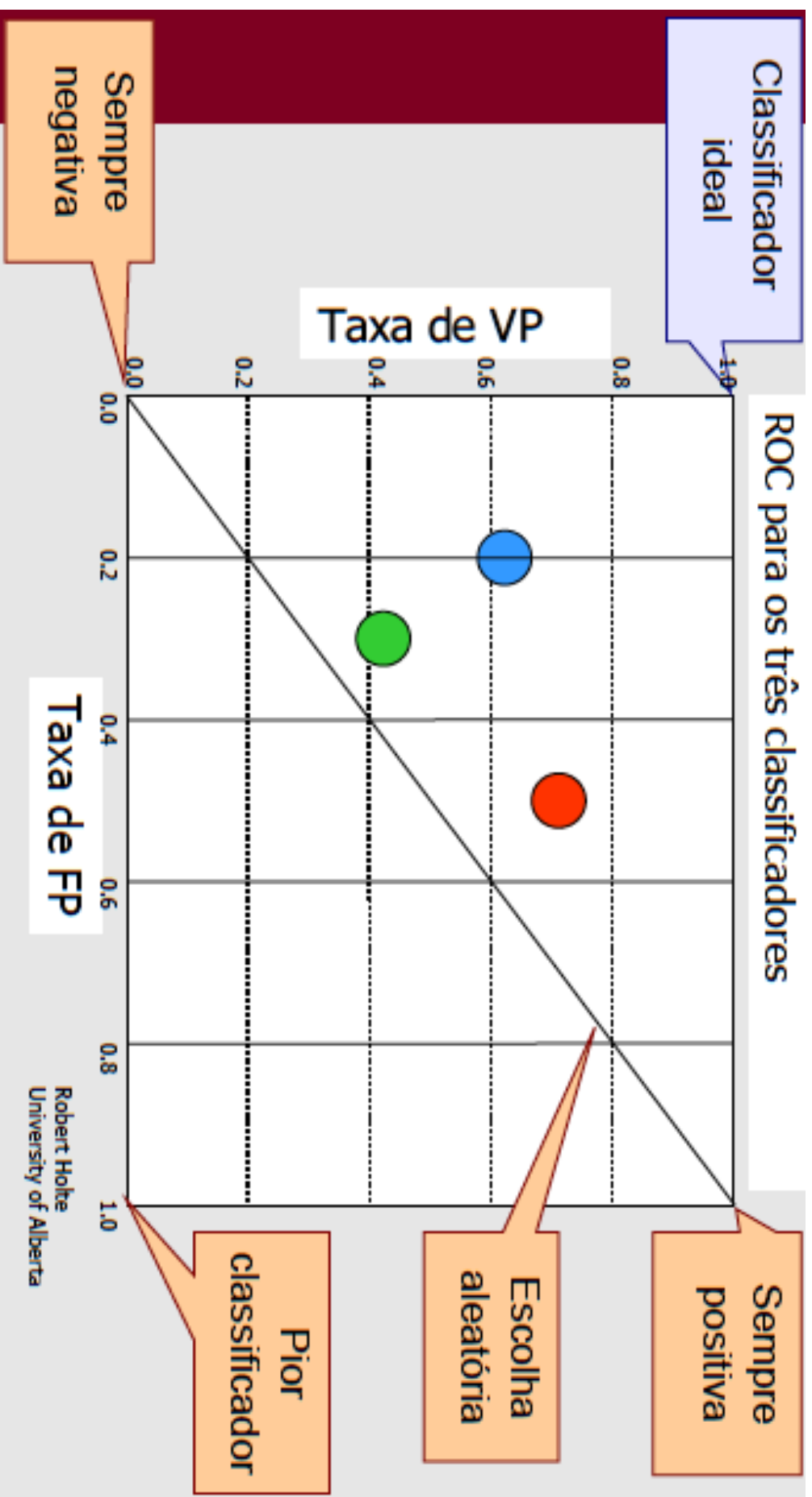
Classificador 2
TFP = 0.5
TVP = 0.7

Classificador 3
TFP = 0.2
TVP = 0.6



Um classificador é considerado melhor que outro se seu ponto no espaço ROC encontra-se acima e à esquerda do ponto correspondente ao segundo classificador

Análise ROC





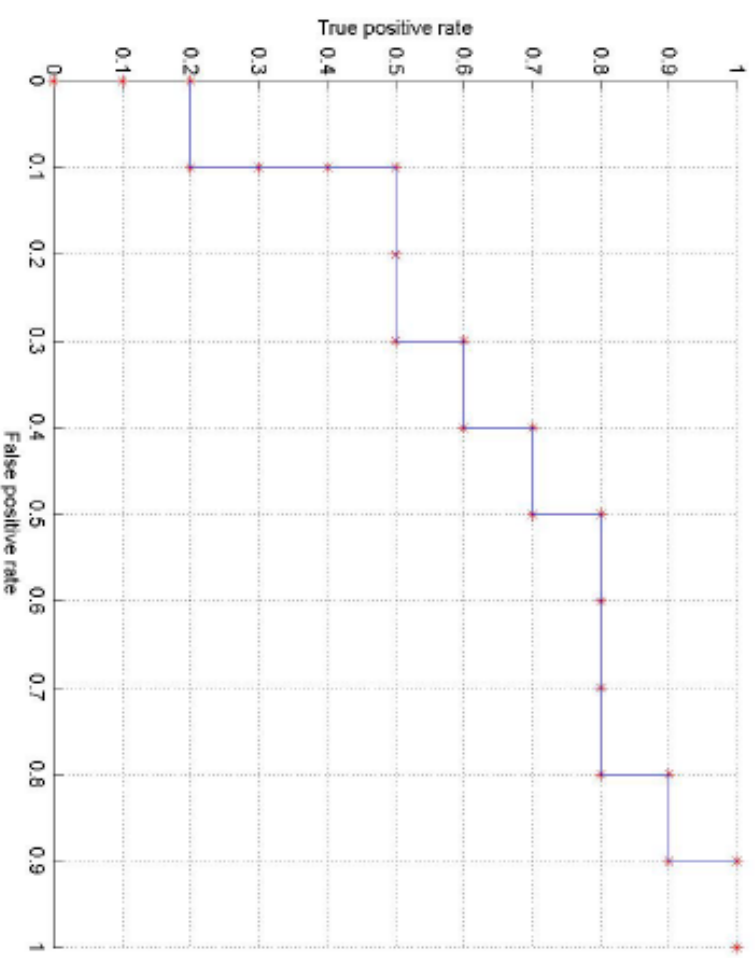
Curvas no espaço ROC

- Muitos classificadores, tais como árvores de decisão ou conjunto de regras, são desenvolvidos para produzir uma única classe, i. e., + ou -, para cada instância.
- Portanto, um classificador discreto gera apenas um ponto no espaço ROC
- Muitos classificadores produzem medidas que permitem ranquear respostas
 - NB ou RNA, produzem uma probabilidade ou valor para cada instância que representa o grau de representatividade da classe sobre aquela instância

Curvas no espaço ROC

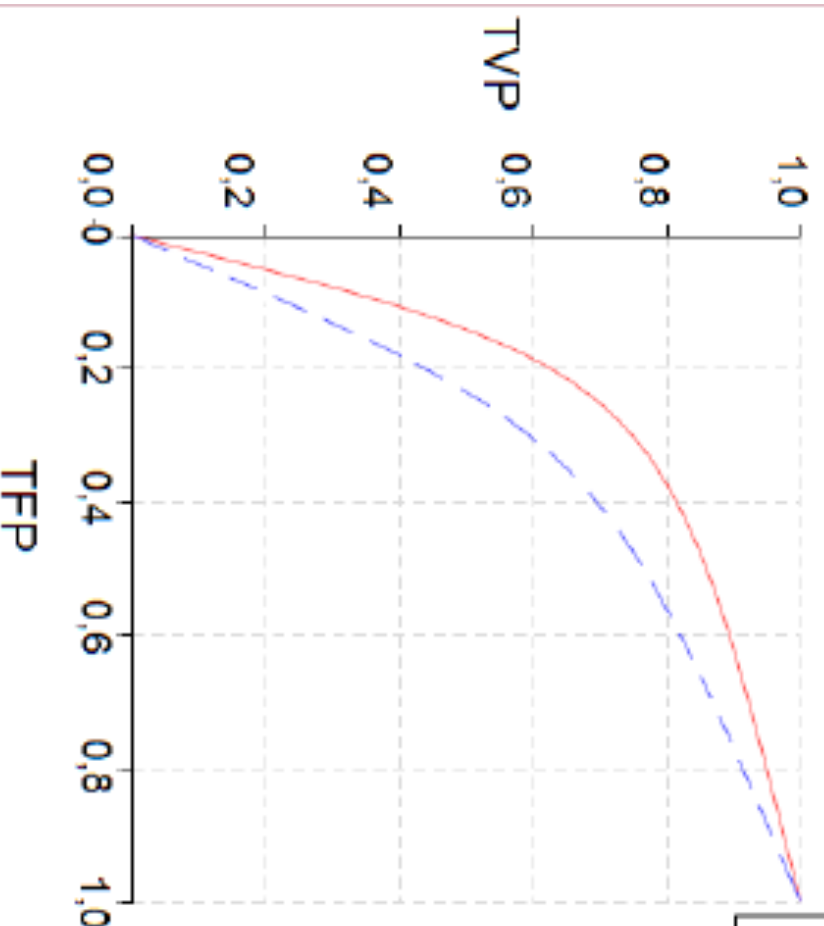
- Exemplo de uma curva ROC produzida a partir de 20 instâncias descritas na Tabela

Inst	Classe	Valor	Inst	Classe	Valor
1	P	0.9	11	P	0.4
2	P	0.8	12	n	0.39
3	n	0.7	13	P	0.38
4	P	0.6	14	n	0.37
5	P	0.55	15	n	0.36
6	P	0.54	16	n	0.35
7	n	0.53	17	P	0.34
8	n	0.52	18	n	0.33
9	P	0.51	19	P	0.30
10	n	0.505	20	n	0.1



Curvas no espaço ROC

▪ Exemplo:



— Classificador A
- - - Classificador B

Comparando duas curvas:

- **Sem intersecção:** a que se mais se aproxima do ponto (0,1) é de melhor desempenho
- **Com intersecções:** cada algoritmo tem uma região em que é melhor que o outro

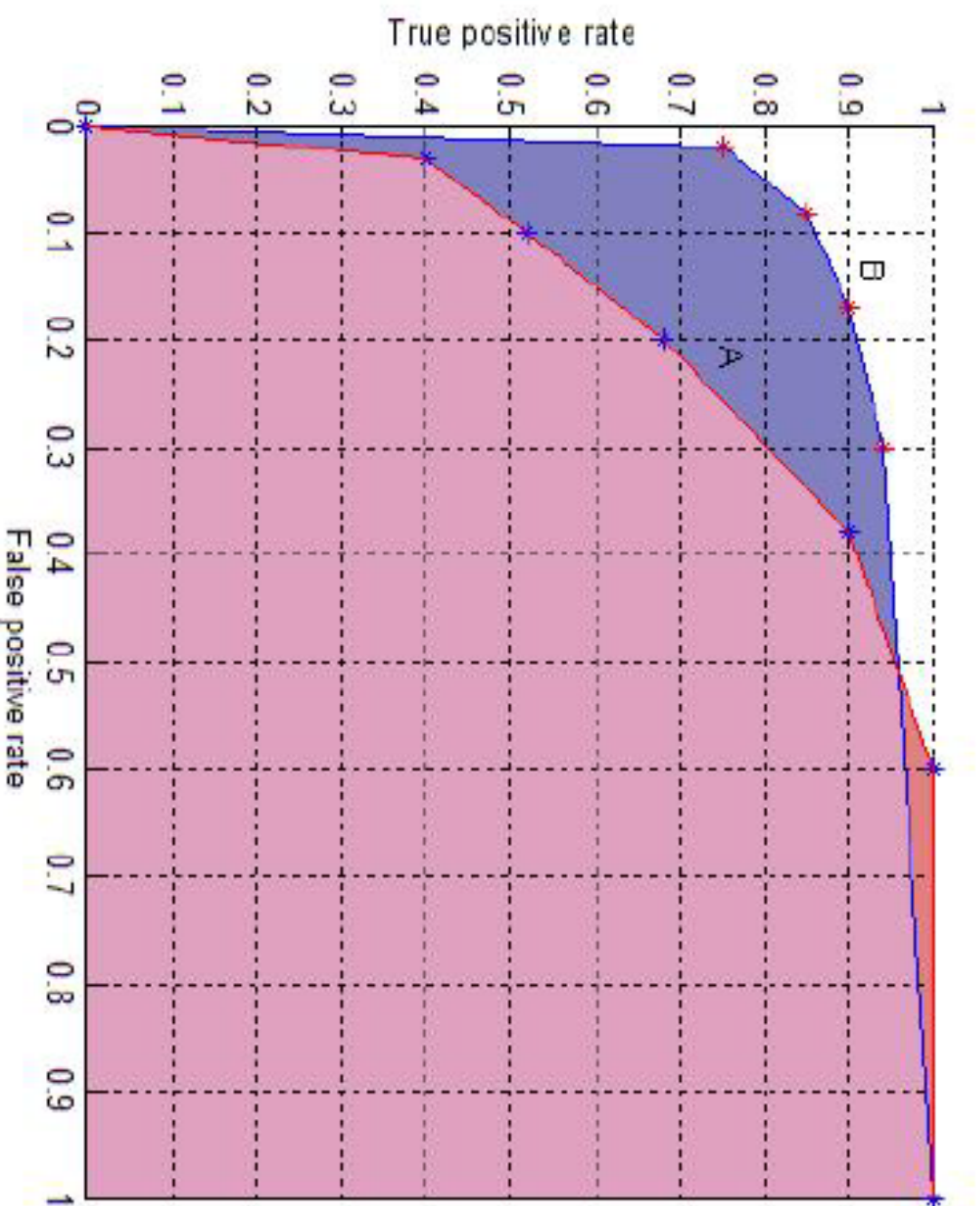


Área abaixo da curva ROC

- Para comparar classificadores é preciso reduzir a curva ROC a um valor escalar
- Método comum: calcular a área abaixo da curva ROC (*Area Under ROC curve* –UAC)
- Produz valores no intervalo [0,1]
 - Valores mais próximos de 1 são considerados melhores
- É mais confiável usar média de AUCs
 - Em validação cruzada

Área abaixo da curva ROC

■ Exemplo



Análise ROC

- Vantagens:
 - Realizar medidas de desempenho independentes do limiar de classificação e de custos associados às classificações incorretas e distribuição das classes
 - Uso de diferentes limiares representa maior ou menor ênfase à classe positiva
 - Permite lidar com desbalanceamento e diferentes custos de classificação

Taxa de erro/acerto é bastante sensível a desbalanceamentos (ex. Conjunto com 90 + e 10 -, taxa de acerto de 0,90 não necessariamente indica bom desempenho preditivo)



Análise ROC

- Desvantagens:
 - Análise originalmente limitada a classificação binária