

**UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA**

**PTC 3467
Análise de Desempenho de
modelos**

**Pós-Graduação em Engenharia Elétrica
2o. Semestre de 2020**

Agenda

- 1. Modelos Preditivos;**
- 2. Avaliação de Modelos Preditivos;**
- 3. Medidas de Desempenho Classificação Binárias**
- 4. Resumo;**
- 5. Referências.**

Modelos preditivos

- **Algoritmo de Análise preditivo**: função que, dado um conjunto de exemplos rotulados, constrói um estimador

Classificação

- Rótulos nominais (conjunto discreto e não ordenado de valores)

Ex. {doente, saudável}, {bom pagador, mau pagador}, {iris setosa, iris versicolor, iris virginica}

- Estimador é chamado

classificador

Regressão

- Rótulos contínuos (conjunto infinito ordenado de valores)

Ex. peso, temperatura, vazão de água

- Estimador é chamado

regressor

Estimadores podem ser vistos como funções

Modelos preditivos

- Exemplos de conjuntos de dados:

Conjunto de dados iris

TamP	LargP	TamS	LargS	Espécie
5,1	3,5	1,4	0,2	Setosa
4,9	3	1,4	0,2	Setosa
7	3,2	4,7	1,4	Versicolor
6,4	3,2	4,5	1,5	Versicolor
6,3	3,3	6	2,5	Virgínica
5,8	2,7	5,1	1,9	Virgínica

Classificação

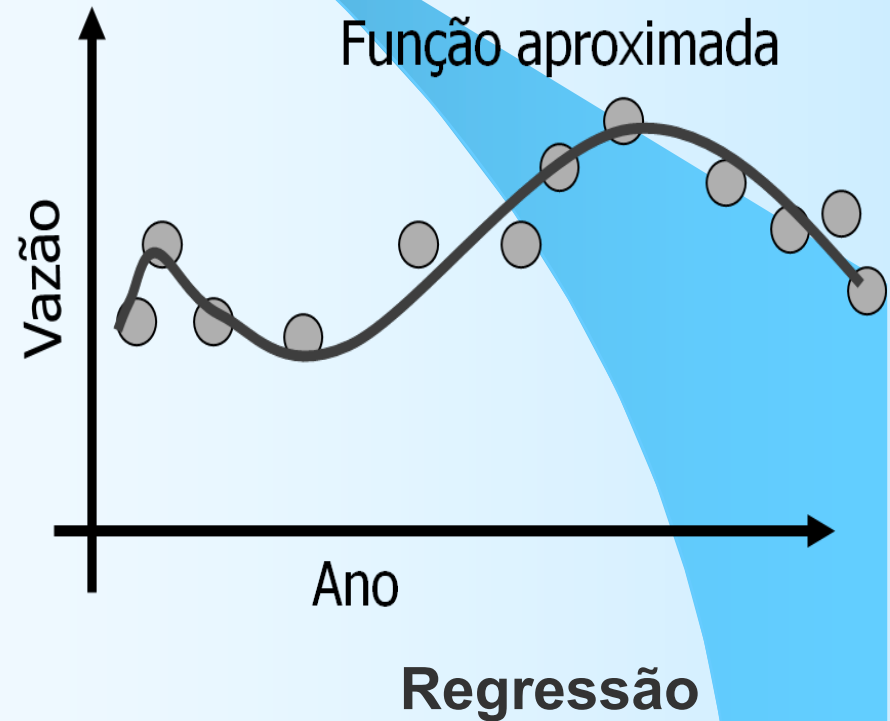
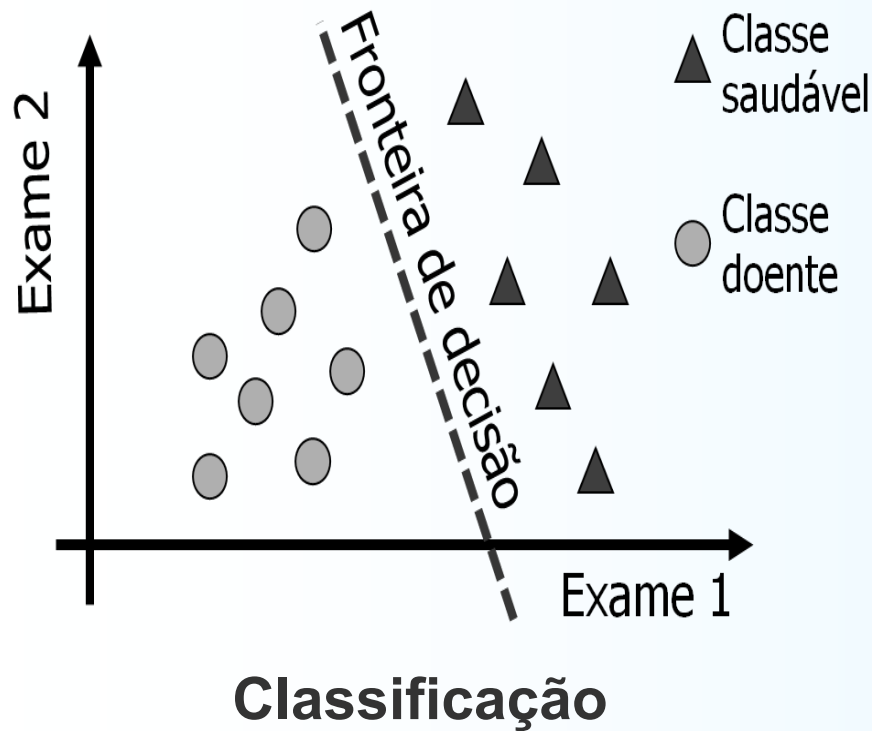
Conjunto de dados swiss

Fertilidade	Agricultura	Educação	Renda	Mortalidade
80,2	17	12	9,9	22,2
83,1	45,1	9	84,8	22,2
92,5	39,7	5	93,4	20,2
85,8	36,5	7	33,7	20,3
76,9	43,5	15	5,2	20,6

Regressão

Modelos preditivos

- Ilustração das tarefas:



Modelos preditivos: classificação

- **Classificação:**

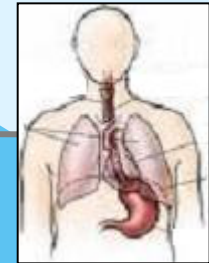
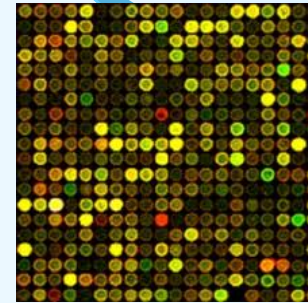
- **Meta:** encontrar fronteira de decisão que separe classes
- Diferentes algoritmos de Análise podem encontrar diferentes fronteiras
- Mesmo algoritmo pode também encontrar fronteiras diferentes
 - Diferenças nos dados de treinamento
 - Variações na ordem de apresentação dos exemplos
 - Processos estocásticos internos

Classificação

- Exemplos:

- Diagnóstico de doenças

- Paciente é doente ou não?
- Qual a sua doença?



- Distribuição geográfica de espécies

- Espécie está presente na região?



Classificação

- Técnicas de AM:
 - Árvores de Decisão
 - Conjuntos de regras
 - Redes Neurais Artificiais
 - Máquinas de Vetores de Suporte
 - K-vizinhos mais próximos
 - Redes Bayesianas
 - Etc.

Modelos preditivos: regressão

- **Regressão:**

- **Meta:** aprender função (curva aproximada) que relacione entradas a valores contínuos de saídas
- Também há diferentes algoritmos de AM para definir essas curvas

- **Exemplos:**

- Prever valor de mercado de um imóvel
- Prever o lucro de um empréstimo bancário

Regressão

- Outros exemplos:
 - Previsão de séries temporais
 - Fusão de sensores
 - Previsão em bolsa de valores
 - Etc



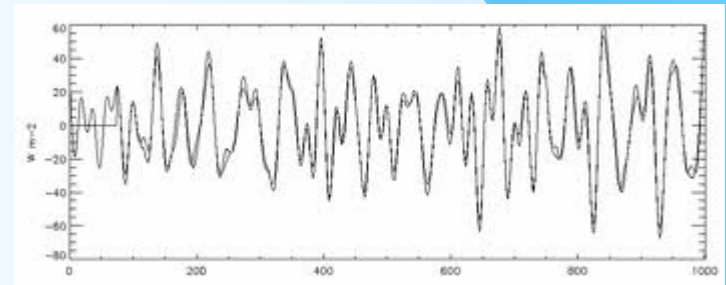
Previsão de Séries Temporais

- Tarefa

- Dado um conjunto de exemplos $\{y(t_1), y(t_2), \dots, y(t_n)\}$, prever a saída $y(t_{n+1})$ no instante de tempo t_{n+1}
- Caso especial de regressão

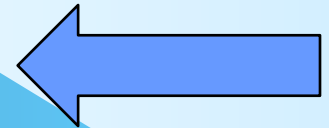
- Exemplos

- Previsão de tempo
- Preço de ações na bolsa
- Desgaste de peças
- Taxas de câmbio
- Etc.



Agenda

1. Modelos Preditivos;
2. Avaliação de Modelos Preditivos;
3. Amostragem
4. Medidas de Desempenho para a Classificação Binária;
5. Resumo;
6. Referências.



Avaliação Modelos Preditivos

- Não existe técnica de Análise 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 \Rightarrow técnicas simbólicas, dados possuem alta dimensão \Rightarrow SVM, etc.
 - Mesmo assim diversos algoritmos podem ser candidatos

Avaliação Modelos Preditivos

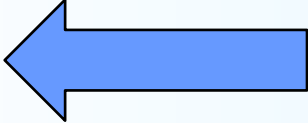
- Mesmo que um único algoritmo seja escolhido
 - Variações de parâmetros produzem diferentes modelos
- ⇒ Domínio de AM: necessidade de **experimentação**
 - Experimentos controlados
 - Procedimentos que garantem a **corretude** e **reprodutibilidade** dos experimentos

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****

Agenda

1. Modelos Preditivos;
2. Avaliação de Modelos Preditivos;
3. Amostragem 
4. Medidas de Desempenho para a Classificação Binária;
5. Resumo;
6. Referências.

Amostragem

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

Treinamento

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

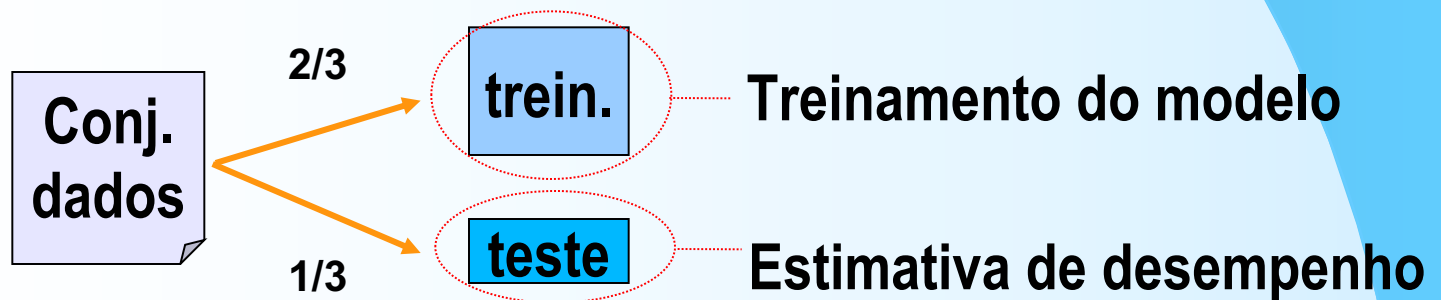
Teste

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

Somente avaliar o modelo obtido

Holdout

- Método mais simples:
 - Divide conjunto de dados em proporção p para treinamento e $(1-p)$ 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

Agenda

1. Modelos Preditivos;
2. Avaliação de Modelos Preditivos;
3. Amostragem
4. Medidas de Desempenho para a Classificação
5. Resumo;
6. Referências.

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

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

- 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_{ij}* : número de exemplos da classe c_i 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*)

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

Distância absoluta média (*Mean Absolute Distance MDA*)

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

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

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

- Ex.: avaliação de três classificadores

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

Exemplo

- 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

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

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

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

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

		Predito	
		p	n
Verdadeiro	p	VP	FN
	n	FP	VN
		p	n
p	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	n
p	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

$$\text{erro (+)} = \frac{\text{FN}}{\text{TP} + \text{FN}} = 0.02$$

$$\text{erro (-)} = \frac{\text{FP}}{\text{FP} + \text{TN}} = 0.00$$

ROC

- Forma de avaliar classificadores binários
- Curvas ROC (*Receiving Operating Characteristics*)
- 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

ROC

- Ex.: 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

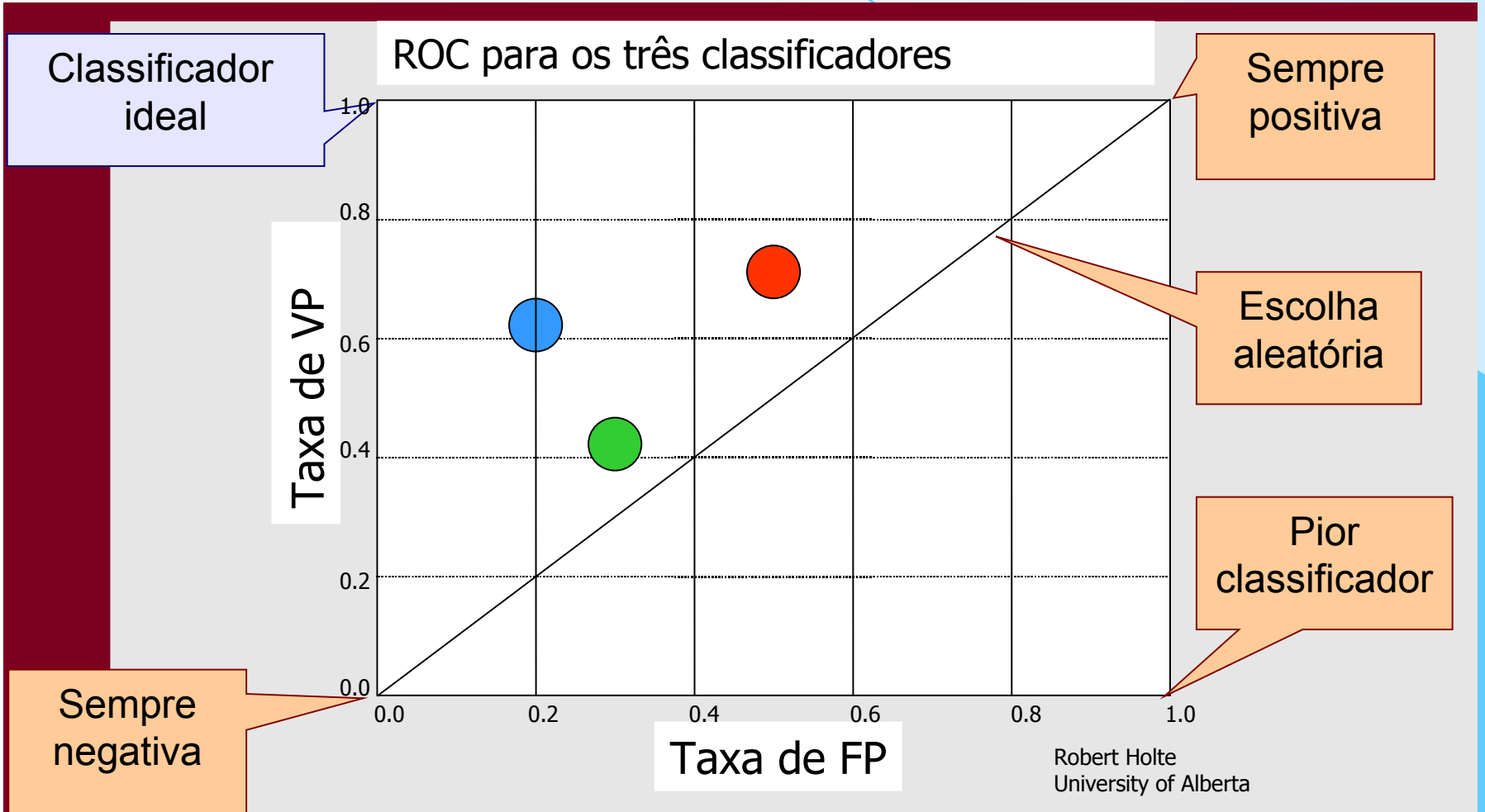


Classificador 3
TFP = 0.2
TVP = 0.6



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

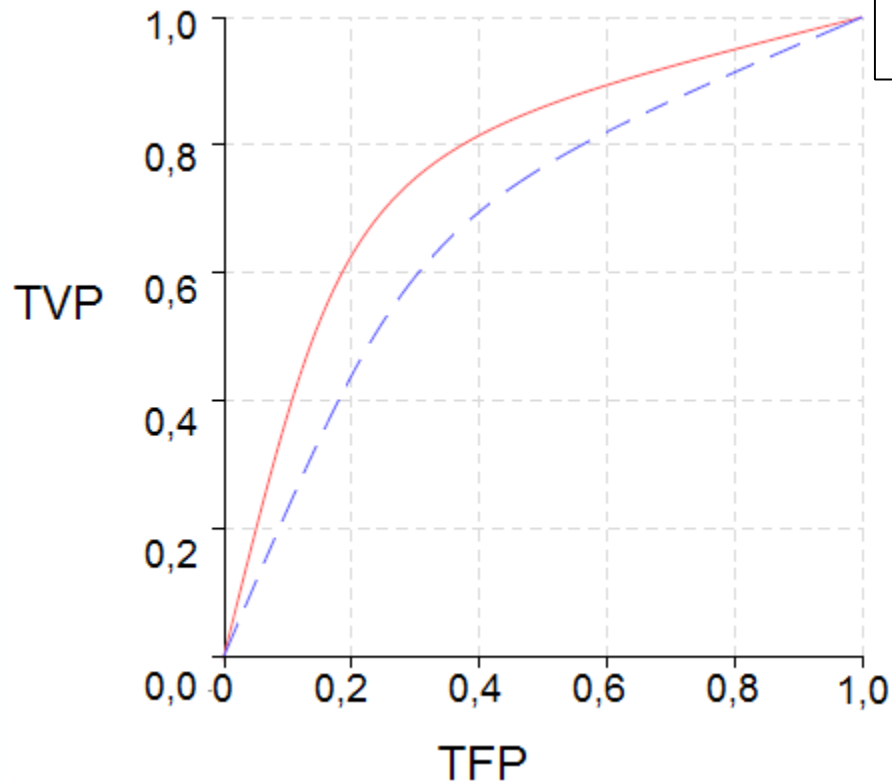


Análise ROC

- Mais usual: gerar curva ROC
- É necessário empregar um ranqueamento na classificação
- Muitos classificadores produzem medidas que permitem ranquear respostas

Análise ROC

Exemplo:



— Classificador A
- - - Classificador B

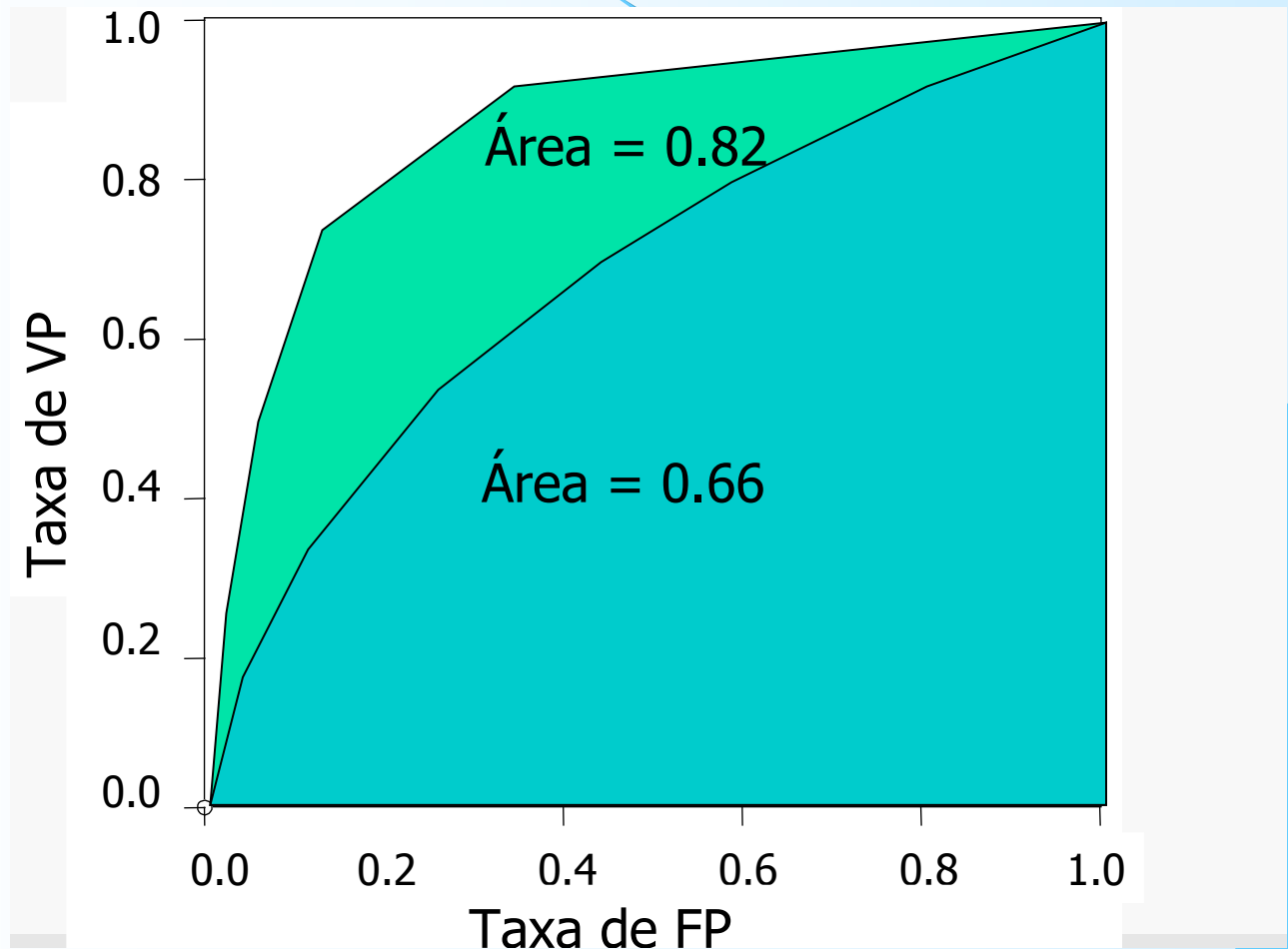
Comparando duas curvas:

- **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

Análise ROC

- Mais usual: gerar curva ROC e calcular área abaixo da curva ROC
- AUC (*Area Under ROC Curve*) Produz valores no intervalo $[0,1]$
- Valores mais próximos de 1 são considerados melhores

AUC



Resumo sobre Análise de Desempenho

- Reportar média e desvio-padrão de desempenhos
- Conjunto de teste deve ser **somente** para teste
 - Simular a chegada de dados totalmente novos ao modelo
 - Indução devem ser feitos usando os dados de treinamento
- Realizar testes estatísticos para comparar desempenhos de diferentes algoritmos
 - Diferenças podem não ser relevantes estatisticamente

Referências

- FACELI, K.; LORENA, A.C.;GAMA J.; CARVALHO, A.C.L.F. Inteligência Artificial: uma abordagem de aprendizado de máquina. Capítulos 4 e 9
- JAIN R. The Art of Computer Systems Performance Analysis, John Wiley & Sons, 1991. Capítulos:1, 2, 3 e 5.
- slides baseados em apresentações de:
 - Prof Dr André C. P. L. F. Carvalho, ICMC-USP e Profa. Dra. Ana Carolina Lorena

**UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA**

**PTC 3467
Análise de Desempenho de
modelos**

**Pós-Graduação em Engenharia Elétrica
2º, semestre de 2020**