

Curso 2 – CD, AM e DM

MBA EM IA e BIGDATA

PROFA. ROSELI AP. FRANCELIN ROMERO



PLANEJAMENTO DE EXPERIMENTOS PARTE II

CURVAS ROC

Roseli Ap. Francelin Romero
SCC – ICMC - USP



Principais tópicos

- Acurácia
- Avaliação do desempenho
- Curvas ROC



Acurácia

- Uma das mais usadas
- Taxa de objetos corretamente classificados
 - Trata as classes igualmente
 - Pode não ser adequada para dados desbalanceados
 - Pode induzir modelo com baixa taxa de acerto para classe minoritária
 - Geralmente mais interessante que a majoritária
 - Acurácia balanceada



Revocação X Precisão

- Revocação (*recall*)

- Porcentagem de exemplos positivos classificados como positivos
 - Nenhum exemplo positivo é deixado de fora
 - Todos são lembrados

$$\frac{VP}{VP + FN}$$

- Precisão

- Porcentagem de **exemplos classificados** como positivos que são realmente positivos
 - Nenhum exemplo negativo é incluído
 - Não tem intrusos

$$\frac{VP}{VP + FP}$$



Sensibilidade X Especificidade

- Sensibilidade

- Porcentagem de **exemplos positivos classificados** como positivos
 - Igual a revocação

$$\frac{VP}{VP + FN}$$

- Especificidade

- Porcentagem de **exemplos negativos classificados** como negativos
 - Nenhum exemplo negativo é deixado de fora
 - Todos são lembrados

$$\frac{VN}{VN + FP}$$



Medida-F

- Média harmônica ponderada da precisão e da revocação

- Medida-F1

- Precisão e revocação têm o mesmo peso

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

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



Observação

- A utilização das métricas **Precisão e Revocação** são recomendadas para análise da ocorrência de **Falsos Positivos** e **Falsos Negativos** respectivamente.
- Quanto maior a ocorrência de Falsos Positivos, menor é a precisão.
- Quanto maior a ocorrência de Falsos Negativos, menor é a revocação.
- Já o F1 consiste na combinação das duas métricas descritas de forma a mensurar a performance de um classificador em uma única métrica.

$$\frac{VP}{VP + FP}$$

$$\frac{VP}{VP + FN}$$



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
Verdadeiro	P	70	30
	N	40	60



Exemplo

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

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

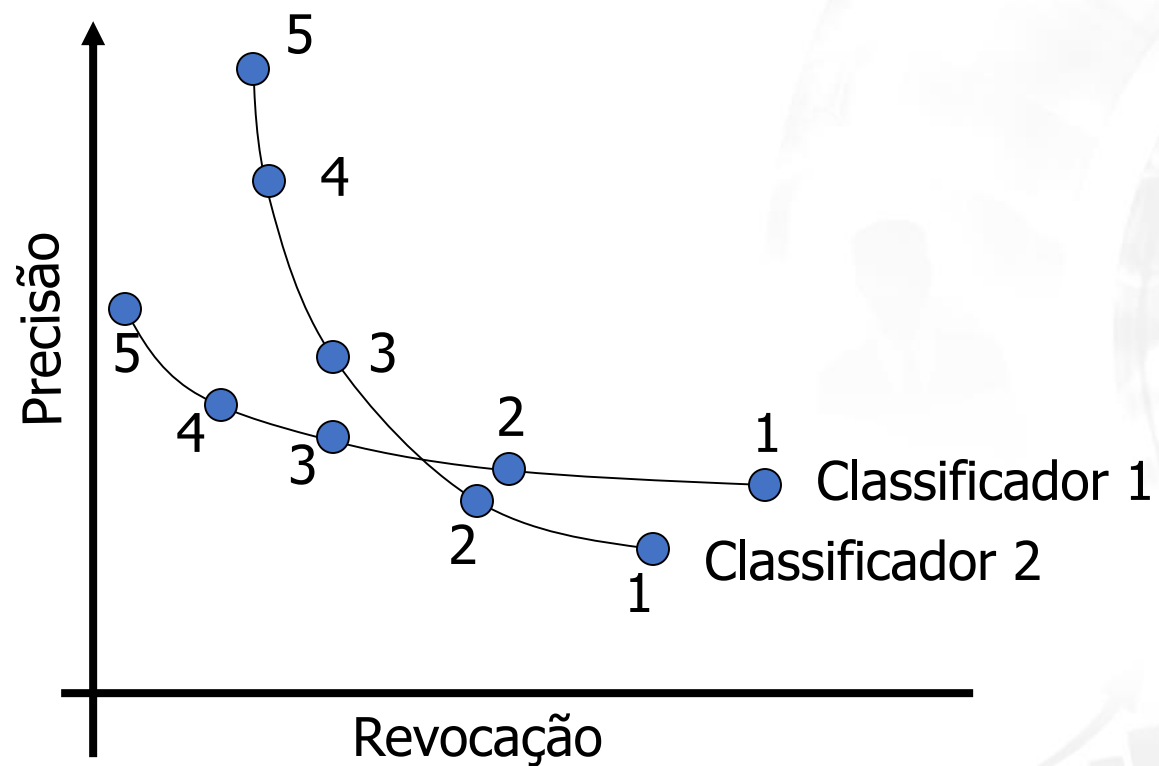
$$\text{Revocação} = \frac{VP}{VP + FN} = 70 / (70 + 30) = 0.70$$

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

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



Observação



Outras medidas

- Média geométrica de taxas positivas
 - G-mean
- Acurácia balanceada $\sqrt{\textit{precisão} \times \textit{revocação}}$
- Kappa



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)



Curvas de ROC

- A [curva ROC](#) (ou curva de ROC) foi desenvolvida por engenheiros elétricos e engenheiros de sistemas de radar durante a [Segunda Guerra Mundial](#) para detectar objetos inimigos em campos de batalha, também conhecida como [teoria de detecção de sinais](#).
- A análise ROC tem sido utilizada em medicina, radiologia, psicologia e outras áreas por muitas décadas e, mais recentemente, foi introduzida à áreas como [aprendizado de máquina](#) e [mineracao de dados](#).



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

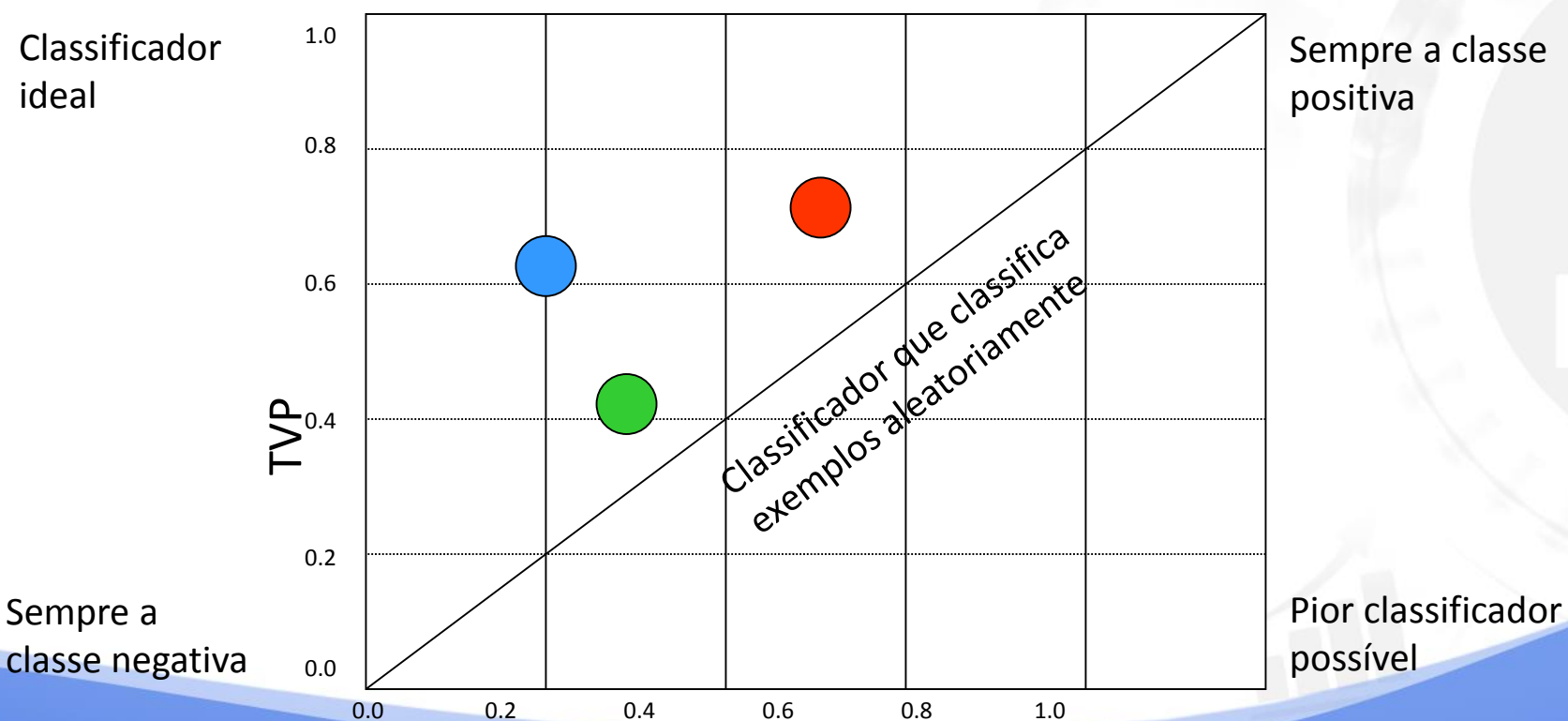


Classificador 3
TFP = 0.2
TVP = 0.6



Gráficos ROC

ROC para três classificadores



Gráficos ROC

- Classificadores discretos produzem um simples ponto no gráfico ROC
 - ADs e conjuntos de regras
- Outros classificadores produzem uma probabilidade ou score
 - RNAs e NB
- Curvas ROC permitem uma melhor comparação de classificadores
 - São insensíveis a mudanças na distribuição das classes

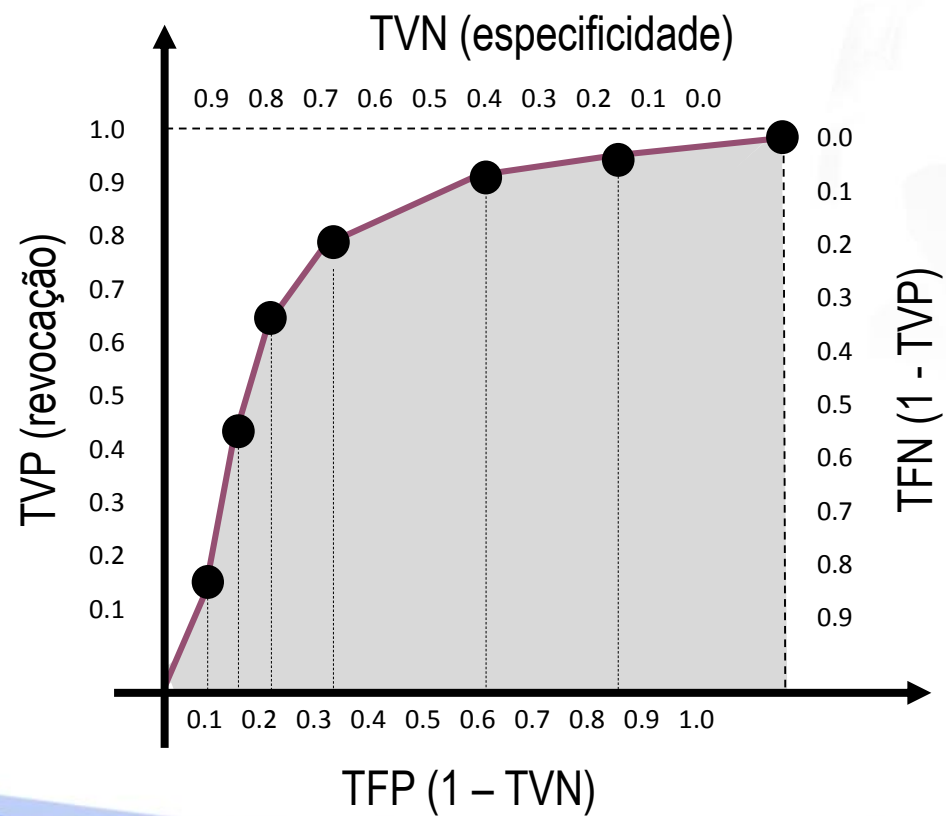


Curvas ROC

- Mostram ROC para diferentes variações
- Classificadores que geram valores contínuos (*threshold*, probabilidade)
 - Diferentes valores de *threshold* podem ser utilizados para gerar vários pontos
 - Ligação dos pontos gera uma curva ROC
- Classificadores discretos
 - Convertidos internamente ou comitês



Curva ROC



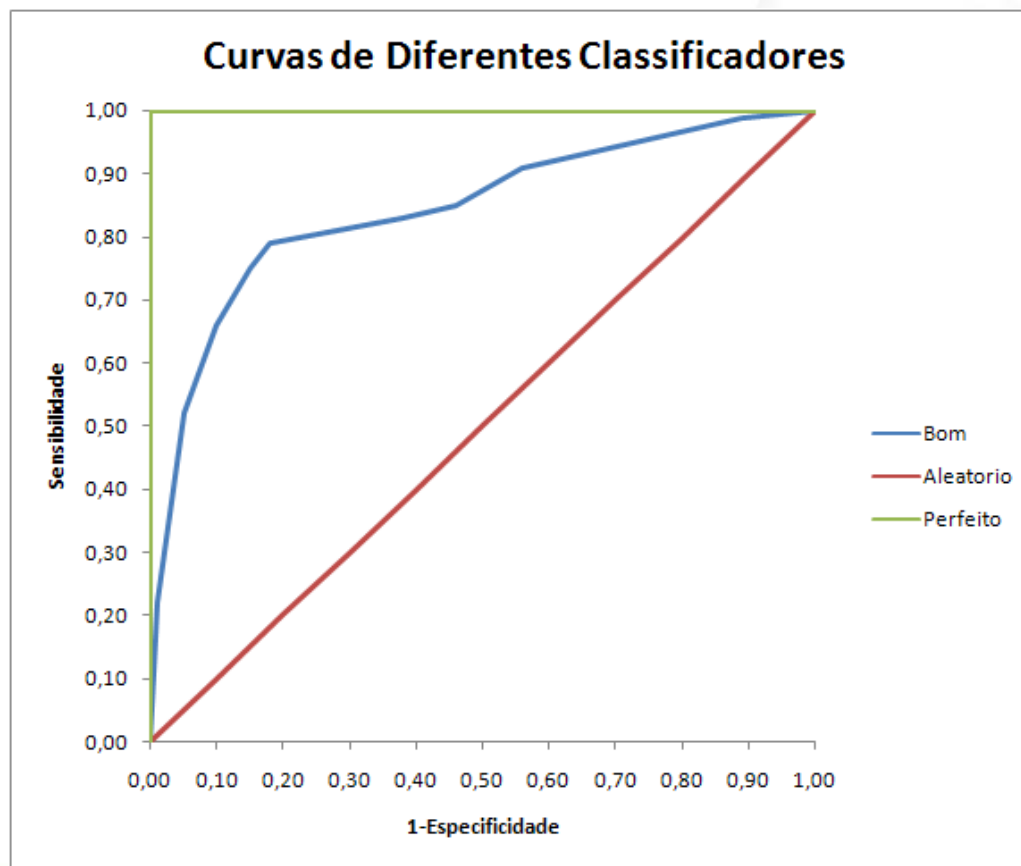
Curvas de ROC

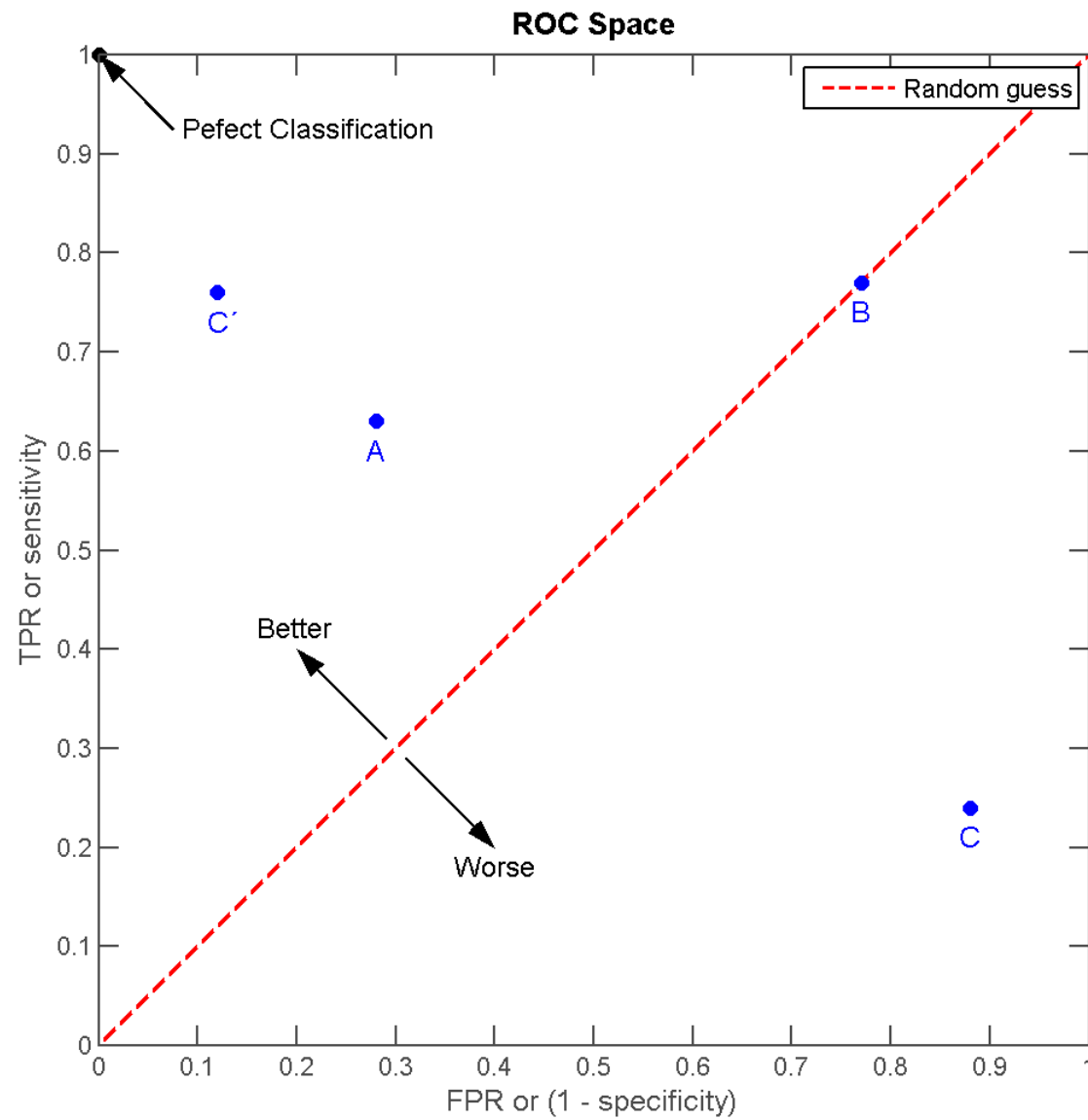
O melhor método de previsão possível produziria um ponto no canto superior esquerdo ou coordenada $(0,1)$ do espaço ROC, representando 100% de sensibilidade (sem falsos positivos) e

100% de especificidade (sem falsos negativos).



Curvas de ROC





Wikipedia

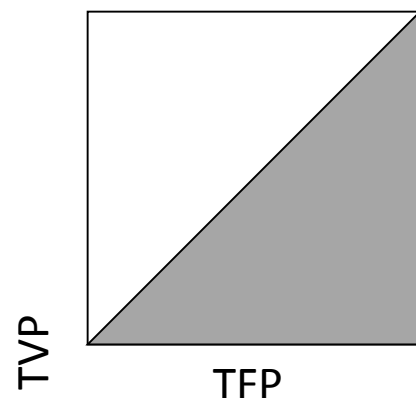


Área sob a curva ROC (AUC)

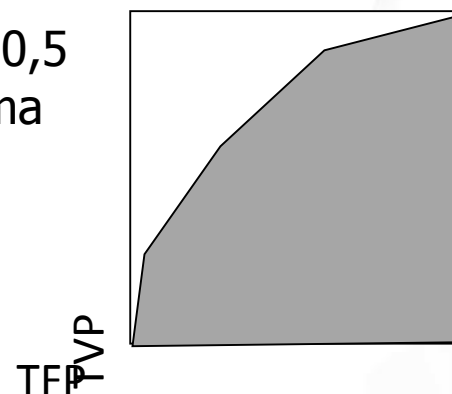
- Fornece uma estimativa do desempenho de classificadores
- Gera um valor contínuo no intervalo $[0, 1]$
 - Quanto maior melhor
 - Adição de áreas de sucessivos trapezóides
- Um classificador com maior AUC pode apresentar AUC pior em trechos da curva
- É mais confiável utilizar médias de AUCs



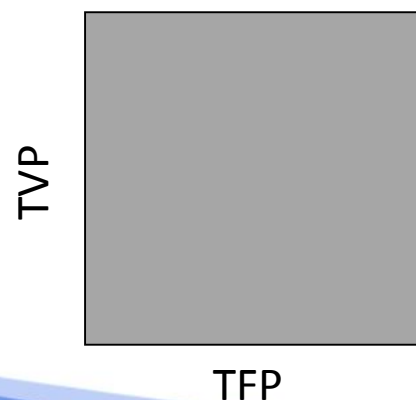
Área Sob Curvas ROC



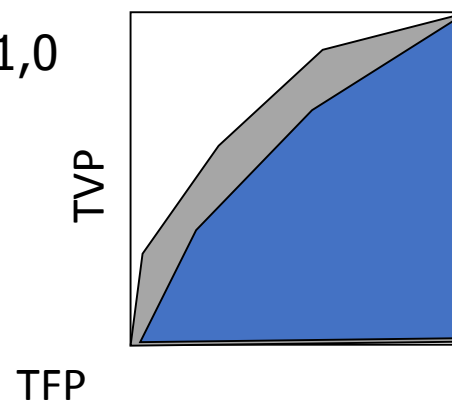
Área = 0,5
Nenhuma



Área = 0,74



Área = 1,0
Perfeita



Área = 0,74
Área = 0,67



Teste de hipótese

- Permite afirmar que uma técnica é melhor que outra com X% de confiança
- Podem assumir que os dados seguem uma dada distribuição de probabilidade
 - Paramétricos
 - Não paramétricos
- Número de técnicas comparadas
 - Duas
 - Mais que duas



Teste de hipótese

- Testes usados atualmente são baseados na verificação da hipótese nula
 - Várias deficiências para uso em AM
 - Não geram probabilidades de ocorrência da hipótese nula e da hipótese alternativa
 - E de uma técnica ser melhor que outra
- Se $p < 0.05$ – rejeita-se a Hipotese Nula



Como escolher o TESTE

- A decisão muitas vezes é arbitrária e subjetiva, a fim de reduzir essa subjetividade e arbitrariedade deve considerar algumas características.
- Como cada teste tem um conjunto de requisitos próprios que precisam ser satisfeitos, para fazer sentido a utilização de determinado teste, primeiro é preciso identificar no cenário que se está avaliando se existe alguma função de distribuição dos dados.



Como escolher o TESTE

- A distribuição das amostras é um fator importante para a escolha do teste a ser utilizado. O mais comum é assumir que os dados do teste seguem uma distribuição normal, mas nem sempre isso é verdadeiro.



Como escolher o TESTE

- Os testes que assumem alguma distribuição são chamados paramétricos, os não paramétricos não supõem uma distribuição dos dados.
- A quantidade de exemplos em cada amostra também é importante para a escolha do teste.



Como escolher o TESTE

- A escolha da quantidade de amostras que se está avaliando é importante para determinar qual teste utilizar.
- O mais comum é comparar dois classificadores, mas as vezes temos vários classificadores e queremos determinar se algum é melhor dentre os vários.

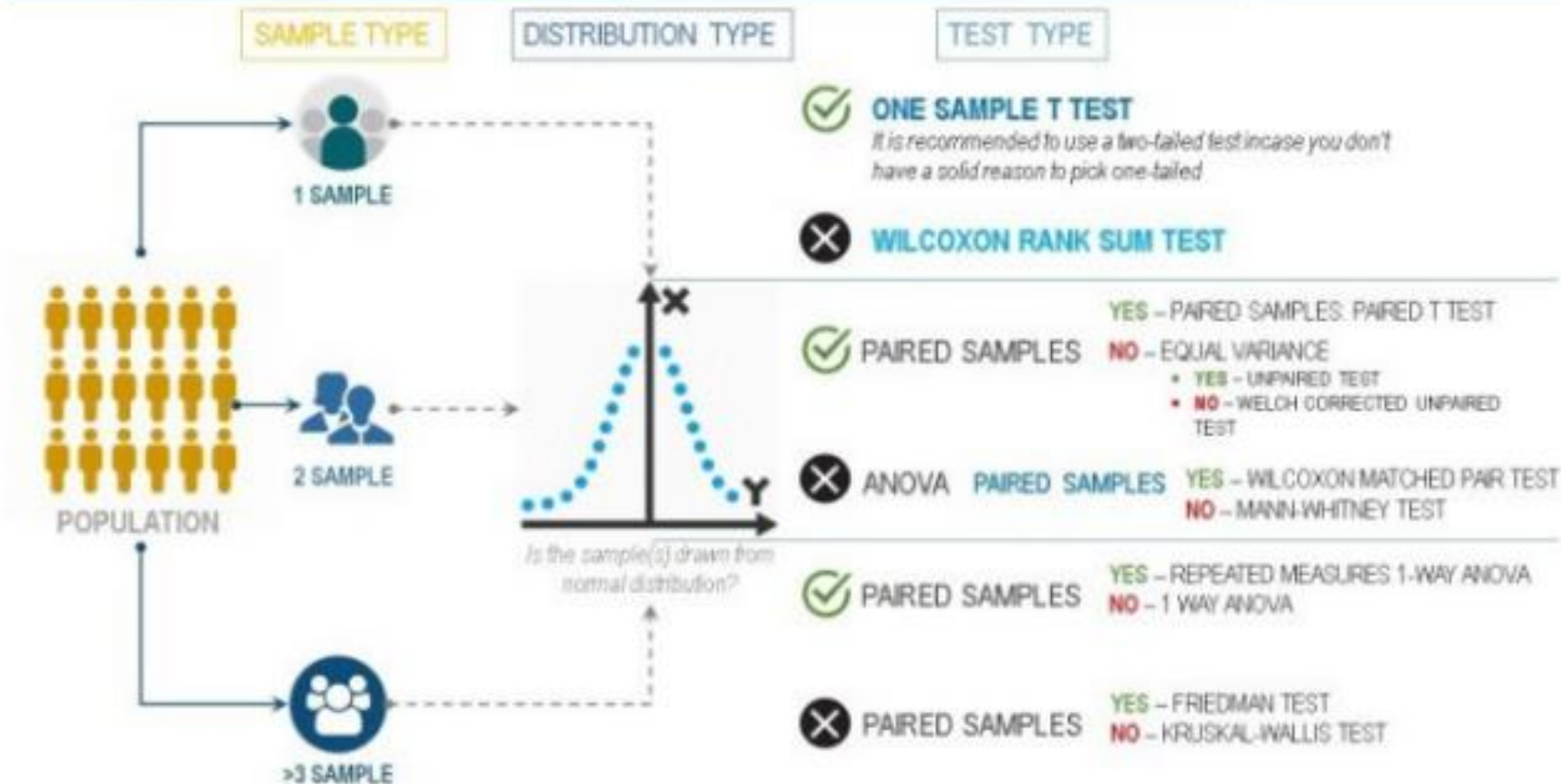


Como escolher o TESTE

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116565/>
- Dois classificadores: testes Não-pareados
- Se os dados seguem uma normal utilizamos **o teste t não pareado**, senão utilizar o teste U Mann-Whitney ou teste da soma dos ranks de Wilcoxon.



A SIMPLE GUIDE FOR SELECTING STATISTICAL TEST WHEN COMPARING GROUPS



Considerações Finais

- Desempenho preditivo
- Avaliação do desempenho
 - Erro
 - Tempo de resposta
 - Memória
 - Representação
- Medidas
- Gráficos e curvas ROC
- Teste de hipóteses

