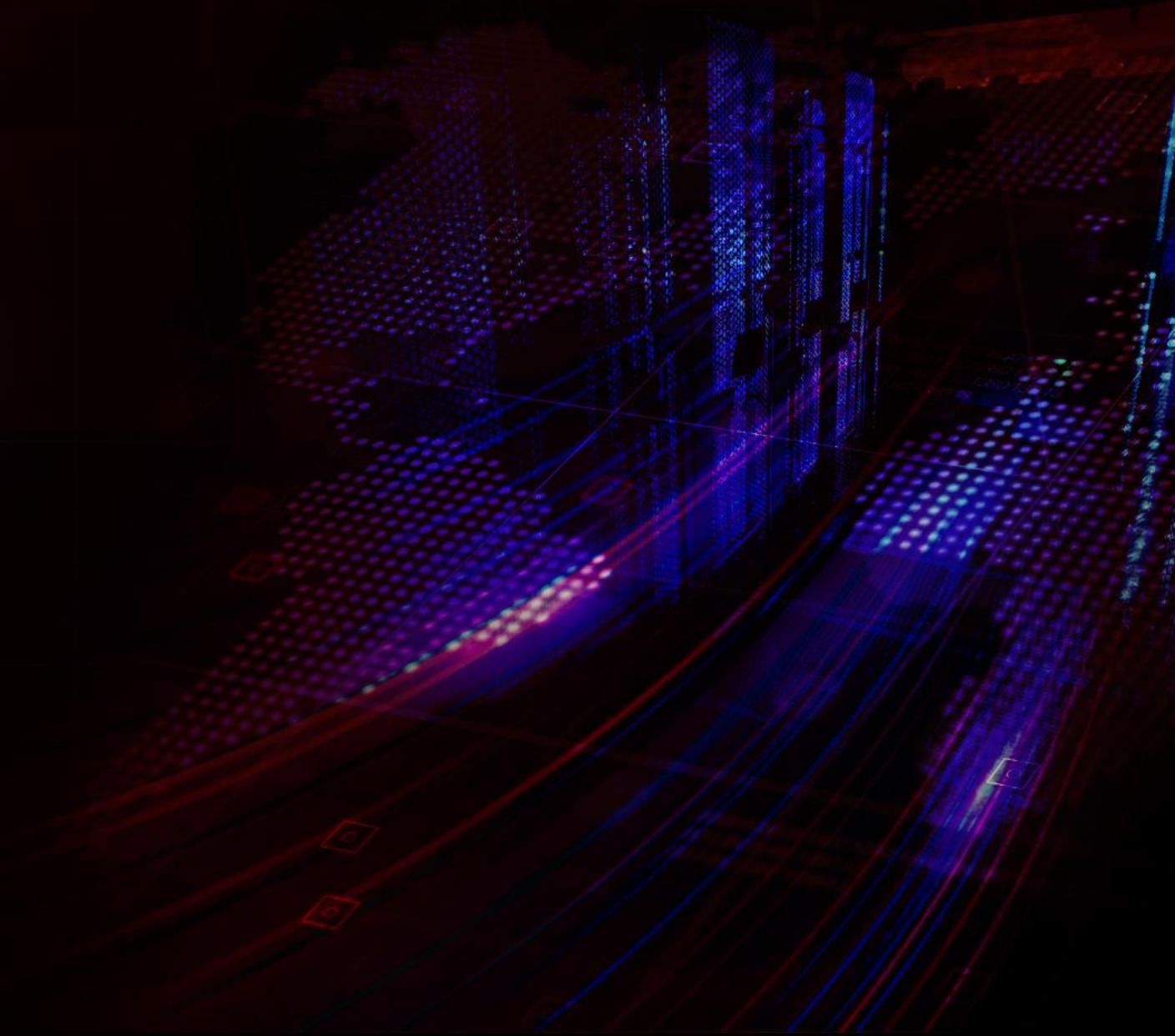


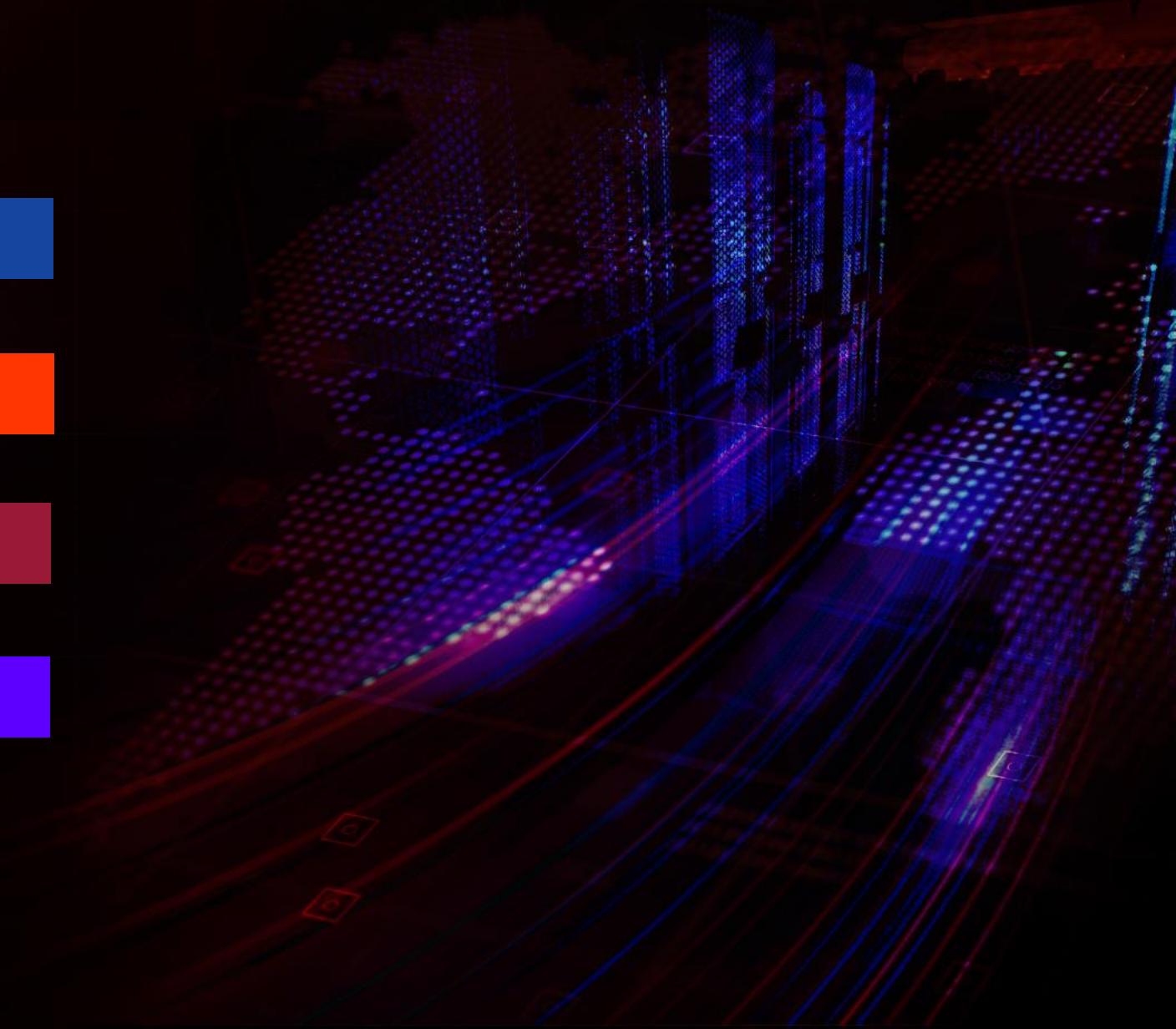
APRENDIZADO DE MÁQUINAS

Avaliação de Modelos Preditivos



TÓPICOS

- Métricas de erro
- Medidas de desempenho
- Problemas de duas classes
- Testes de hipóteses



MODELOS PREDITIVOS

Não existe um único algoritmo melhor em todos os casos

Características do conjunto de dados

Características do problema

Parâmetros dos algoritmos



MODELOS PREDITIVOS

Dados	X	Técnica
Alta dimensionalidade	→	SVMs
Baixa dimensionalidade	→	k-NN
Regras bem definidas	→	Árvores de Decisão

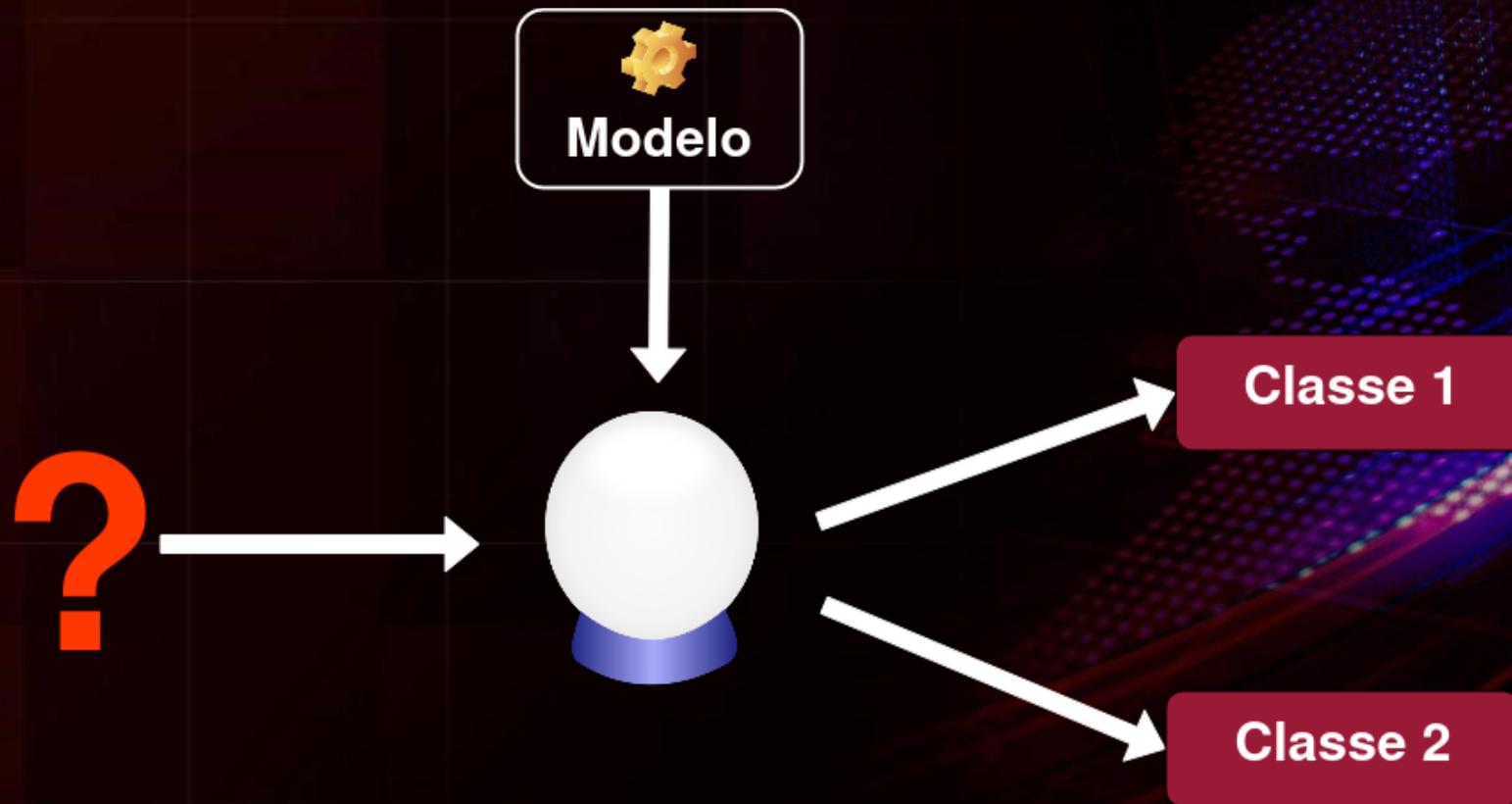
MODELOS PREDITIVOS

Dados	X	T
Alta dimensionalidade	→	SVMs
Baixa dimensionalidade	→	k-NN
Regras binárias	→	Árvores de Decisão

Experimentação

MÉTRICAS DE ERRO

Desempenho na rotulação de objetos do conjunto de teste



MÉTRICAS DE ERRO - CLASSIFICAÇÃO

- C : classificador
- n : total de objetos (dados)
- I : função de comparação
 - $I(\text{verdadeiro}) = 1$
 - $I(\text{falso}) = 0$
- y_i : classe verdadeira
- x_i : classe predita
- i : objeto avaliado



MÉTRICAS DE ERRO - CLASSIFICAÇÃO

Erro (E)

% de objetos errados

$$E(C) = \frac{1}{n} \sum_{i=1}^n I(y_i \neq C(x_i))$$

Acurácia (ACC)

% de objetos corretos

$$ACC(C) = \frac{1}{n} \sum_{i=1}^n I(y_i = C(x_i))$$

0.0

pior

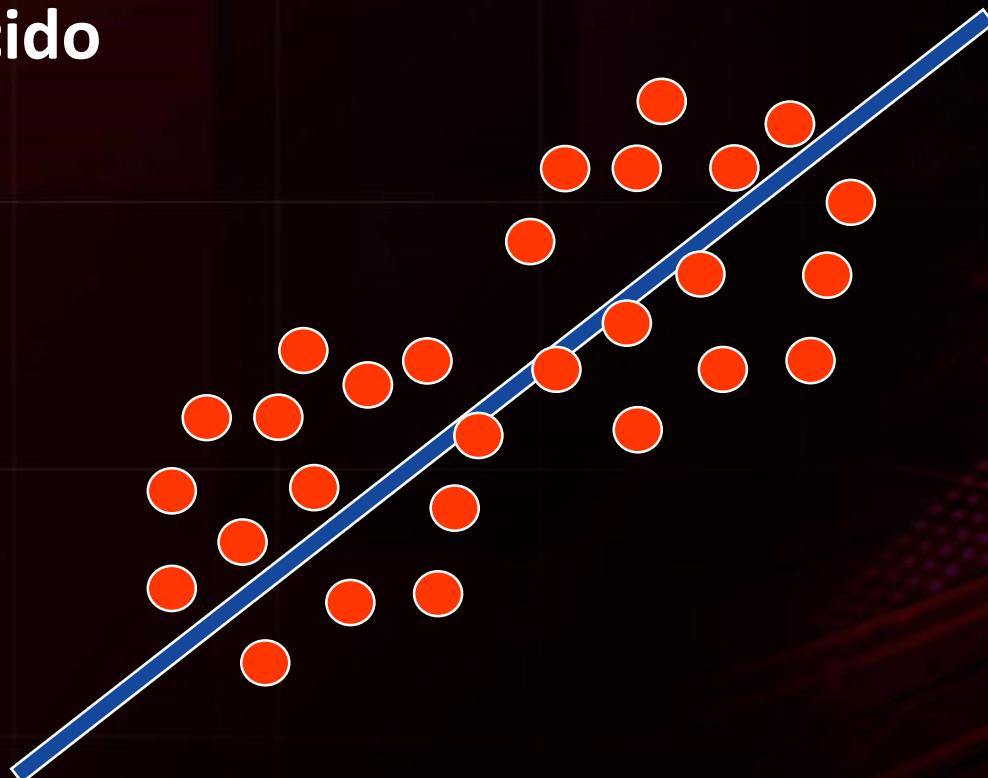
1.0

melhor

MÉTRICAS DE ERRO - REGRESSÃO

O erro é a distância entre o valor conhecido (y_i) e o valor predito (x_i)

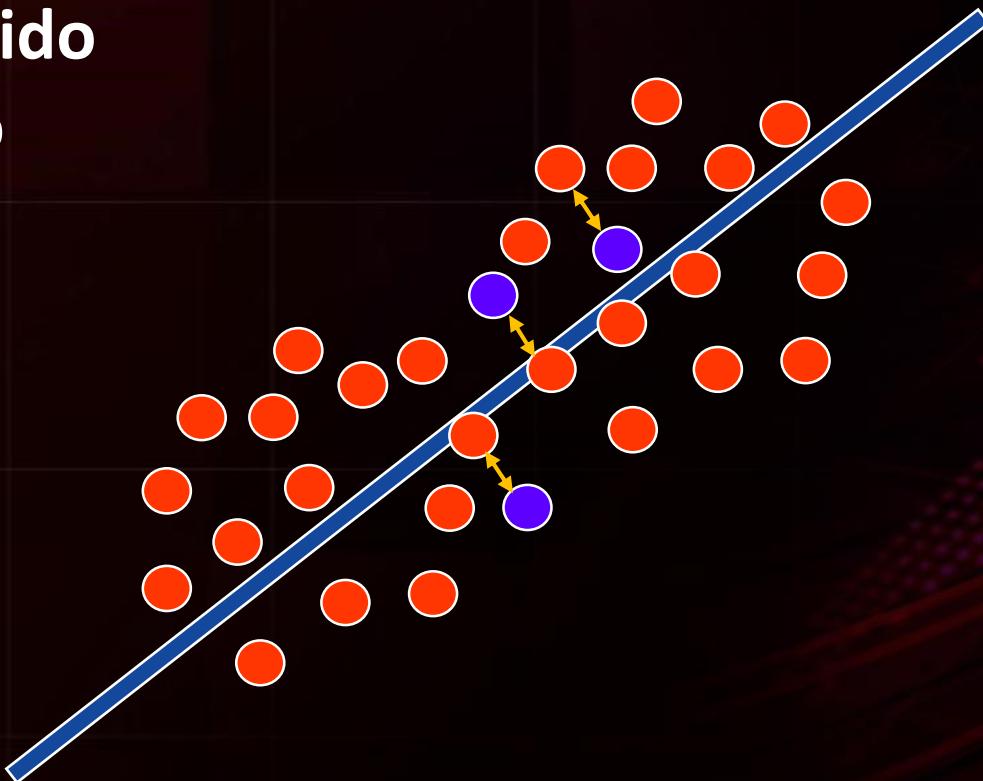
- conhecido



MÉTRICAS DE ERRO - REGRESSÃO

O erro é a distância entre o valor conhecido (y_i) e o valor predito (x_i)

- conhecido
- predito



MÉTRICAS DE ERRO - REGRESSÃO

Erro quadrático médio
(MSE)

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

Distância absoluta média
(MAD)

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

0.0

melhor

+

pior

MÉTRICAS DE ERRO - REGRESSÃO

Erro quadrático médio normalizado (NMSE)

$$NMSE(C) = \frac{\sum_{i=1}^n (y_i - C(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



CLASSIFICAÇÃO - DUAS CLASSES

Matriz de confusão		Classe predita	
Classe original	Positiva	Positiva	Negativa
	Negativa	FP	VN

- VP: Verdadeiros Positivos
- VN: Verdadeiros Negativos
- FP: Falsos Positivos
- FN: Falsos Negativos

CLASSIFICAÇÃO - ERRO E ACURÁCIA

Acurácia (ACC)

% global de objetos corretos

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

Erro (E)

% global de objetos errados

$$E = 1 - ACC$$

CLASSIFICAÇÃO - MEDIDAS DE DESEMPENHO

Taxa de falsos negativos (TFN)

**% objetos da classe positiva
classificados incorretamente**

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

Taxa de falsos positivos (TFP)

**% objetos da classe negativa
classificados incorretamente**

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

CLASSIFICAÇÃO - MEDIDAS DE DESEMPENHO

Precisão

% objetos da classe positiva
classificados corretamente

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

Revocação (recall)

Taxa de acerto na
classe positiva (TVP)

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

Especificidade

Taxa de acerto na
classe negativa (TFP)

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



CLASSIFICAÇÃO - MEDIDAS DE DESEMPENHO

Medida F (F-score)

w: peso entre precisão (prec)
e revocação (rev)

$$\frac{(w+1) \times \text{prec}(C) \times \text{rev}(C)}{\text{prec}(C) + w \times \text{rev}(C)}$$

Medida F_1 (F_1 score)

precisão = revocação

$$\frac{2 \times \text{prec}(C) \times \text{rev}(C)}{\text{prec}(C) + \text{rev}(C)}$$

ANÁLISE ROC

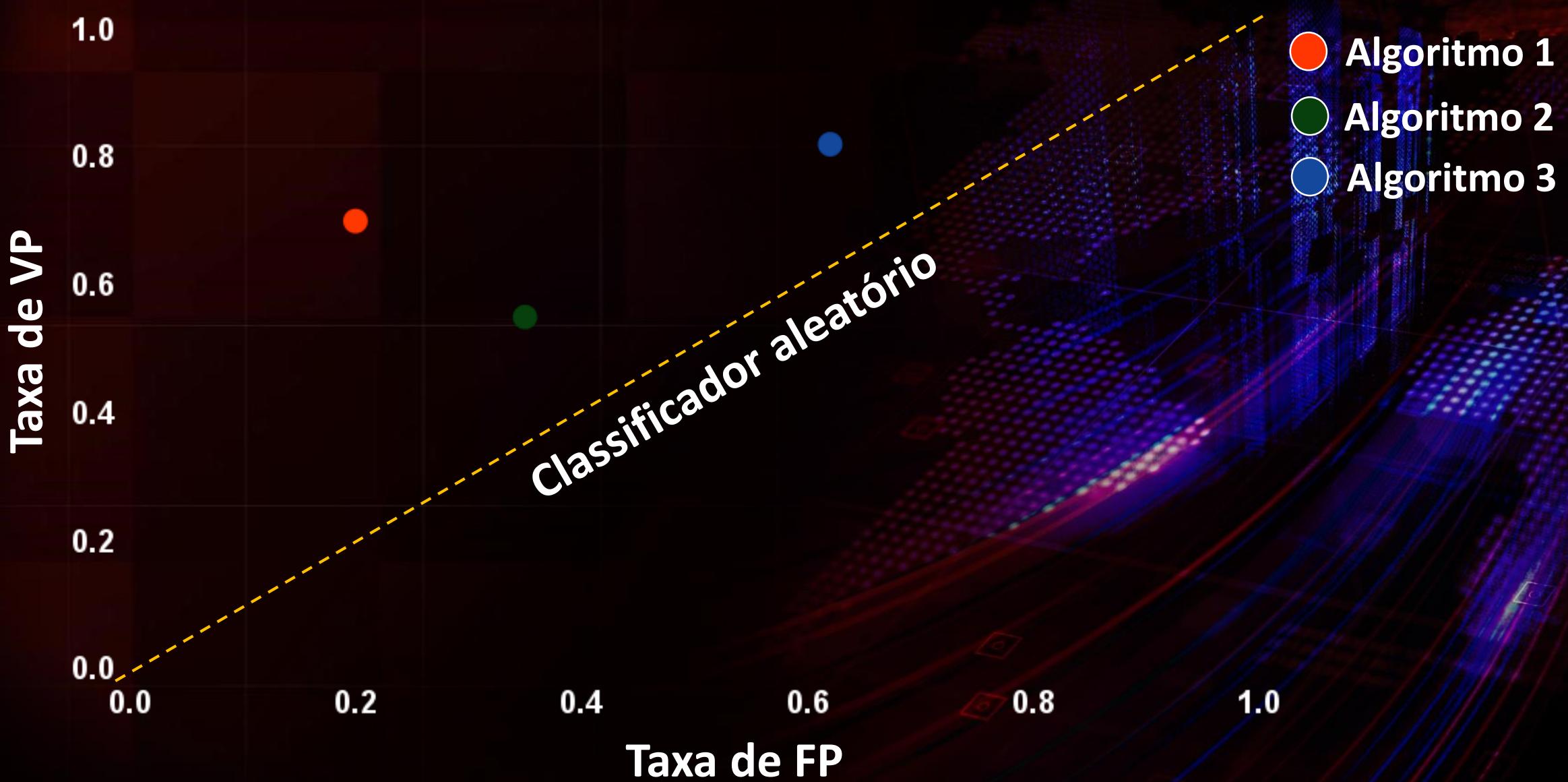
Receiving Operating Characteristics

Comparação de desempenho de classificadores

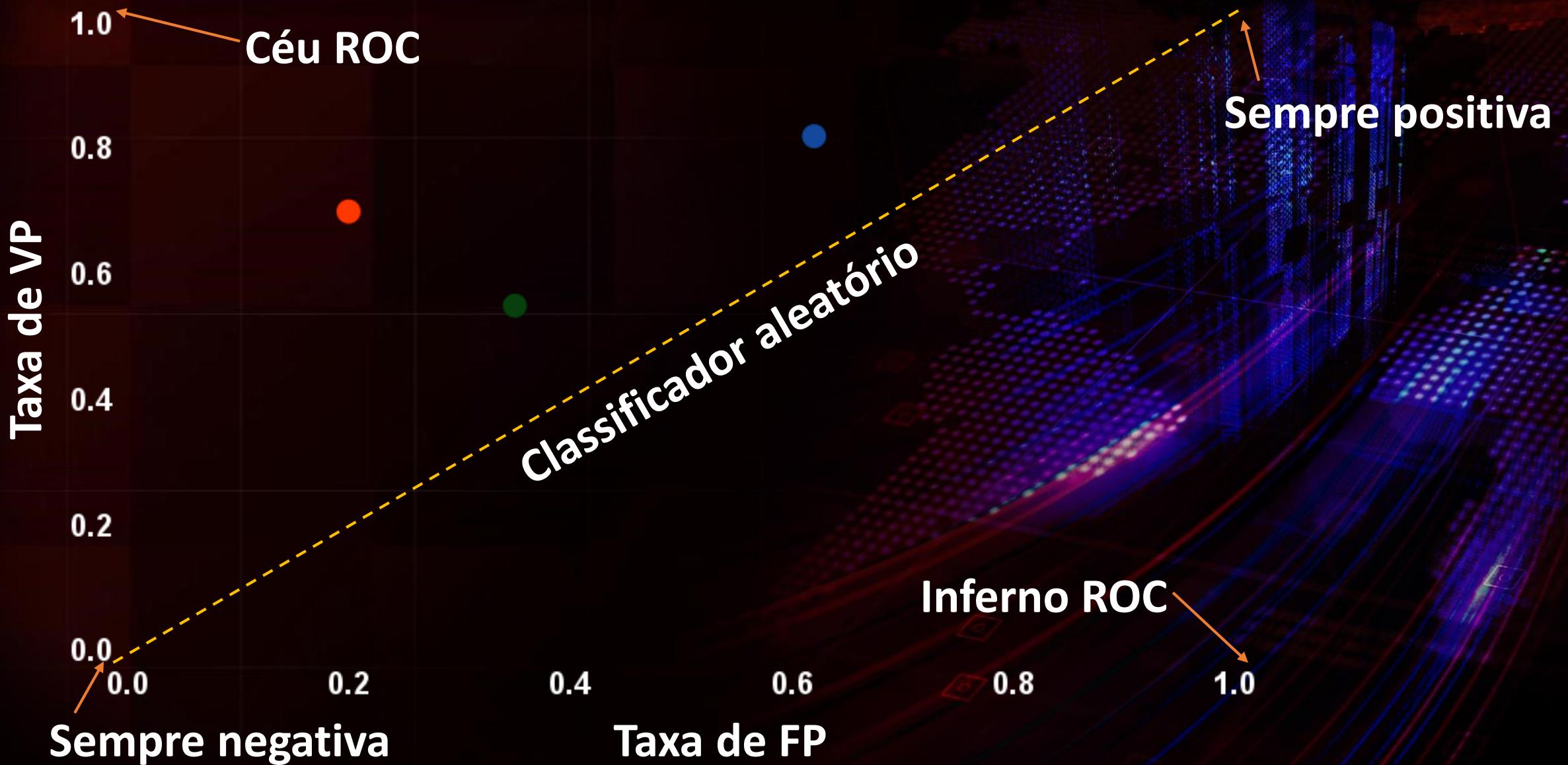
Gráfico bidimensional: X (TFP), Y (TVP)



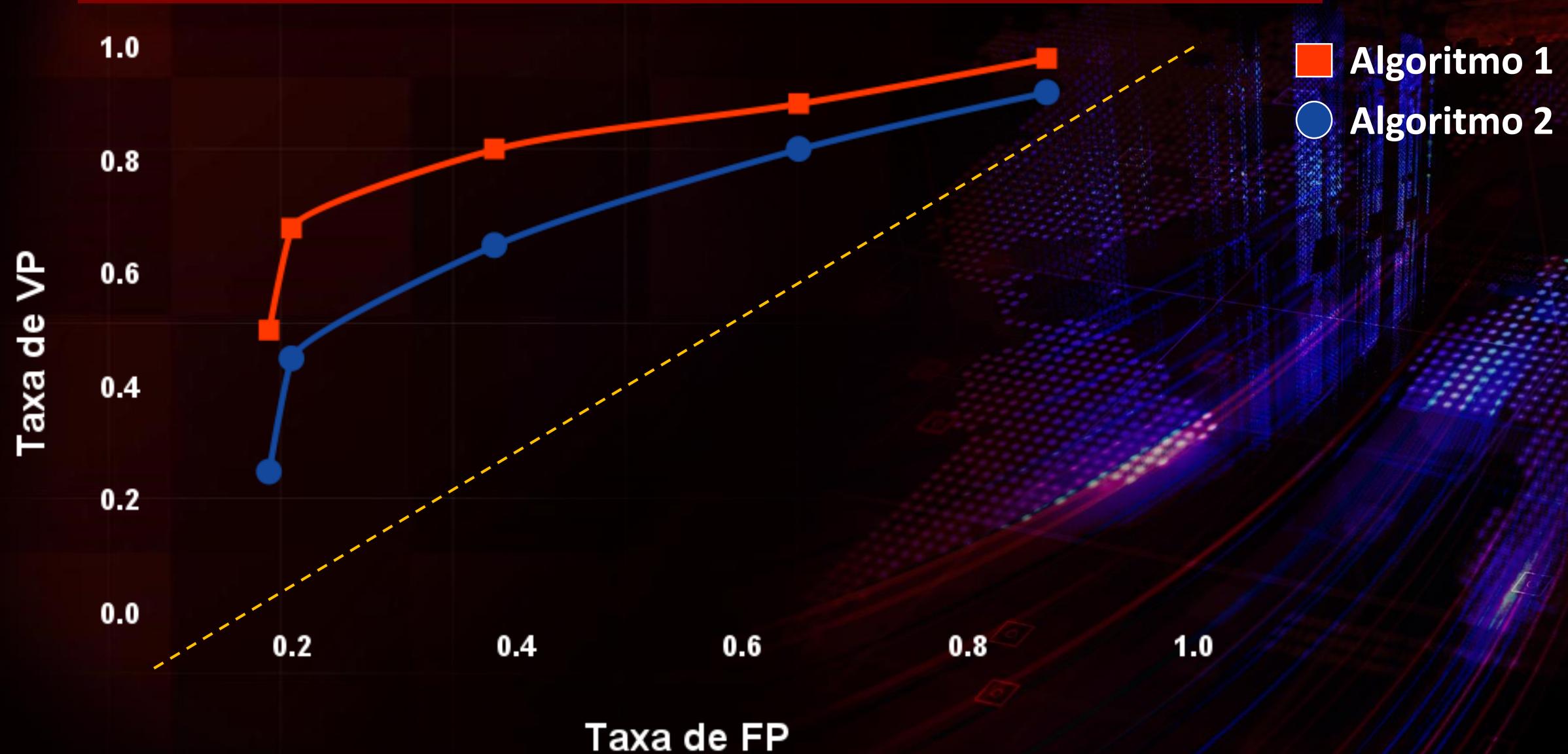
ANÁLISE ROC



ANÁLISE ROC



CURVA ROC



TESTES DE HIPÓTESES

- Comparação de dois ou mais algoritmos
- Igualdade de condições
 - Mesmas partições dos dados (validação cruzada)
- Média da taxa de erro

Nem sempre as diferenças entre os algoritmos são significativas

TESTES DE HIPÓTESES

Hipótese nula ($H_0: \mu_1 - \mu_2 = 0$): modelos são equivalentes

Hipótese alternativa ($H_1: \mu_1 - \mu_2 \neq 0$): modelos são diferentes

μ_1 : erro médio do algoritmo 1

μ_2 : erro médio do algoritmo 2

Significância: $\alpha = 0,05$

Nível de confiança de 95% de não rejeitar H_0 verdadeira

TESTES DE HIPÓTESES

Teste Wilcoxon signed-rank

- Comparação de dois algoritmos (teste pareado)
- Não paramétrico
- Calcula a diferença no desempenho dos algoritmos
- Ordena em ordem crescente de diferença
- Compara as posições das diferenças positivas e negativas

TESTES DE HIPÓTESES

Teste Wilcoxon signed-rank

```
from scipy import stats  
stats.wilcoxon(alg1_erro_medio, alg2_erro_medio)
```

```
>WilcoxonResult(statistic=2235.3, pvalue=0.049)
```

*`alg1_erro_medio, alg2_erro_medio`: listas com ao menos 20 resultados de cada algoritmo

TESTES DE HIPÓTESES

Teste Wilcoxon signed-rank

```
from scipy import stats  
  
stats.wilcoxon(alg1_erro_medio, alg2_erro_medio)  
  
>WilcoxonResult(statistic=2235.3, pvalue=0.049)
```

Hipótese nula rejeitada. Há diferenças de desempenho entre os algoritmos 1 e 2

REFERÊNCIAS

Inteligência Artificial - Uma Abordagem de Aprendizado de
Máquina. Katti Facelli et al. SAGAH, 2021.

APRENDIZADO DE MÁQUINAS

Avaliação de Modelos Preditivos

