



Métricas de avaliação I: Matriz de confusão e acurácia

≡ Ciclo	Ciclo 02: Aprendizado supervisionado - Classificação
# Aula	14
🕒 Created	@January 5, 2023 4:41 PM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ A matriz de confusão
- ☐ Acurácia
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. A matriz de confusão

A matriz de confusão é uma ferramenta para medir a performance de um algoritmo de Machine Learning ao realizar uma classificação.

A ideia geral é contar o número de vezes em que os exemplos da classe A são classificados erroneamente como classe B.

▼ 1.1 Matrix de confusão

▼ 1.1.1 Exemplo:

O conjunto de dados possui 2 classes de fotos, a foto de cachorros e gatos. O algoritmo de Machine Learning foi treinado para classificar corretamente as imagens de cachorro como classe "Cachorro" e de gato como "Gato".

A matriz de confusão revela a performance do algoritmo.

▼ 1.1.2 Matriz de confusão

		Previsão do Rótulo	
		Classe A	Classe B
Rótulo Real	Classe A	5	2
	Classe B		

	Classe B	3	3
--	----------	---	---

A interpretação é a seguinte algoritmo fez 13 previsões:

- 5 previsões como classe A, sendo que era realmente classe A.
- 3 previsões como classe A, sendo que era classe B.
- 2 previsões como classe B, sendo que era classe A.
- 3 previsões como classe B, sendo que era class B.

▼ 1.1.3 Denominações da matriz de confusão

		Previsão do Rótulo	
		Classe Positiva	Classe Negativa
Rótulo Real	Classe Positiva (P)	True Positive (TP)	False Negative (FN)
	Classe Negativa (N)	False Positive (FP)	True Negative (TN)

- **Total de previsões** = P + N
- **True Positive** = Classificação correta da classe positiva
- **True Negative** = Classificação correta da classe negativa
- **False Positive** = Classificação errada da classe positiva
- **False Negative** = Classificação errada da class negativa

▼ 2. Acurácia

Acurácia é a quantidade de previsões acertadas da Classe A e da Classe B.

Em outras palavras, quantas vezes o algoritmo classificou uma observação como Classe A, sendo que ela era realmente da Classes A, mais a quantidade de vezes que o classificador classificou uma observação como Classe B.

Olhando para a matrix de confusão, a acurácia é a soma dos acertos tanto da Classe A quanto da classe B, representado pela diagonal principal, dividido pelo total de classificações realizadas, que é a soma de todas as células.

$$Acurácia = \frac{TP + TN}{P + N}$$

▼ 2.1 O problema da acurácia

O conjunto de dados no qual o classificador **foi treinando é desbalanceado**. A classe A possui 100 exemplos e a classe B possui 20 exemplos.

		Previsão	
		Classe A	Classe B
Rótulo Real	Classe A (98)	90	8
	Classe B (12)	2	10

$$\text{Acurácia} = (90 + 10) / (90 + 8 + 2 + 10) = 100 / 110 = 91\%$$

A acurácia atribui o mesmo peso para qualquer classificação. Portanto, o valor da acurácia para problemas com classes desbalanceadas é dominado pelo acerto da classe majoritária (com mais exemplos)

▼ 3. Métricas na prática

```
import pandas as pd
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics as mt

# 1.0 Load dataset
df = pd.read_csv( '../dataset/train.csv' )

# 2.0 Seleção de Features
features = ['idade', 'saldo_atual', 'divida_atual', 'renda_anual', 'valor_em_investimentos',
            'taxa_utilizacao_credito', 'num_emprestimos', 'num_contas_bancarias', 'num_cartoes_credito',
            'dias_atraso_dt_venc', 'num_pgto_atrasados', 'num_consultas_credito', 'taxa_juros']

label = ['limite_adicional']

x_train = df.loc[:, features]
y_train = df.loc[:, label].values.ravel()

# 3.0 Treinamento do KNN
knn_classifier = KNeighborsClassifier( n_neighbors = 8 )
knn_classifier.fit( x_train, y_train )

y_pred = knn_classifier.predict( x_train )
# 4.0 Performance

# 4.1 Confusion Matrix
mt.confusion_matrix( y_train, y_pred )

# 4.2 Accuracy
mt.accuracy_score( y_train, y_pred )
```

▼ 4. Resumo

1. A matriz de confusão é uma matriz que mapeia os acertos e os erros de um algoritmo de Machine Learning que realiza uma tarefa de classificação.
2. A acurácia é a quantidade de acertos que o classificador realizou tanto da classe A quanto da classe B, dividido pelo total de classificações que ele realizou.

▼ 5. Próxima aula

Métricas de avaliação II: Recall e Precision