Linear Regression

IBM

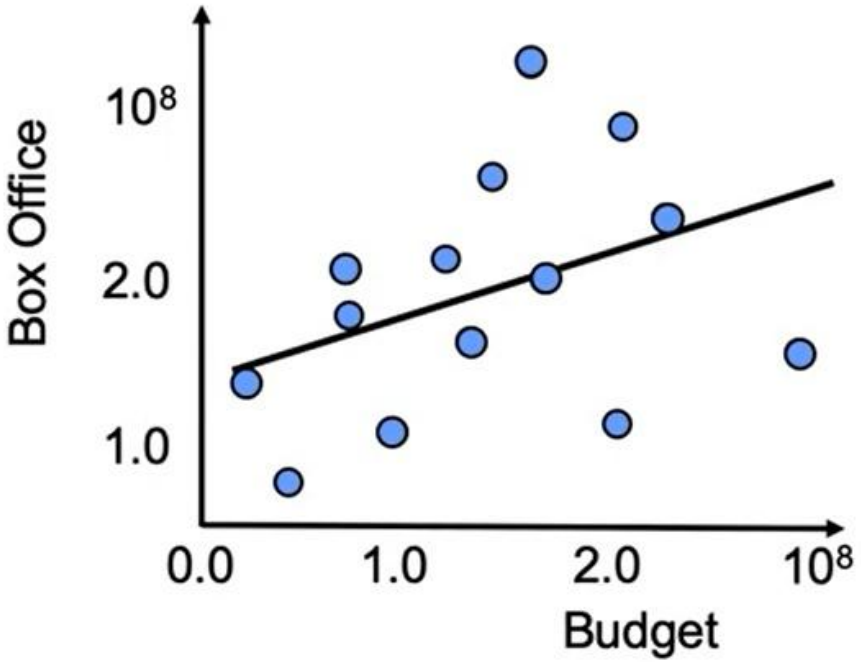
Learning Goals

In this section, we will cover:

- Linear Regression
- Modeling Best Practice
- Measuring Errors

IBM

Introduction to Linear Regression



$$y_\beta(x) = \beta_0 + \beta_1 x$$

box office revenue

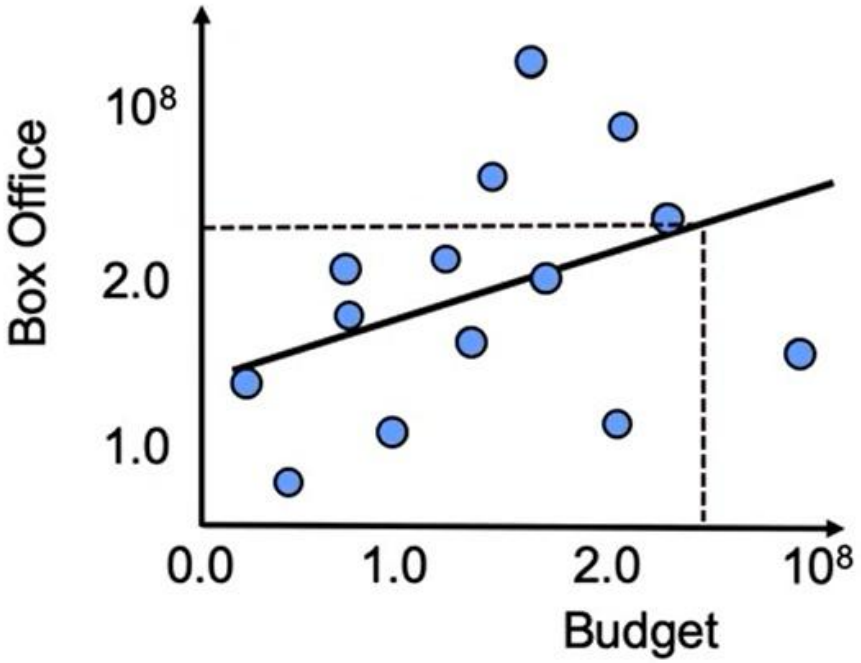
coefficient 0

coefficient 1

movie budget

IBM

Introduction to Linear Regression



$$y_\beta(x) = \beta_0 + \beta_1 x$$

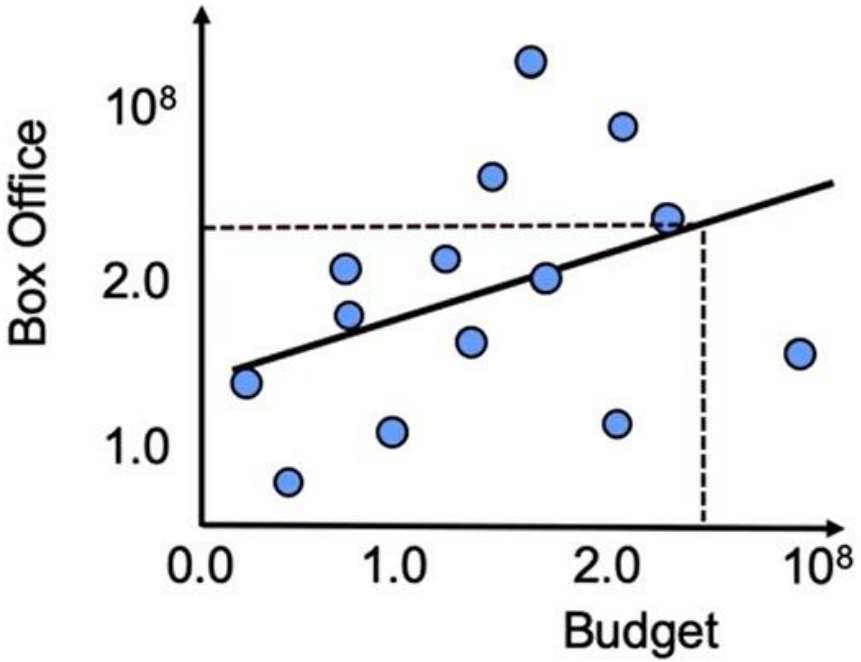
box office revenue

80 million

movie budget

0.6

Introduction to Linear Regression



$$y_\beta(x) = \beta_0 + \beta_1 x$$

175 million

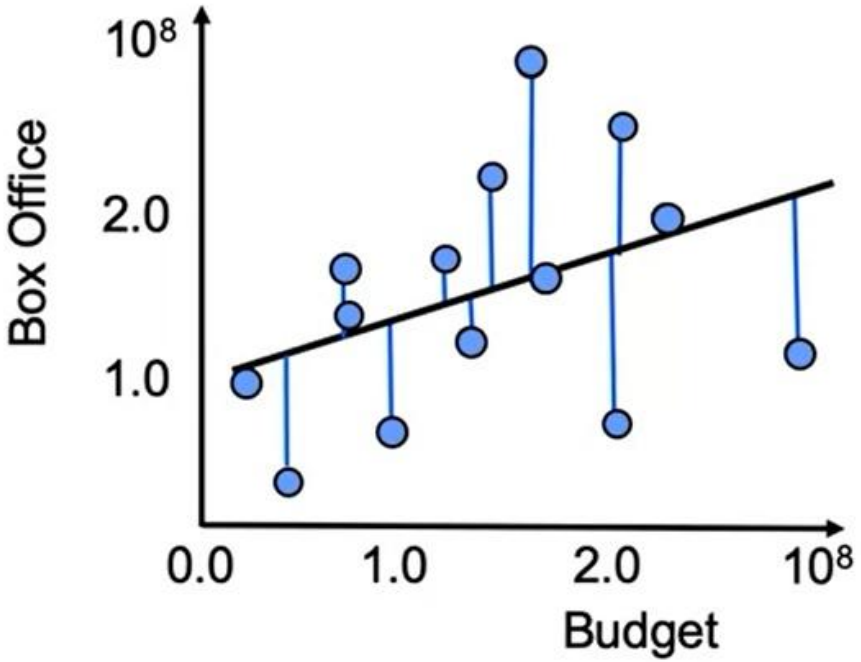
80 million

0.6

160 million

IBM

Calculating the Residuals



$$y_\beta\left(x_{obs}^{(i)}\right) - y_{obs}^{(i)}$$

predicted value

observed value

IBM

Minimizing the Error Function

$$\min_{\beta_0,\beta_1} \frac{1}{m} \sum_{i=1}^{m} \left(\left(\beta_0 + \beta_1 x_{obs}^{(i)} \right) - y_{obs}^{(i)} \right)^2$$

IBM

Minimizing the Error Function

$$J(\beta_0, \beta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\left(\beta_0 + \beta_1 x_{obs}^{(i)} \right) - y_{obs}^{(i)} \right)^2$$

IBM

Modeling Best Practice

Use cost function to fit model

Develop multiple models

Compare results and choose best one

IBM

Other Measures of Error

Sum of Squared Error (SSE): $$\sum_{i=1}^{m} \left(y_\beta(x^{(i)}) - y_{obs}^{(i)} \right)^2$$

Total Sum of Squares (TSS): $$\sum_{i=1}^{m} \left(\overline{y_{obs}} - y_{obs}^{(i)} \right)^2$$

Coefficient of Determination (R^2): $$1 - \frac{SSE}{TSS}$$

IBM

Linear Regression: The Syntax

Import the class containing the regression method

```
from sklearn.linear_model import LinearRegression
```

Create an instance of the class

```
LR = LinearRegression()
```

Fit the instance on the data and then predict the expected value

```
LR = LR.fit(X_train, y_train)
y_predict = LR.predict(X_test)
```

IBM