



# Métricas de avaliação II: ROC Curve

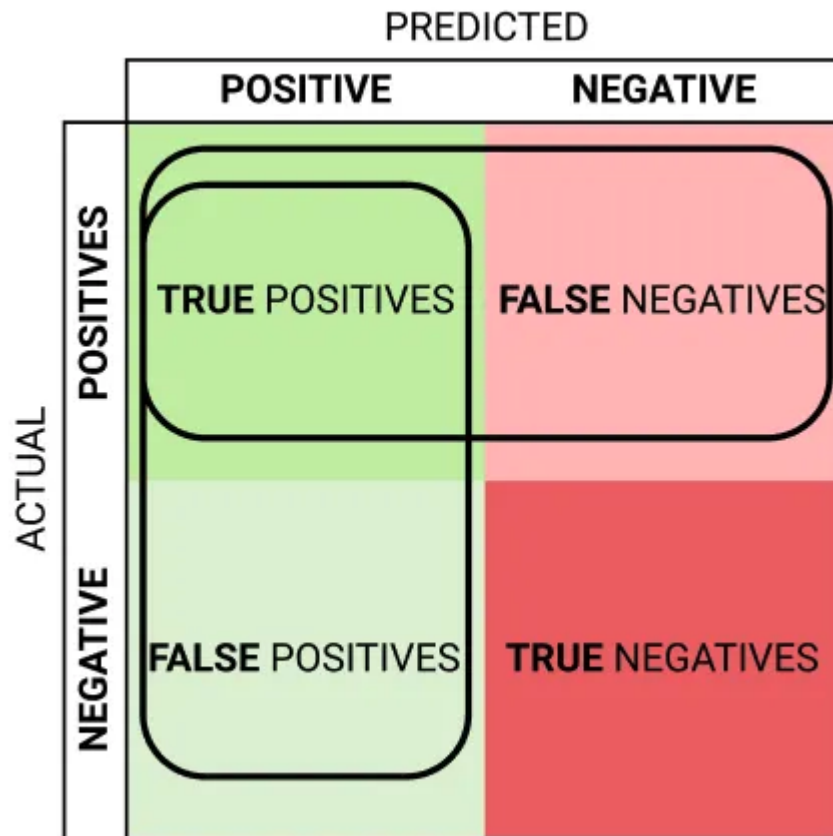
☰ Ciclo	Ciclo 06: Algoritmos baseado em árvores
# Aula	45
🕒 Created	@March 10, 2023 4:55 PM
☑ Done	☑
☑ Ready	☑

## Objetivo da Aula:

- ☐ Revendo a Matriz de confusão
- ☐ Resumo
- ☐ Próxima aula

## Conteúdo:

- ▼ **1. Revendo a matriz de confusão**
  - ▼ **1.1 Matriz de confusão**



### ▼ 1.2 True Positive Rate ( TPR )

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

### ▼ 1.3 False Positive Rate ( FPR )

$$FPR = \frac{FP}{FP + TN}$$

## ▼ 2. Curva ROC

A curva ROC (Receiver Operating Characteristic) é uma curva que representa a habilidade de um modelo em distinguir entre classes positivas e negativas variando o threshold de classificação.

A curva ROC é construída plotando a taxa de verdadeiros positivos (True Positive Rate - TPR) em função da taxa de falsos positivos (False Positive Rate - FPR) para diferentes valores de threshold. A área sob a curva ROC (AUC) é uma métrica utilizada para comparar diferentes modelos.

## ▼ 2.1 Curva ROC no Python

```
# Imports
from sklearn import datasets      as dt
from sklearn import linear_model  as lm
from sklearn import model_selection as ms
from sklearn import metrics       as mt
from matplotlib import pyplot     as plt

# Load dataset
X, y = dt.make_classification( n_samples=1000,
                              n_classes=2, random_state=1)

# split into train/test sets
X_train, X_test, y_train, y_test = ms.train_test_split(X, y, test_size=0.5, random_state=2)

# define a model e fit
model = lm.LogisticRegression(solver='lbfgs')
model.fit(X_train, y_train)

# predict probabilities
yhat_test = model.predict_proba(X_test)
yhat_test = yhat_test[:, 1]

# calculate roc curve
fpr, tpr, thresholds = mt.roc_curve(y_test, yhat_test)

# calculate AUC
roc_auc = auc(fpr, tpr)

# plot the roc curve for the model
plt.plot(fpr, tpr, marker='.', label='Model(AUC = %0.2f)' % roc_auc)

# axis labels
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')

# show the legend
plt.legend()
plt.show()
```

## ▼ 2.2 Curva Threshold, TPR x FPR no Python

```
# The optimal threshold is highest TPR e lowest FPR
plt.plot( thresholds, tpr, 'b--', label='TPR' );
plt.plot( thresholds, 1-fpr, 'g-', label='FPR' );
plt.xlabel( 'True Positive Rate' );
plt.ylabel( 'False Positive Rate' );

plt.legend()
plt.grid()
```

---

## ▼ 3. Resumo

1. A curva ROC mede a habilidade de separação do algoritmo em 2 classes.
2. A métrica AUC é o valor da área abaixo da curva, quanto maior esse valor, maior a performance de separação do algoritmo.

## ▼ 4. Próxima aula