



Métricas de avaliação I: R2 - Teoria

☰ Ciclo	Ciclo 03: Aprendizado supervisionado - Regressão
# Aula	21
🕒 Created	@January 24, 2023 9:35 AM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ A reta de regressão
- ☐ Os tipos de erros da regressão
- ☐ Os coeficientes de determinação
- ☐ R2 na prática
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. A reta de regressão

1. Reta sem comportamento (média)
2. Reta com comportamento (Regressão)

▼ 2. Os tipos de erros da regressão

▼ 2.1 SSR

O erro SSR é o erro da entre a **reta de regressão estimada**, conseguida através dos ajuste aos dados, e a **reta média** que representa a reta conseguida através da média dos pontos. Esse erro, em inglês, é chamado de “Regression Sum of Square”.

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

▼ 2.2 SSE

O erro SSE é o erro entre **os pontos** do conjunto de dados e a **reta de regressão estimada**. Ela quantifica quanto os dados variam em torno da reta de regressão estimada (modelo). Esse, em inglês, é chamado de “Error Sum of Squares”.

$$SSE = \sum_{i=1}^n (y_i - \hat{y})^2$$

▼ 2.3 SSTO

O erro SSTO é o erro entre **os pontos** do conjunto de dados e a **reta média**. Ela quantifica quanto os dados variam em torno da reta média. Esse, em inglês, é chamado de “Total Sum of Squares”.

$$SSTO = \sum_{i=1}^n (y_i - \bar{y})^2$$

▼ 2.4 O coeficiente de determinação

O coeficiente de determinação, conhecido como R², pode ser calculado pela combinação dos erros da regressão.

$$R^2 = 1 - \frac{SSE}{SSTO}$$

▼ 3. O coeficiente de determinação

O R quadrado mede a quantidade de variação em uma variável resposta ou variável alvo, que pode ser explicado pelas variável independentes.

Em outras palavras, o R quadrado indica o quanto eu consigo explicar o fenômeno observado, representado pela variável resposta, a partir das características coletadas da observação do fenômeno.

O valor do R^2 varia de 0 a 1, sendo que 0 significa que o modelo (conjunto de variáveis e suas relações) não explica nada da variação na variável alvo, enquanto que o valor igual a 1 significa que o modelo explica toda a variação.

▼ 3.1 Característica do R^2

1. R^2 é uma proporção que varia de 0 a 1.
2. Se o R^2 é igual a 1, todos os pontos de dados estão perfeitamente em cima da reta de regressão. Os preditores do fenômeno explicam TODA a variação do alvo.
3. Se o R^2 é igual a 0, significa que a reta de regressão estimada é perfeitamente horizontal. Os preditores do fenômeno explicam ZERO a variação do alvo.

▼ 3.2 Interpretação do R^2

Imagine que o valor do R^2 é de 57%. Podemos interpretar esse valor de duas formas:

1. 57% da variação da variável alvo y é reduzida, levando em consideração o preditor x .
2. 57% da variação da variável alvo y é “explicada pela variação do preditor x ”.

Cuidado com a interpretação 2: **Correlação não é Causalidade.**

▼ 4. Resumo

1. Na análise de regressão, podemos calcular 3 tipos de erros: SSR, SSE e SSTO
2. O R quadrado mede a quantidade de variação em uma variável resposta, que pode ser explicado pelas variáveis características.

▼ 5. Próxima aula

Métricas de avaliação I: R^2 - Prática