A numerical example of linear regression using a small dataset. We'll use a dataset with one independent variable X and one dependent variable Y.

Step 1: Create the Dataset

Consider the following dataset:

X	Y
1	2
2	3
4	5
3	4
5	6

Step 2: Calculate the Mean of X and Y

First, calculate the mean of the independent variable X and the dependent variable Y.

$$\bar{X} = \frac{1+2+4+3+5}{5} = 3$$

$$\bar{Y} = \frac{2+3+5+4+6}{5} = 4$$

Step 3: Calculate the Slope (m)

The slope m of the best-fit line can be calculated using the formula:

$$m = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

Let's calculate each term:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = (1-3)(2-4) + (2-3)(3-4) + (4-3)(5-4) + (3-3)(4-$$

$$= (-2)(-2) + (-1)(-1) + (1)(1) + (0)(0) + (2)(2)$$

$$= 4 + 1 + 1 + 0 + 4 = 10$$

$$\sum(X_i - \bar{X})^2 = (1-3)^2 + (2-3)^2 + (4-3)^2 + (3-3)^2 + (5-3)^2$$

$$= (-2)^2 + (-1)^2 + (1)^2 + (0)^2 + (2)^2$$

$$= 4 + 1 + 1 + 0 + 4 = 10$$

So, the slope m is:

$$m = \frac{10}{10} = 1$$

Step 4: Calculate the Intercept (b)

The intercept b can be calculated using the formula:

$$b = \bar{Y} - m\bar{X}$$

$$b = 4 - 1 \times 3 = 1$$

Step 5: Form the Regression Equation

Now that we have the slope and intercept, the linear regression equation can be written as:

$$Y = mX + b$$

$$Y = 1 \times X + 1$$

So, the final linear regression equation is:

$$Y = X + 1$$

Step 6: Use the Model for Prediction

Using the model $Y = X + 1$, you can predict the value of Y for any given X. For example, if $X = 6$:

$$Y = 6 + 1 = 7$$

This is a simple example of how to perform linear regression on a small dataset.

Let's go through a simple numerical example of multiple linear regression. We'll use a dataset with two independent variables X_1 and X_2, and one dependent variable Y.

Step 1: Create the Dataset

Consider the following dataset:

X_1	X_2	Y
1	2	3
2	1	3
4	3	7
3	5	8
5	4	9

Step 2: Set Up the Multiple Linear Regression Model

The general form of the multiple linear regression model is:

$$Y = b_0 + b_1X_1 + b_2X_2$$

where:

- Y is the dependent variable.
- X_1 and X_2 are the independent variables.
- b_0 is the intercept.
- b_1 and b_2 are the coefficients of X_1 and X_2, respectively.

Step 3: Organize the Data into Matrices

We can organize the data into matrix form, which will make it easier to calculate the coefficients:

$$X = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 4 & 3 \\ 1 & 3 & 5 \\ 1 & 5 & 4 \end{bmatrix}, \quad Y = \begin{bmatrix} 3 \\ 3 \\ 7 \\ 8 \\ 9 \end{bmatrix}$$

Here, the first column of matrix X is a column of ones to account for the intercept b_0.

Step 4: Calculate the Coefficients

The coefficients b_0, b_1, and b_2 can be calculated using the normal equation:

$$\beta = (X^T X)^{-1} X^T Y$$

where $\beta = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$.

Let's break this down step by step.

1. **Calculate $X^T X$:**

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 & 5 \\ 2 & 1 & 3 & 5 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 4 & 3 \\ 1 & 3 & 5 \\ 1 & 5 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 15 & 15 \\ 15 & 55 & 44 \\ 15 & 44 & 55 \end{bmatrix}$$

2. **Calculate $X^T Y$:**

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 & 5 \\ 2 & 1 & 3 & 5 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 30 \\ 99 \\ 97 \end{bmatrix}$$

3. **Calculate $(X^TX)^{-1}$:**
   - The inverse of X^TX can be calculated (though it's complex and often done with software tools). Let's assume it's calculated to be:

$$(X^TX)^{-1} = \begin{bmatrix} 6.1 & -2.5 & -2.5 \\ -2.5 & 0.5 & 0.5 \\ -2.5 & 0.5 & 0.5 \end{bmatrix}$$

4. **Calculate the coefficients β:**

$$\beta = (X^TX)^{-1}X^TY = \begin{bmatrix} 6.1 & -2.5 & -2.5 \\ -2.5 & 0.5 & 0.5 \\ -2.5 & 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 30 \\ 99 \\ 97 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

So, the coefficients are $b_0 = 1$, $b_1 = 1$, and $b_2 = 1$.

Step 5: Form the Regression Equation

Using the coefficients, the multiple linear regression equation is:

$$Y = 1 + 1 \times X_1 + 1 \times X_2$$

Simplifying:

$$Y = 1 + X_1 + X_2$$

Step 6: Use the Model for Prediction

Using the model $Y = 1 + X_1 + X_2$, you can predict the value of Y for any given values of X_1 and X_2. For example, if $X_1 = 2$ and $X_2 = 3$:

$$Y = 1 + 2 + 3 = 6$$

This is a simple example of how to perform multiple linear regression on a small dataset.

Problem Statement

Suppose we want to predict the price of a house (`` `Price` ``) based on its size in square feet (`` `Size` ``) and the number of bedrooms (`` `Bedrooms` ``). We have the following dataset:

House	Size (sq. ft.)	Bedrooms	Price ($1000s)
1	1500	3	330
2	1700	3	360
3	1600	2	310
4	1800	4	400
5	1750	3	380

Goal

We want to find a linear relationship of the form:

$$\text{Price} = \beta_0 + \beta_1 \times \text{Size} + \beta_2 \times \text{Bedrooms}$$

Where:

- β_0 is the intercept.
- β_1 is the coefficient for `` `Size` ``.
- β_2 is the coefficient for `` `Bedrooms` ``.

Step 1: Set Up the Linear Equation

Using the above dataset, our equation becomes:

$$330 = \beta_0 + \beta_1 \times 1500 + \beta_2 \times 3$$
$$360 = \beta_0 + \beta_1 \times 1700 + \beta_2 \times 3$$
$$310 = \beta_0 + \beta_1 \times 1600 + \beta_2 \times 2$$
$$400 = \beta_0 + \beta_1 \times 1800 + \beta_2 \times 4$$
$$380 = \beta_0 + \beta_1 \times 1750 + \beta_2 \times 3$$

Step 2: Solve for Coefficients

This system of equations can be solved using linear algebra techniques (like matrix operations) to find the best-fitting values of β_0, β_1, and β_2. However, in practical scenarios, you would typically use statistical software or programming languages like Python or R to perform these calculations.

For simplicity, let's assume that after performing the calculations, we find:

$$\beta_0 = 80, \quad \beta_1 = 0.1, \quad \beta_2 = 10$$

Step 3: Predict Using the Model

Using the estimated coefficients, the model for predicting the price is:

$$\text{Price} = 80 + 0.1 \times \text{Size} + 10 \times \text{Bedrooms}$$

For example, if a house has a size of 1600 square feet and 3 bedrooms, the predicted price would be:

\text{Price} = 80 + 0.1 \times 1600 + 10 \times 3 = 80 + 160 + 30 = 270 \text{ (\$1000s) or } 270,000 \text{ dollars}

Problem:

A company wants to predict the price of a house based on its size (in square feet) and the number of bedrooms. The company collects data on 5 houses:

House	Size (sq ft)	Number of Bedrooms	Price ($)
1	1500	3	300,000
2	1700	4	350,000
3	1400	2	280,000
4	1300	3	250,000
5	1600	4	340,000

Objective:

We want to develop a multiple linear regression model to predict the price of a house based on its size and the number of bedrooms.

Step 1: Formulate the Regression Equation

The multiple linear regression equation is:

$$\text{Price} = \beta_0 + \beta_1 \times \text{Size} + \beta_2 \times \text{Bedrooms}$$

Where:

- β_0 is the intercept (the base price when both predictors are zero).
- β_1 is the coefficient for the size of the house.
- β_2 is the coefficient for the number of bedrooms.

Step 2: Organize the Data

Let's organize the data into matrices. For regression, we define matrix X as:

$$X = \begin{bmatrix} 1 & 1500 & 3 \\ 1 & 1700 & 4 \\ 1 & 1400 & 2 \\ 1 & 1300 & 3 \\ 1 & 1600 & 4 \end{bmatrix}$$

And the output vector Y as:

$$Y = \begin{bmatrix} 300000 \\ 350000 \\ 280000 \\ 250000 \\ 340000 \end{bmatrix}$$

Step 3: Calculate the Coefficients

The coefficients β_0, β_1, and β_2 can be calculated using the formula:

$$\beta = (X^T X)^{-1} X^T Y$$

Let's compute this step by step:

1. **Compute $X^T X$:**
   - Multiply the transpose of matrix X by matrix X.
2. **Compute $(X^T X)^{-1}$:**
   - Invert the matrix obtained from the previous step.
3. **Compute $X^T Y$:**
   - Multiply the transpose of matrix X by vector Y.
4. **Calculate β:**
   - Multiply the inverted matrix $(X^T X)^{-1}$ by $X^T Y$.

I'll calculate these coefficients next.

Step 3.1: Compute $X^T X$

$$X^T X = \begin{bmatrix} 5 & 7500 & 16 \\ 7500 & 11350000 & 24400 \\ 16 & 24400 & 54 \end{bmatrix}$$

Step 3.2: Compute $(X^T X)^{-1}$

$$(X^T X)^{-1} = \begin{bmatrix} 29.23333333 & -0.02433333 & 2.33333333 \\ -0.02433333 & 0.00002333 & -0.00333333 \\ 2.33333333 & -0.00333333 & 0.83333333 \end{bmatrix}$$

Step 3.3: Compute $X^T Y$

$$X^T Y = \begin{bmatrix} 1520000 \\ 2306000000 \\ 4970000 \end{bmatrix}$$

Step 3.4: Calculate the Coefficients β

$$\beta = (X^T X)^{-1} X^T Y = \begin{bmatrix} -81333.33 \\ 253.33 \\ 1666.67 \end{bmatrix}$$

So, the coefficients are:

- $\beta_0 = -81,333.33$ (Intercept)
- $\beta_1 = 253.33$ (Size in square feet)
- $\beta_2 = 1666.67$ (Number of bedrooms)

Step 4: Interpret the Coefficients

- $\beta_0 = -81,333.33$: This is the base price when both the size and the number of bedrooms are zero, which is not practically relevant but is necessary for the equation.
- $\beta_1 = 253.33$: For each additional square foot of size, the price of the house increases by $253.33, assuming the number of bedrooms remains constant.
- $\beta_2 = 1666.67$: For each additional bedroom, the price of the house increases by $1,666.67, assuming the size remains constant.

Step 5: Make Predictions

Let's predict the price of a house that is 1550 square feet with 3 bedrooms:

$$\text{Price} = -81333.33 + 253.33 \times 1550 + 1666.67 \times 3$$