

Objective: To perform simple linear regression with one-dimensional input and output

Approach:

Steps in building regression model

Create data using scikit-learn (Number of samples = 1000, Add noise). Do not download data from internet

1. Pre-process it.
2. Creating train and test datasets
3. Visualization and descriptive analytics of patterns present in the data
4. Model building (simple linear regression)
5. Validation and evaluation of model.

Might be helpful:

The diagram shows the equation $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ with labels and brackets identifying its components:

- Dependent Variable**: Points to Y_i
- Population Y intercept**: Points to β_0
- Population Slope Coefficient**: Points to β_1
- Independent Variable**: Points to X_i
- Random Error term**: Points to ϵ_i
- Linear component**: A bracket under $\beta_0 + \beta_1 X_i$
- Random Error component**: A bracket under ϵ_i

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x};$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$