05

Estimando um modelo de regressão linear

Transcrição

Temos uma função de regressão para três variáveis explicativas neste projeto, mas podemos ter mais ou menos. Temos Y, nossa variável dependente para consumo de cerveja, beta 1 que caracteriza o momento em que a nossa reta de regressão corta o eixo Y, os betas restantes são os coeficientes angulares parciais e o X as variáveis explicativas, o U é o termo de erro, isto é, o quanto da variação de Y não conseguiu ser explicada pela nossa equação.

$$Yi = \beta1 + \beta2X2i + \beta3Xei + \beta4X4i + Ui$$

Importaremos as ferramentas necessárias da biblioteca do sckit-learn. São elas `linear_model`, `LinearRegression`, `metrics`. Lembrando que todas as etapas estão organizadas no notebook disponibilizado.

Instanciaremos a classe `LinearRegression()`. Em seguida utilizaremos o método `fit()`, que nos auxiliará a saber os passos a serem trilhados. Esse método precisará receber o os dados x e y de treino. Em seguida, calcularemos o coeficiente de determinação, uma medida resumida do grau de ajuste da regra de digressão.

```
print('R² = {}'.format(modelo.score(X_train, y_train).rou
```

COPIAR CÓDIGO

A medida resumida trabalha com valores entre 0 e 1, então quanto mais próximo de 1 melhor. Podemos testá-lo e gerar previsões para este modelo, o

objetivo de todo o processo. Teremos as observações das variáveis explicativas e conseguiremos fazer previsões da variável dependente. Criaremos uma variável chamada `y_previsto` que receberá `modelo`. Em seguida evocaremos o método `predict()` pera realizar de fato a previsão, utilizando como parâmetro X de teste.

Para conseguirimos explicar os resultados que serão obtidos na previsão, utilizaremos o `metrics()`, e o `r2_score()`. Precisaremos passar os parâmetros `y_test_` e `y_previsto`. Ao final, teremos a seguinte estrutura:

```
print('R² = %s' % metrics.r2_score(y_test, y_previsto).ro
```

COPIAR CÓDIGO

Teremos como resultado o valor `R² = 0.69`. Ao aumentarmos o número de variáveis as estatísticas poderão ser melhoradas e assim teremos um modelo mais eficiente.