

ANA RAQUEL

POSTECH

DATA ANALYTICS

MACHINE LEARNING AVANÇADO

AULA 07

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	4
HANDS ON	8
O QUE VOCÊ VIU NESTA AULA?	9
REFERÊNCIAS.....	10
PALAVRAS-CHAVE	11

O QUE VEM POR AÍ?

Você aprendeu, na aula passada, as principais métricas de validação de algoritmos de Machine Learning. Agora, você irá aprender mais duas poderosas métricas de validação: as curvas ROC (Receiver Operating Characteristic) e AUC (Area under the ROC curve).

CONHEÇA SOBRE O ASSUNTO

Vamos, inicialmente, entender a curva ROC, que significa “Receiver Operating Characteristic”, pois a AUC é uma derivada da curva ROC. A curva ROC é uma ferramenta muito comum, utilizada com classificadores binários. As curvas ROC só podem ser usadas para avaliar classificadores que retornam uma pontuação (ou probabilidade) de previsão entre as classes. A curva ROC plota a taxa de verdadeiros positivos versus a taxa de falsos positivos, então podemos dizer que a curva ROC plota a **sensibilidade** do modelo (revocação) versus a **especificidade**.

Antes de plotar a curva ROC, é necessário ter em mãos os verdadeiros positivos e os falsos positivos. Quanto mais próximo de 1, e consequentemente longe de 0.5, melhor o ajuste da curva ROC. Vamos ver como funciona com um exemplo no Python:

```
from sklearn import metrics
```

```
from sklearn.metrics import roc_curve, auc
```

```
y_prob = modelo_classificador.predict_proba(x_test_escalonado)[: , 1]
```

```
false_positive_rate, true_positive_rate, thresholds = roc_curve(y_test,  
y_prob)
```

```
roc_auc = auc(false_positive_rate, true_positive_rate)
```

```
roc_auc
```

```
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(10,10))

plt.title('Receiver Operating Characteristic')

plt.plot(false_positive_rate,true_positive_rate, color='red',label =
'AUC = %0.2f' % roc_auc)

plt.legend(loc = 'lower right')

plt.plot([0, 1], [0, 1],linestyle='--')

plt.axis('tight')

plt.ylabel('True Positive Rate')

plt.xlabel('False Positive Rate')
```

```
false_positive_rate, true_positive_rate, thresholds = roc_curve(y_test, y_prob)
roc_auc = auc(false_positive_rate, true_positive_rate)
roc_auc
```

```
0.8769230769230769
```

Figura 1 - Exemplo de ROC AUC
Fonte: Elaborado pela autora (2023)

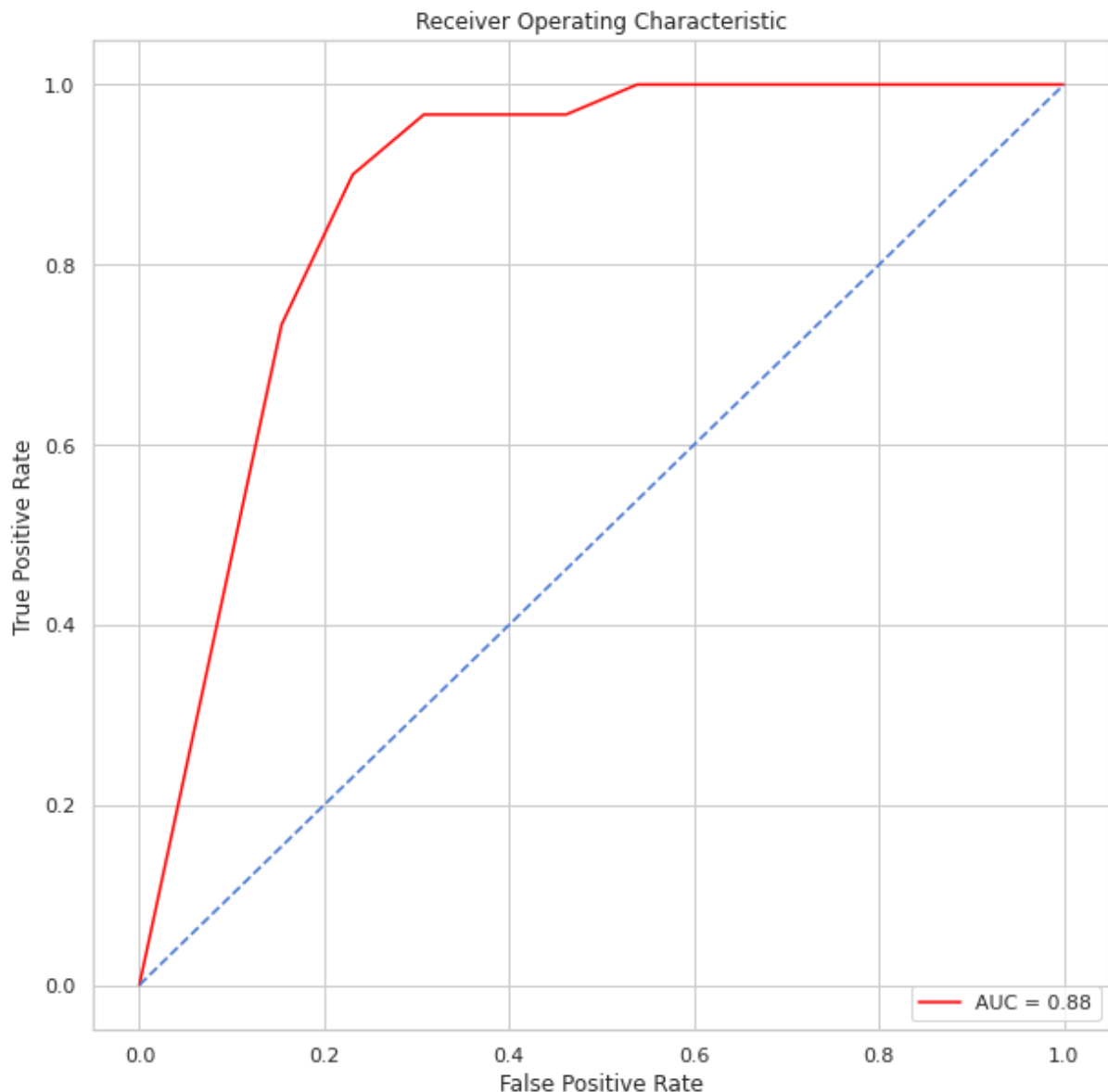


Figura 2 - Plotando a curva ROC AUC
Fonte: Elaborado pela autora (2023)

Observe que, no modelo executado, a curva ROC está distante do limiar (threshold 0.5), ou seja, podemos concluir que o modelo possui um equilíbrio entre sensibilidade e especificidade. Logo, podemos afirmar que o modelo é preciso em ranquear exemplos positivos acima de exemplos negativos. Quanto maior o valor da curva ROC, melhor o modelo está ranqueando. Quando temos um modelo muito próximo de 0,5m significa que o modelo não possui um poder preditivo.

Uma maneira de comparar classificadores é medir a área abaixo da curva AUC ("area under the ROC curve"). Um classificador perfeito terá uma ROC AUC igual a 1,

enquanto um classificador puramente aleatório terá um ROC AUC Igual a 0,5. A AUC resume a curva ROC agregando todos os limiares da ROC, calculando a “área sob a curva”.

HANDS ON

Agora chegou o momento de ver, na prática, como podemos validar um modelo utilizando as curvas ROC e AUC. Vamos nessa?

Para essa aula, temos um notebook para você. Acesse abaixo:

- [Notebook 1](#)

Além disso, também disponibilizamos a base de dados, para te ajudar com os estudos e exercícios.

- [Base de dados 1](#)

O QUE VOCÊ VIU NESTA AULA?

Curva ROC e AUC.

Daqui em diante, é importante que você replique os conhecimentos adquiridos para fortalecer ainda mais suas bases e conhecimentos.

IMPORTANTE: não esqueça de praticar com o desafio da disciplina, para que, assim, você possa aprimorar seus conhecimentos!

Você não está só nesta jornada! Te esperamos no Discord e nas *lives* com especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

REFERÊNCIAS

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**. 2. ed. [s.l.]: O'Reilly Media, Inc., 2019.

SCIKIT-LEARN. **Documentação Scikit-Learn**. 2023. Disponível em: <<https://scikit-learn.org/stable/>>. Acesso em: 13 abr. 2023.

PALAVRAS-CHAVE

Palavras-chave: Curva ROC. Curva AUC.

The background is a dark blue field filled with numerous small, light blue dots. Overlaid on this are several large, wavy, translucent lines in shades of blue, yellow, and red. These lines flow from the left side towards the right, creating a sense of movement. Various geometric shapes are scattered throughout: a thin vertical line on the left, a circle containing the number '7' in the upper center, a small circle on the left, an 'X' mark near the bottom left, a small circle near the bottom left, and a hexagon in the bottom right corner.

POSTECH