



Métricas de avaliação III: RMSE

≡ Ciclo	Ciclo 03: Aprendizado supervisionado - Regressão
# Aula	24
🕒 Created	@January 30, 2023 2:51 PM
☑ Done	<input type="checkbox"/>
☑ Ready	<input checked="" type="checkbox"/>

Objetivo da Aula:

- ☐ A reta de regressão
- ☐ O erro RMSE
- ☐ Vantagens e desvantagens do RMSE
- ☐ RMSE na prática
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. A reta de regressão

▼ 2. O erro RMSE

RMSE (Root Mean Square Error) calcula a raiz quadrada do erro médio quadrático (MSE) entre as previsões e os valores reais.

O valor do RMSE está na mesma unidade de medida da variável resposta, o que faz a interpretação do erro ser direta. Essa métrica de erro atribui um erro maior para previsões com altos valores de erro.

▼ 2.1 Fórmula

A raiz quadrada do erro médio quadrático, conhecido como RMSE, é calculado da seguinte forma:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

▼ 2.2 Exemplo:

Posição	Altura atual (y)	Altura predita (\hat{y})	Error (y - \hat{y})	Error (y - \hat{y}) ^2	((y - \hat{y}) / y) ^2
1	188	180	8	64	(0.04) ^ 2 = 0.0018
2	180	160	20	400	
3	175	170	5	25	
4	148	150	-2	4	
5	203	200	3	9	
6	184	190	-6	36	
7	150	140	10	100	
				MSE = 91,14 m2	
	Média = 170			RMSE = 9,54m	
				RMSPE = 5%	

O erro médio ponderado entre as previsões e os valores reais neste conjunto de dados é 9.55, o que provavelmente é um bom valor, uma vez que a altura real média no conjunto de dados é 170.

▼ 2.2.1 Como interpretar o RMSE

O RMSE é uma medida de performance do modelo na mesma escala da variável alvo. O RMSE pode ser interpretado como o erro médio que as previsões do modelo tem com os dados reais, sendo que o erro está na mesma escala.

▼ 2.2.1.1 Por exemplo:

Um valor de RMSE de R\$1.000 para a previsão do preço de venda de uma casa parece bom, uma vez que os preços das casas tendem a ser maior do que R\$ 100.000.

Entretanto, o mesmo RMSE de R\$ 1.000 para a previsão de vendas de um computador entre R\$ 800 e R\$ 5.000 é terrível .

▼ 3. Vantagens e Desvantagens do RMSE

▼ 3.1 Vantagens:

1. Atribui um peso maior para grandes erros.
2. Apresenta a mesma unidade de medida da variável resposta.

▼ 3.2 Desvantagens:

1. Não é robusto na presença de outliers.

▼ 4. RMSE na prática

```

import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn import metrics as mt

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

# 2.0 Seleção de Features
features = ['idade', '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 = ['saldo_atual']

x_train = df.loc[:, features]
y_train = df.loc[:, label]

# 3.0 Model Training
lr_model = LinearRegression()
lr_model.fit( x_train, y_train )

y_pred = lr_model.predict( x_train )

df1 = df.loc[:, ['id_cliente', 'saldo_atual']]
df1['predicted'] = y_pred

# 4.0 Model Performance
## 4.1 R squared
r2_squared = np.round( 100*mt.r2_score( y_train, y_pred ), 2 )
print( 'R2 square: {}'.format( r2_squared ) )

## 4.2 MSE
mse = np.round( mt.mean_squared_error( y_train, y_pred ), 2 )
print( 'A cada previsão, o erro médio é de: US{}'.format( mse ) )

## 4.3 RMSE
rmse = np.sqrt( mse )
print( 'A cada previsão, o erro médio é de: US{}'.format( rmse ) )

# 5.0 Conclusão
print( '{}% da variação da variável alvo y é reduzida, levando em consideração o preditor'.format( r2_squared ) )
print( '{}% da variação da variável alvo y é "explicada pela variação do preditor x'.format( r2_squared ) )

```

▼ 5. Resumo

1. RMSE apresenta o erro médio na mesma unidade de medida da variável resposta. Portanto, facilita a interpretação.
2. RMSE ainda atribui um peso maior aos grandes erros de previsão.

▼ 6. Próxima aula

Exercícios