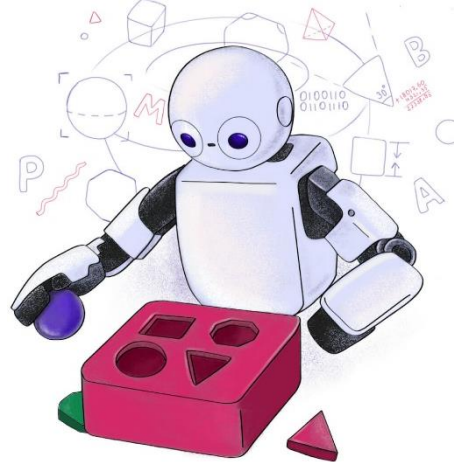


T320 - Introdução ao Aprendizado de Máquina II: *Classificação (Parte V)*



Recapitulando

- Anteriormente, vimos como lidar com problemas de classificação que envolvem mais de duas classes, também chamados de problemas de ***classificação multi-classes*** através das abordagens:
 - Um-Contra-Resto
 - Um-Contra-Um
 - Regressão Softmax
- Nesta aula, veremos as ***métricas*** mais utilizadas para medir o ***desempenho de classificadores***.

Métricas para avaliação de classificadores

- As métricas para avaliação do desempenho de classificadores que estudaremos são:
 - Matriz de confusão
 - Várias métricas podem ser extraídas da matriz.
 - Pontuação-F1 (*F1-score*)
 - Curva Característica Operacional do Receptor (do inglês, *Receiver Operating Characteristic* - ROC)

Matriz de confusão

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1Q} \\ C_{21} & C_{22} & & C_{2Q} \\ & \vdots & \ddots & \vdots \\ C_{Q1} & C_{Q2} & \dots & C_{QQ} \end{bmatrix}$$

- O nome, ***matriz de confusão*** mostra o quanto um classificador está se ***confundindo***.
- A matriz permite verificar quais classes o ***classificador*** tem ***maior dificuldade em classificar***.
- A ***matriz de confusão contabiliza o número de classificações corretas e incorretas*** para cada uma das Q classes existentes.
- É uma ***matriz quadrada*** com dimensões $\mathbb{R}^{Q \times Q}$.

Matriz de confusão

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1Q} \\ C_{21} & C_{22} & & C_{2Q} \\ & \vdots & \ddots & \vdots \\ C_{Q1} & C_{Q2} & \cdots & C_{QQ} \end{bmatrix}$$

- A **diagonal principal** de \mathbf{C} fornece o número de **classificações corretas**.
- A q -ésima **linha** indica o **total de exemplos que foram classificados como pertencentes a q -ésima classe**, incluindo exemplos que pertencem e não pertencem à classe.
- A q -ésima **coluna** indica o **total de exemplos que realmente pertencem à q -ésima classe**, incluindo classificações corretas e incorretas.

Matriz de confusão

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1Q} \\ C_{21} & C_{22} & \dots & C_{2Q} \\ \vdots & & \ddots & \vdots \\ C_{Q1} & C_{Q2} & \dots & C_{QQ} \end{bmatrix}$$

Exemplos classificados como pertencentes à classe 1.

Quantidade de exemplos realmente pertencentes à classe 1.

- C_{11} indica quantos exemplos da classe 1 foram corretamente atribuídos à classe 1.
- C_{12} indica quantos exemplos da classe 2 foram atribuídos à classe 1.

Matriz de confusão

Predições: classes a que os exemplos foram atribuídos

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1Q} \\ C_{21} & C_{22} & & C_{2Q} \\ & \vdots & \ddots & \vdots \\ C_{Q1} & C_{Q2} & \dots & C_{QQ} \end{bmatrix}$$

Rótulos: classes a que realmente pertencem os exemplos

Matriz de confusão para caso binário ($Q = 2$)

Classes Estimadas	Classes Verdadeiras	
	- (C_1)	+ (C_2)
- (C_1)	Verdadeiro Negativo (TN)	Falso Negativo (FN)
+ (C_2)	Falso Positivo (FP)	Verdadeiro Positivo (TP)

- **Verdadeiro Positivo (TP)**: número de exemplos da classe positiva, C_2 , classificados corretamente.
- **Verdadeiro Negativo (TN)**: número de exemplos da classe negativa, C_1 , classificados corretamente.
- **Falso Positivo (FP)**: número de exemplos atribuídos à classe positiva, mas que pertencem à classe negativa.
- **Falso Negativo (FN)**: número de exemplos atribuídos à classe negativa, mas que pertencem à classe positiva.

Nós podemos calcular algumas métricas de desempenho importantes a partir das informações contidas na matriz de confusão.

Acurácia

$$\text{acurácia} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}$$

- Acurácia mede a *proporção de exemplos classificados corretamente em relação ao total de exemplos* avaliados.
- Em outras palavras, a acurácia fornece uma indicação de quão bem o modelo está fazendo suas *predições corretas em comparação com todas as previsões feitas*.
- É uma métrica útil para avaliar a *performance geral do modelo*.

Acurácia

$$\text{acurácia} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}$$

- A **acurácia** é, geralmente, a **primeira escolha** para medir a qualidade de um classificador.
- Entretanto, ela pode ser **enganosa com problemas desbalanceados**.
 - Problemas onde uma ou algumas classes têm muito mais exemplos do que as demais.
- Nesses casos, ela pode nos levar a concluir que um **classificador ruim é muito bom**.

Acurácia

- Analisando a equação da acurácia, o que aconteceria se TP fosse muito maior do que TN, FN e FP?

$$\lim_{TP \rightarrow \infty} \text{acurácia} = \lim_{TP \rightarrow \infty} \left(\frac{TP + TN}{TP + TN + FN + FP} \right) = \frac{TP}{TP} = 1.$$

- Portanto, quando temos *classes desbalanceadas, precisamos analisar outras métricas*.
 - O mesmo aconteceria se TN fosse muito maior do que TP, FN e FP
- Ela também é enganosa quando os custos de falsos positivos e falsos negativos não são iguais.

Precisão

$$\text{precisão} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

TN	FN
FP	TP

- **Precisão** é a proporção de exemplos da classe positiva corretamente classificados (TP) em relação a todos os exemplos atribuídos à classe positiva (TP + FP).
- É uma **boa medida** para determinar a qualidade do classificador **quando os custos de falsos positivos são altos**.
 - Por exemplo, na classificação de **spams** (**verdadeiro positivo**), um **falso positivo** significa que um **ham** (**verdadeiro negativo**) foi classificado como **spam**. O usuário de *email* pode perder *emails* importantes se a **precisão** for baixa.

Recall

$$recall = \frac{TP}{FN + TP}$$

TN	FN
FP	TP

- **Recall** ou sensibilidade é a proporção de exemplos da classe positiva corretamente classificados.
- O *recall* calcula quantos exemplos realmente positivos o classificador captura em relação a todos exemplos positivos.
- É uma **boa medida** para determinar a qualidade de um classificador **quando os custos de falsos negativos são altos**.
 - Por exemplo, na classificação de doenças, se um paciente doente (**verdadeiro positivo**) for classificado como não doente (**falso negativo**). O custo associado ao **falso negativo** será extremamente alto se a doença for contagiosa.

Especificidade

$$\text{especificidade} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

TN	FN
FP	TP

- **Especificidade** ou *taxa de verdadeiros negativos* é a proporção de exemplos da classe negativa corretamente classificados.

Matriz de confusão para caso multiclass ($Q > 2$)

Classe C_1 é a positiva.

Classes Preditas	$+(C_1)$	Verdadeiro Positivo (TP)	Falso Positivo (FP)	Falso Positivo (FP)
	$-(C_2)$	Falso Negativo (FN)	Verdadeiro Negativo (TN)	Verdadeiro Negativo (TN)
	$-(C_3)$	Falso Negativo (FN)	Verdadeiro Negativo (TN)	Verdadeiro Negativo (TN)
		$+(C_1)$	$-(C_2)$	$-(C_3)$
		Classes Verdadeiras		

- É possível calcular as métricas anteriores para o cenário multiclass (i.e., $Q > 2$).
- Para isso, basta selecionar, uma vez, cada classe como sendo a classe positiva, enquanto todas as demais classes formam a classe negativa.
 - Estratégia um-contra-o-resto.
- Assim, obtém-se os valores das *métricas para cada uma das Q classes*.
- Vejamos um exemplo para $Q = 3$, ou seja, C_1 , C_2 e C_3 .

Matriz de confusão para caso multiclass ($Q > 2$)

Classe C_2 é a positiva.

Classes Preditas	- (C_1)	Verdadeiro Negativo (TN)	Falso Negativo (FN)	Verdadeiro Negativo (TN)
	+ (C_2)	Falso Positivo (FP)	Verdadeiro Positivo (TP)	Falso Positivo (FP)
	- (C_3)	Verdadeiro Negativo (TN)	Falso Negativo (FN)	Verdadeiro Negativo (TN)
		- (C_1)	+ (C_2)	- (C_3)
		Classes Verdadeiras		

Classe C_3 é a positiva.

Classes Preditas	- (C_1)	Verdadeiro Negativo (TN)	Verdadeiro Negativo (TN)	Falso Negativo (FN)
	- (C_2)	Verdadeiro Negativo (TN)	Verdadeiro Negativo (TN)	Falso Negativo (FN)
	+ (C_3)	Falso Positivo (FP)	Falso Positivo (FP)	Verdadeiro Positivo (TP)
		- (C_1)	- (C_2)	+ (C_3)
		Classes Verdadeiras		

Precisão versus *recall*

$$\text{precisão} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

$$\text{recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

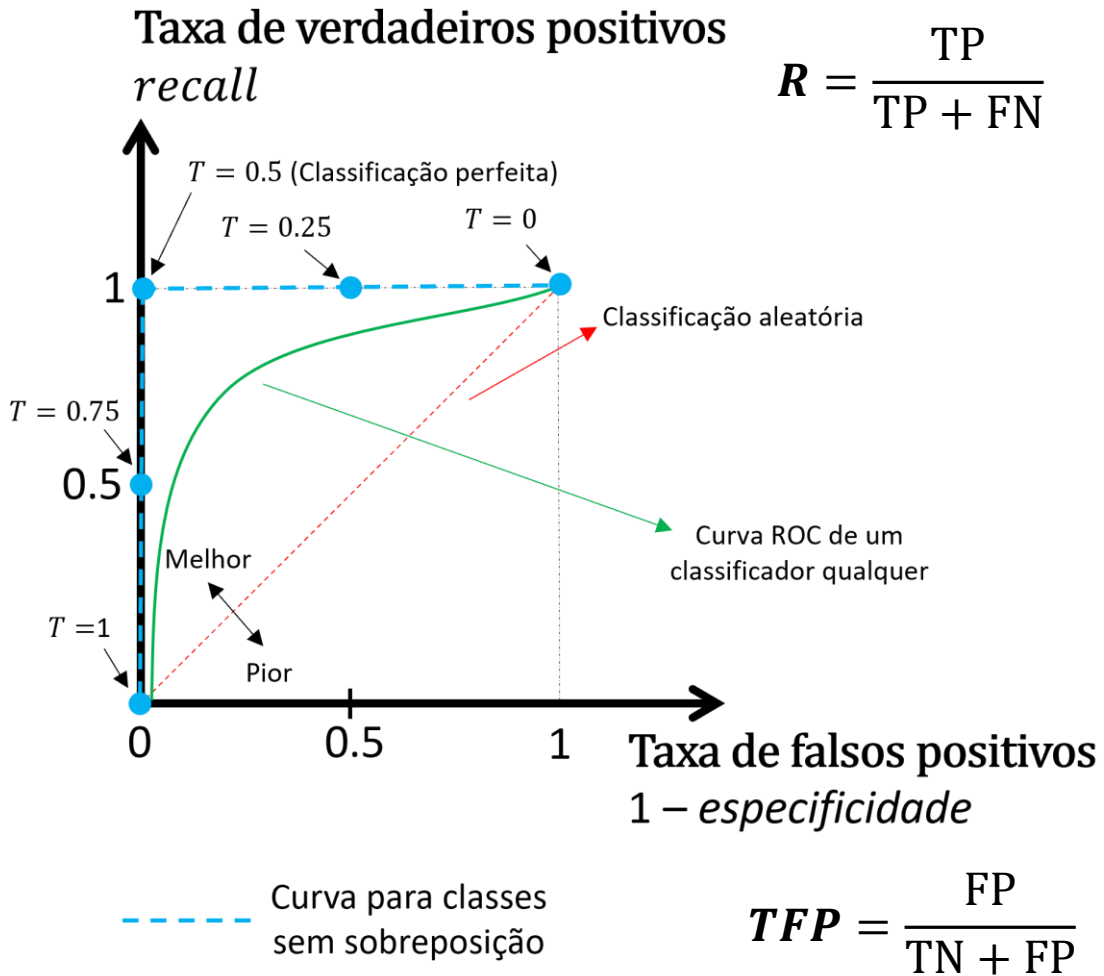
- A **precisão** não fornece informações a respeito da quantidade de **falsos negativos**.
- Por outro lado, o **recall** não fornece informações a respeito da quantidade de **falsos positivos**.
- Mas e se os **custos associados a falsos positivos e negativos são iguais**?
- Nesse caso, para analisarmos o desempenho de um classificador, precisamos de uma métrica que **combine as duas**.

F1-score

- O F1-score combina as duas métricas em uma única.
- A métrica dá a mesma importância a **precisão** e para o **recall**.
- Valores de F_1 próximos de 1 indicam que o **classificador** obteve ótimos resultados tanto de **precisão** quanto de **recall**.
- Em outras palavras, $F_1 \approx 1$ significa que FN e FP ≈ 0 .

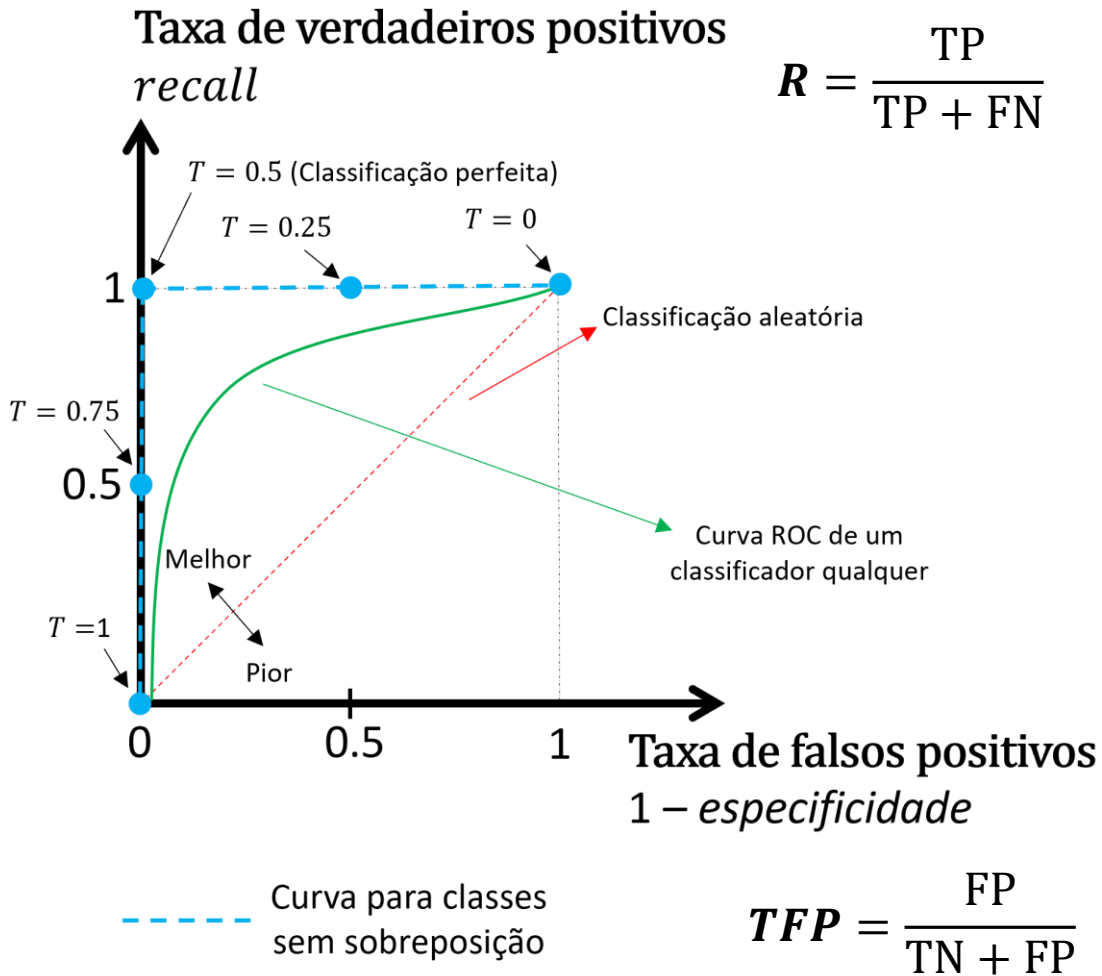
$$F_1 = \frac{TP}{TP + \frac{FN + FP}{2}} = 2 \frac{\text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}}.$$

Curva característica de operação do receptor



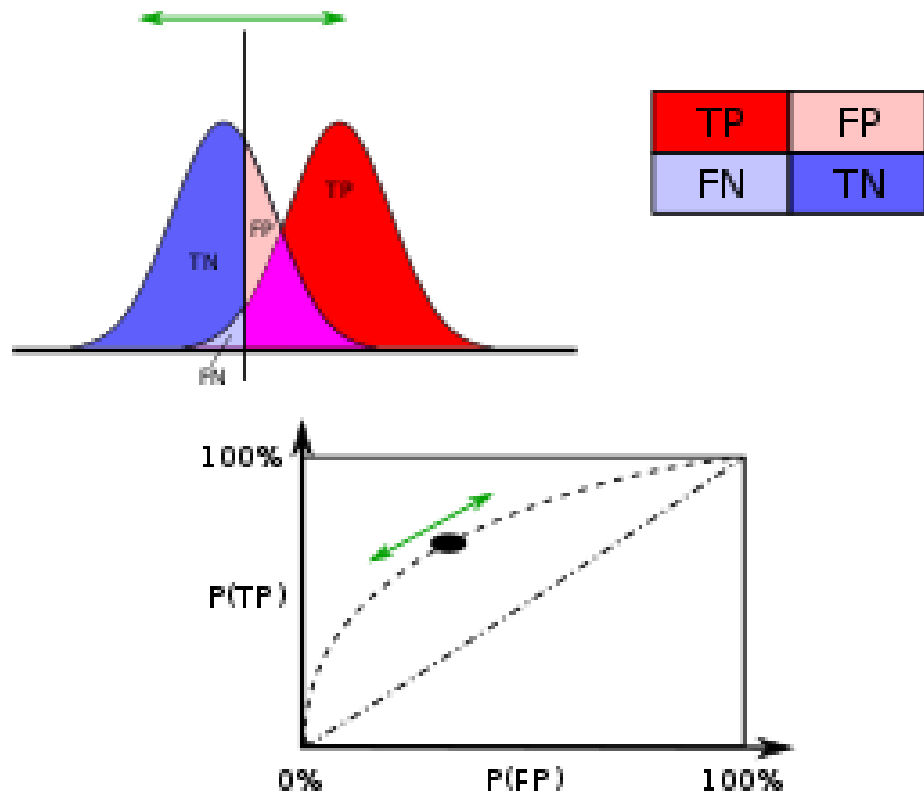
- Também conhecida como curva ROC.
- É um gráfico que mostra o desempenho de um **classificador binário** conforme seu **limiar de quantização (ou discriminação)**, T , é variado.
- A curva é criada plotando-se o **recall** em função da **taxa de falsos positivos** para **vários valores de T** .
- Quanto **mais à esquerda e para cima** estiver a **curva ROC** de um **classificador**, **melhor será o seu desempenho**.

Curva característica de operação do receptor



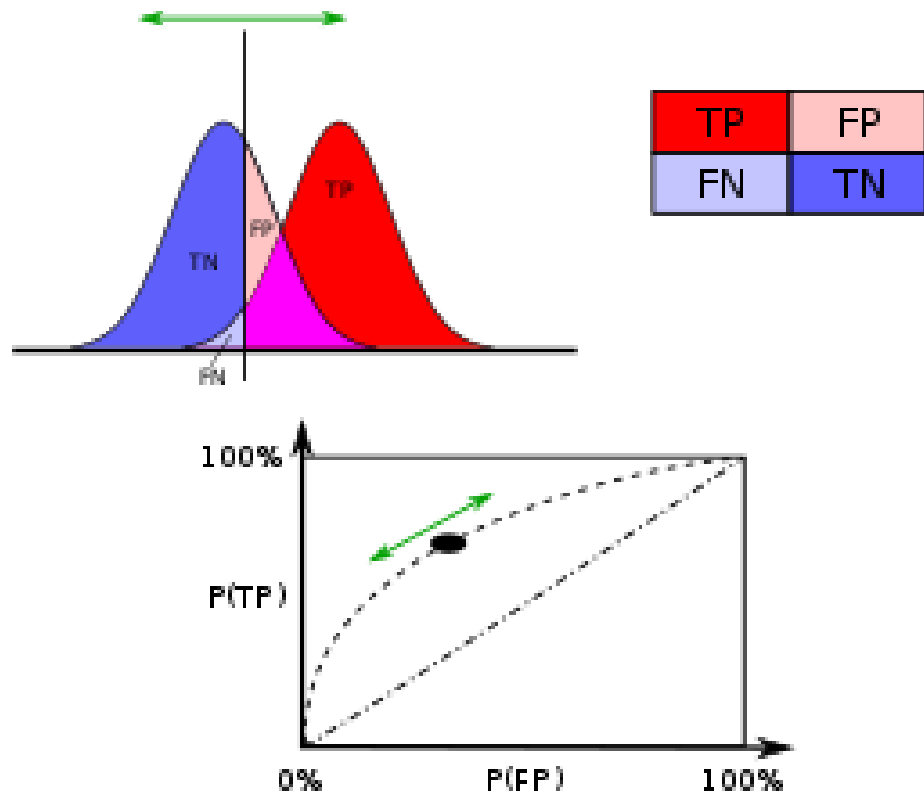
- A linha em vermelho, indica um **classificador puramente aleatório**.
 - Um bom **classificador** fica o mais à esquerda possível dessa linha.
- Classes **sem sobreposição** apresentam uma curva ROC **paralela aos eixos do recall e da TFP** (linha azul tracejada).
- Classes sem sobreposição têm classificação perfeita quando $T = 0.5$, representando 100% de **recall** (i.e., sem falsos negativos) e 100% de **especificidade** (i.e., sem falsos positivos).

Curva característica de operação do receptor



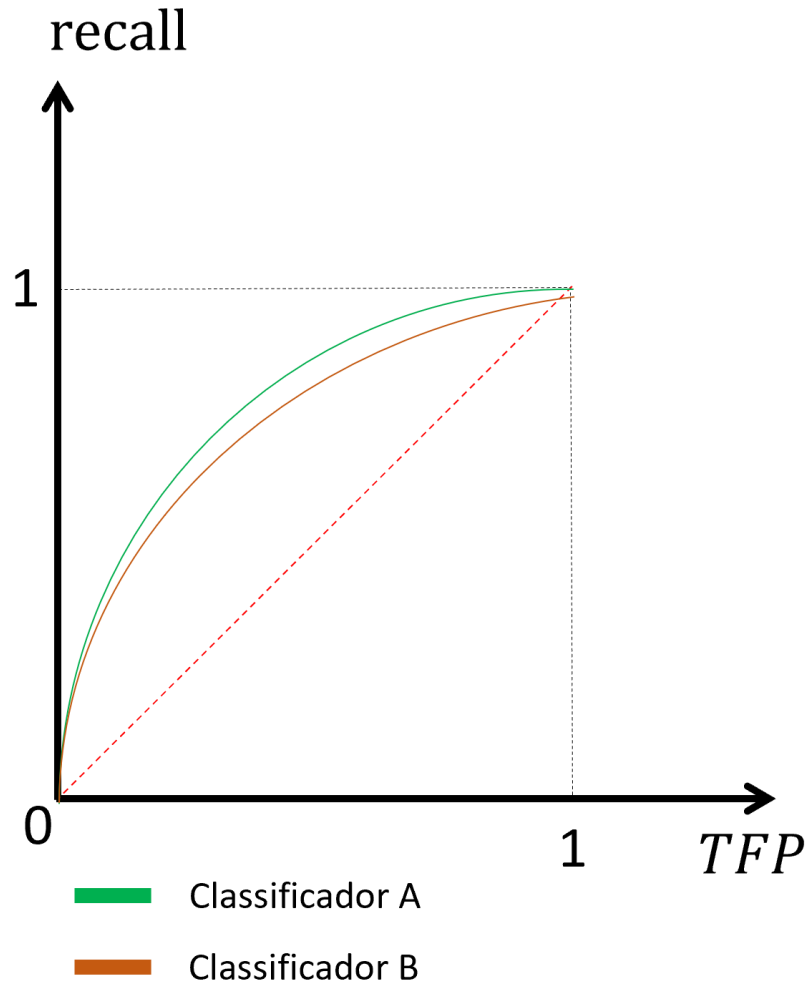
- Em geral, *classificadores binários apresentam em sua saída a probabilidade* para um exemplo de entrada.
- Em seguida, essas *probabilidades são discretizadas* para que se tenha a decisão final.
- Por exemplo, se o valor de $h_a(x(i))$ ultrapassa um determinado **limiar**, T , ele é mapeado no valor 1 (classe positiva, C_2); caso contrário, ele é mapeado no valor 0 (classe negativa, C_1).

Curva característica de operação do receptor



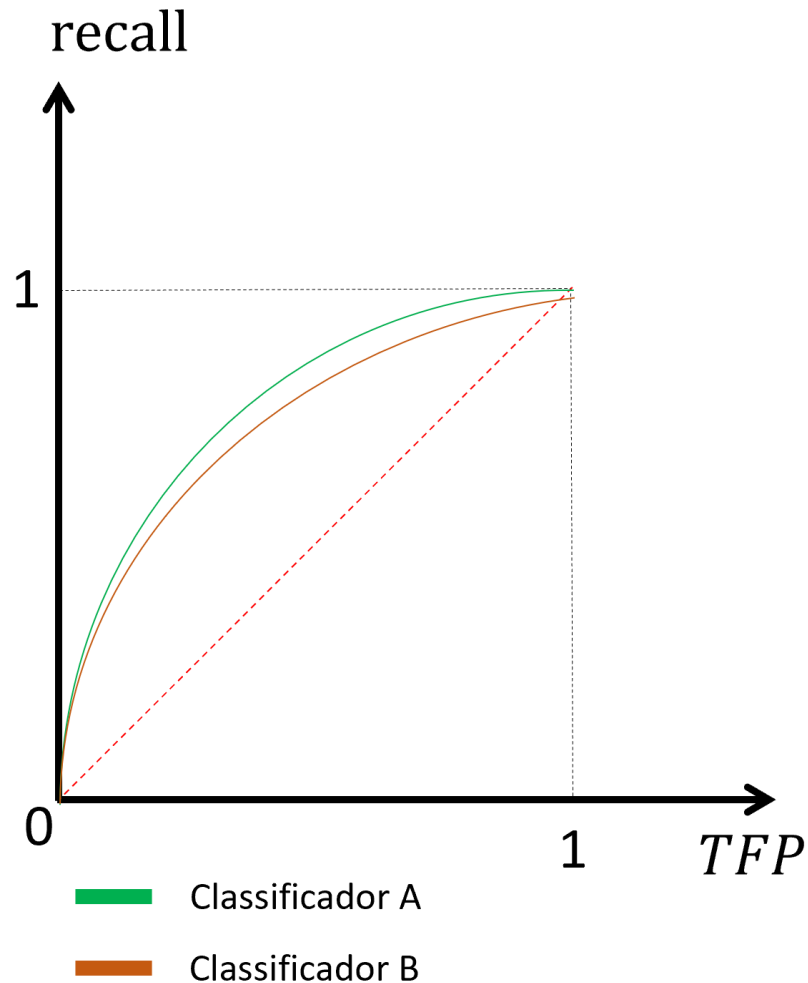
- Sendo assim, ao plotarmos a ***taxa de verdadeiro positivo*** (ou ***recall***) versus a ***taxa de falso positivo*** para diferentes valores de ***limiar, T***, obtemos a ***curva ROC*** associada a um ***classificador binário***.
- A ***escolha do limiar de decisão diferente de 0.5 pode melhorar o desempenho de um classificador*** binário em situações de desbalanceamento de classes, custos assimétricos de erro, preferência por precisão ou ***recall***, etc.

Comparando classificadores com a curva ROC



- Comparamos **classificadores binários** criando **curvas ROC** para cada um deles.
- Por exemplo, considerem as **curvas ROC** na figura ao lado.
- Para decidir qual o melhor **classificador**, podemos analisar a **área sob a curva (ASC) ROC**.
- **ASC** é outra métrica da qualidade de um classificador.
- É um número entre 0 e 1. Quanto maior a **ASC**, melhor será o classificador.

Comparando classificadores com a curva ROC



- O **classificador A** tem melhor desempenho, pois tem **ASC** maior do que a do **classificador B**.
- **Vantagens da curva ROC**
 - Possibilita a análise de diferentes métricas de desempenho independente do **limiar de quantização**.
 - Auxilia o estudo de diferentes **limiares** para lidar com problemas de **desbalanceamento** nos dados.
- **Desvantagens**
 - Usada em problemas de **classificação binária**.
 - No caso **multiclasses**, devemos utilizar as estratégias **um-contra-o-resto** ou **um-contra-um** e plotar várias **curvas ROC**.

Tarefas

- **Quiz:** “*T320 - Quiz - Classificação (Parte V)*” que se encontra no MS Teams.
- **Exercício Prático:** [Laboratório #5](#).
 - Pode ser baixado do MS Teams ou do GitHub.
 - Pode ser respondido através do link acima (na nuvem) ou localmente.
 - [Instruções para resolução e entrega dos laboratórios](#).
 - **Atividades podem ser feitas em grupo, mas as entregas devem ser individuais.**

Obrigado!

verdadeiro positivo



falso positivo



falso negativo



verdadeiro negativo

