

## Questões de Regressão:

Davi Ventura C. Perdigão - 82148

### Questão 1:

O que é o R quadrado (R<sup>2</sup>) e qual a sua importância na avaliação de modelos de regressão? Como ele é calculado e interpretado?

R: O R quadrado (R<sup>2</sup>) é uma métrica que avalia a qualidade de um modelo de regressão, ou seja, a quantidade da variância dos dados que é explicada pelo modelo construído. O R<sup>2</sup> é importante porque permite avaliar se o modelo escolhido é adequado para explicar a relação entre as variáveis.

O cálculo do R<sup>2</sup> é baseado na comparação da variância dos valores observados com a variância dos valores previstos pelo modelo:

- $\hat{y}$  representa o valor predito,
- $\bar{y}$  representa o valor médio das amostras
- $y$  representa o valor real.

$$R^2 = 1 - \frac{\text{Variança Residual}}{\text{Variança Total}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

OU -> **R<sup>2</sup> = 1 - (SE / ST)**, onde SE é a soma dos quadrados dos erros residuais (ou seja, a diferença entre os valores observados e os valores previstos pelo modelo), e ST é a soma dos quadrados totais (ou seja, a diferença entre cada valor observado e a média dos valores observados).

A interpretação desse resultado se dá de 0 a 1, sendo que quanto mais próximo de 1, melhor é o ajuste do modelo aos dados. Por exemplo, um R<sup>2</sup> de 0,8 significa que 80% da variabilidade dos dados é explicada pelo modelo. No entanto, isso não diz respeito a qualidade das previsões individuais do modelo, e pode ser enganoso se o modelo estiver superajustado aos dados.

### Questão 2:

O que é o Erro Quadrático Médio (MSE) e qual a sua importância na avaliação de modelos de regressão? Como ele é calculado e interpretado?

R: O Erro Quadrático Médio (MSE) é a média dos quadrados das diferenças entre os valores previstos pelo modelo e os valores reais observados. Ele pode ser usado para comparar diferentes modelos e ajudar a identificar aqueles que têm melhor desempenho. Além disso, o MSE pode ser facilmente interpretado e comunicado aos usuários finais.

O MSE é calculado pela soma dos quadrados dos erros (ou resíduos), dividida pelo número de observações. Matematicamente, a fórmula para o cálculo do MSE é:

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

n é o número de observações;  
 $\hat{y}_i$  é o valor real observado da variável;  
 $y_i$  é o valor previsto pelo modelo para a observação.

O MSE é interpretado de forma que os valores mais altos indicam que o modelo está fazendo previsões menos precisas. Ou seja, se o MSE for alto, isso significa que o modelo está com dificuldades para prever valores reais.

### Questão 3:

O que é a Raiz do Erro Quadrático Médio (RMSE) e qual a sua importância na avaliação de modelos de regressão? Como ele é calculado e interpretado?

R: A Raiz do Erro Quadrático Médio (RMSE) indica a diferença média entre os valores previstos pelo modelo e os valores reais observados na amostra de dados. O RMSE é uma métrica importante na avaliação de modelos de regressão porque fornece uma medida quantitativa da precisão do modelo.

Matematicamente, a fórmula para o cálculo do RMSE é:

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

n é o número total de observações na amostra;  
 $\hat{y}_i$  é o valor previsto pelo modelo;  
 $y_i$  é o valor real observado.

O RMSE é de forma que quanto menor o valor, melhor é o desempenho do modelo na previsão de valores. Por exemplo, se o valor do RMSE for 2,0, isso indica que o modelo está errando, em média, 2 unidades em suas previsões.

### Questão 4:

O que é o Erro Absoluto Médio (MAE) e qual a sua importância na avaliação de modelos de regressão? Como ele é calculado e interpretado?

R: O Erro Absoluto Médio (MAE) é a média das diferenças absolutas entre os valores previstos pelo modelo e os valores reais. É importante na avaliação de modelos de regressão porque fornece uma medida da magnitude do erro médio do modelo. Em outras palavras, ele mede o quão longe, em média, as previsões estão dos valores reais.

Matematicamente, a fórmula para o cálculo do MAE é:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

n é o número total de observações;

$\hat{y}_i$  é o valor real da variável de resposta;

$y_i$  é o valor previsto da variável de resposta.

O MAE é interpretado de forma que quanto menor o valor, melhor o desempenho do modelo, pois isso indica que as previsões estão mais próximas dos valores reais. No entanto, o MAE não leva em consideração a direção dos erros, o que significa que erros positivos e negativos são tratados da mesma forma. Por esse motivo, o MAE pode ser menos sensível a outliers do que outras medidas de avaliação, como o MSE.

#### Questão 5:

Considere um modelo de regressão linear com os seguintes dados observados:  $y_{\text{true}}$

$= [10, 20, 30, 40, 50, 60, 70, 80, 90, 100]$

$y_{\text{pred}} = [12, 25, 35, 42, 49, 62, 75, 82, 93, 105]$

Utilizando Sklearn, calcule as métricas MSE, RMSE, MAE e  $R^2$  para avaliar o desempenho do modelo.

```
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import numpy as np

y_true = np.array([10, 20, 30, 40, 50, 60, 70, 80, 90, 100])
y_pred = np.array([12, 25, 35, 42, 49, 62, 75, 82, 93, 105])

# MSE - Erro Quadrático Médio
mse = mean_squared_error(y_true, y_pred)
print("MSE:", mse)

# RMSE - Raiz do erro quadrático médio
rmse = np.sqrt(mse)
print("RMSE:", rmse)

# MAE - Erro Absoluto Médio
mae = mean_absolute_error(y_true, y_pred)
print("MAE:", mae)

# R² - R Quadrado
r2 = r2_score(y_true, y_pred)
print("R²:", r2)

MSE: 12.6
RMSE: 3.5496478698597698
MAE: 3.2
R²: 0.9847272727272727
```

R: Segundo as métricas, o modelo parece ter um bom ajuste aos dados observados, com um baixo MSE, baixo RMSE e uma MAE relativamente baixa. O  $R^2$  indica que o modelo explica aproximadamente 98,5% da variação total dos dados.