4 Simple Linear Regression: From a Line to a Model

4.1 Seeing Relationships Through a Line

4.1.1 What Linear Regression Really Means

Linear regression is a fundamental statistical and machine learning technique used to model the relationship between variables. The term **regression** refers to the process of building a mathematical model to describe how a **dependent variable** changes as one or more **independent variables** change. The word **linear** means that the relationship between the variables can be expressed as a weighted sum—essentially a combination of addition and multiplication—of the independent variables.

The key assumption of linear regression is that the effect of each independent variable on the dependent variable is **additive and linear**, rather than exponential, multiplicative, or following a more complicated nonlinear pattern. This assumption makes the model straightforward to interpret and easy to compute.

In **univariate linear regression**, we model the relationship between a single independent variable and a dependent variable using a straight line. For example, we might try to predict a person's weight based on their height, where height is the independent variable. If the model includes two or more independent variables, it is called **multiple linear regression**.

Imagine a set of scattered data points as shown in Figure 1. These points suggest a roughly linear pattern. The goal of linear regression is to find the straight line that best fits the data, minimizing the distance between the line and all points. This line can then be used for prediction, explanation, or understanding the relationship between variables.

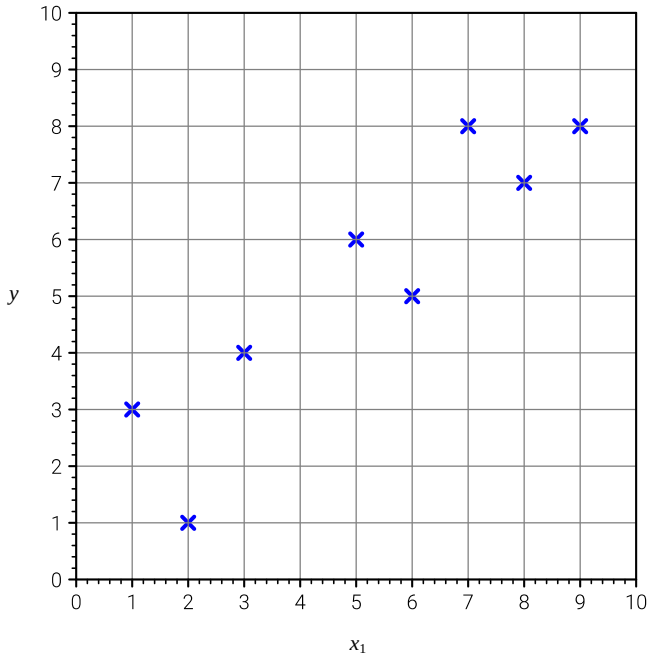

Figure 1. Scatter plot showing a linear relationship between x_1 and y

4.1.2 Correlation Is Not Causation

It is important to note that **linear regression does not imply causation**. Even if a linear association exists between variables, we cannot conclude that changes in the independent variable directly cause changes in the dependent variable. Other factors may be involved, or the observed correlation could be coincidental.

In Figure 1, the variable x_1 is commonly called the independent variable, explanatory variable, regressor, predictor, or exogenous variable, while y is called the dependent variable, response variable, regressand, or endogenous variable.

4.2 Unfolding the Equation of a Line

4.2.1 Intercept and Slope: Shifting and Tilting the Line

In univariate linear regression, the relationship between an independent variable x_1 and a dependent variable y is represented by a straight line.

On a two-dimensional plane, a line can be expressed using the slope-intercept form:

$$\hat{y} = b_0 + b_1 x_1 \tag{1}$$

Here, b_0 is called the intercept, and b_1 is the slope of the line. We write the intercept first to match the order of parameters in the column vector representation used later in this chapter.

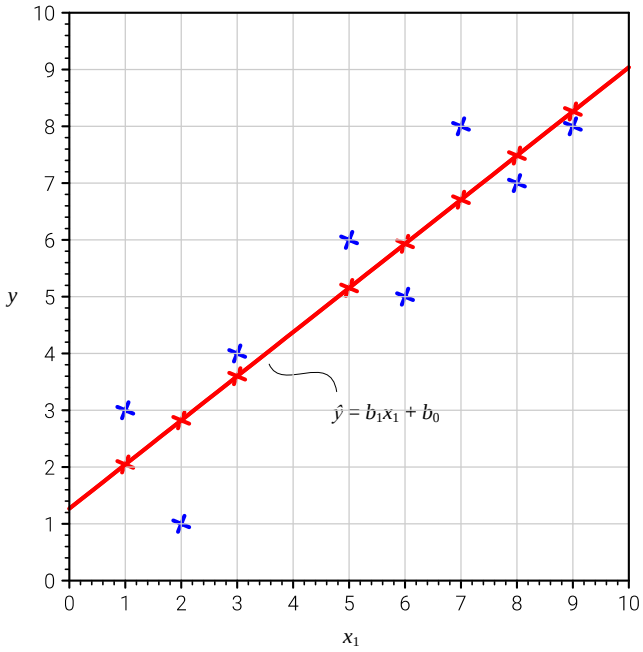

Figure 2. Slope-intercept form of a line

The intercept b_0 determines where the line crosses the vertical axis, as illustrated in Figure 3 (a). When the slope b_1 remains constant, changing b_0 shifts the entire line up or down without affecting its steepness. If b_0 is positive, the line crosses the vertical axis above the origin; if negative, it crosses below.

The slope b_1 controls the tilt of the line, as illustrated in Figure 3 (b). Keeping the intercept fixed, increasing the absolute value of b_1 makes the line steeper. A slope of zero produces a horizontal line, a positive slope tilts the line from the lower left to the upper right, and a negative slope tilts it from the upper left to the lower right.

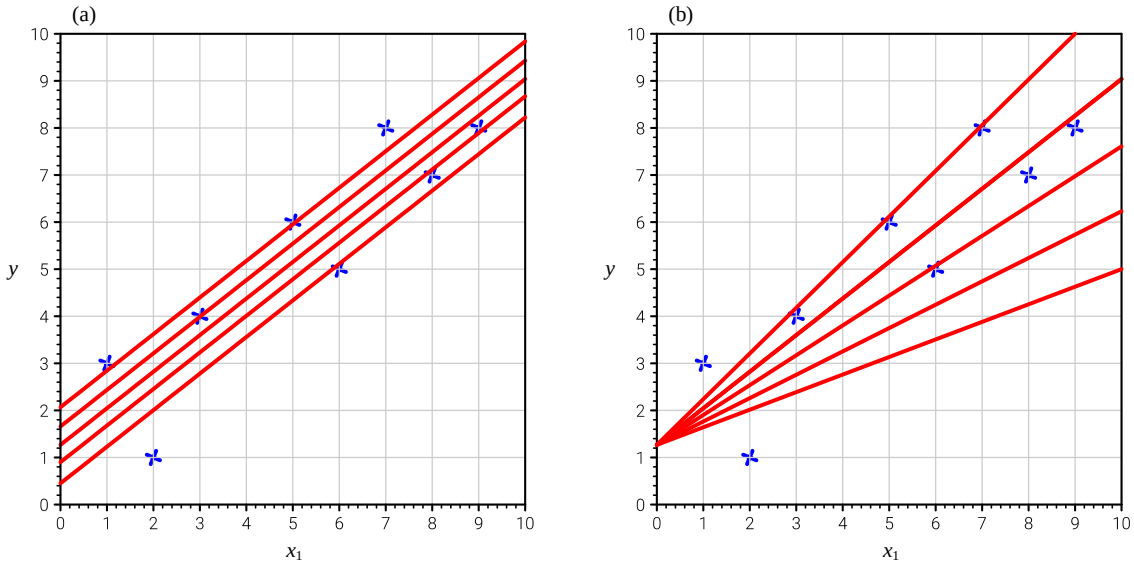

Figure 3. Effects of intercept and slope on a line

4.2.2 Predicted Values and Residuals: The Gap Between Theory and Data

In the equation above, the dependent variable is written as $\hat{y}$, called the predicted value, while y represents the actual observed value from the data. The difference between the observed y and the predicted $\hat{y}$ is called the residual, or error term:

$$\varepsilon = y - \hat{y} = y - \left(\underbrace{b_0 + b_1 x_1}_{\hat{y}} \right) \tag{2}$$

Residuals capture the vertical distance between the observed points and the fitted line, as illustrated in Figure 4.

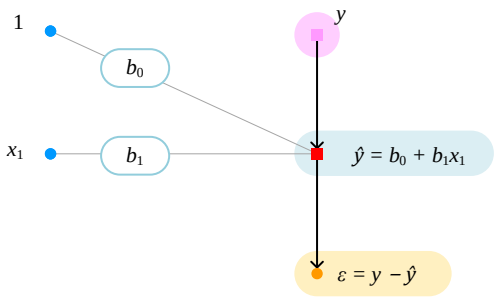

Figure 4. Relationship between x1, y, and error.

For a particular data point indexed by *i*, the horizontal coordinate is $x_1^{(i)}$, the vertical coordinate is $y^{(i)}$, and the corresponding point on the regression line is $\hat{y}^{(i)}$. The difference between $y^{(i)}$ and $\hat{y}^{(i)}$ is the residual for that point, highlighted by the vertical orange line in the Figure 5.

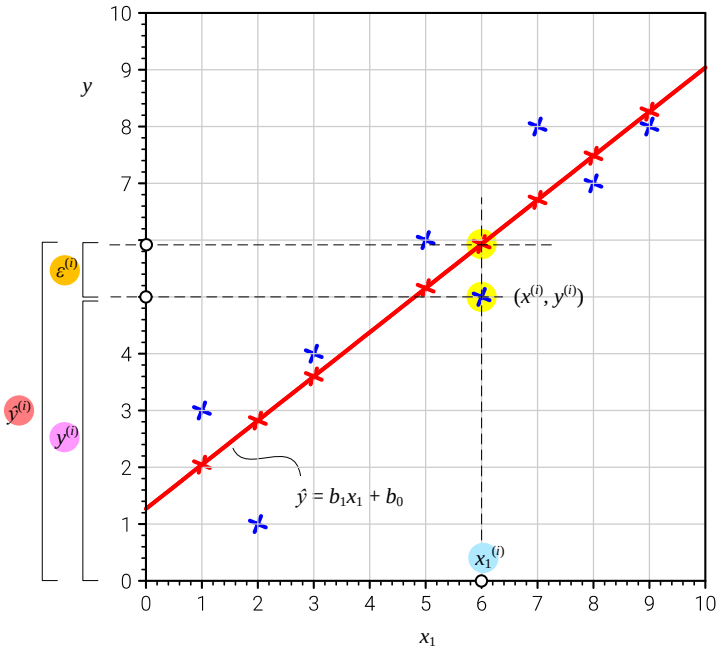

Figure 5. Observed values, predicted values, and residuals. Figure generated by Ch04_01_Simple_Linear_Regression.ipynb.

In simple terms, linear regression tries to find a red line that minimizes all residuals for the data points, making the line fit the scatter points as closely as possible. How exactly to **measure and minimize these residuals** will be discussed in the following sections.

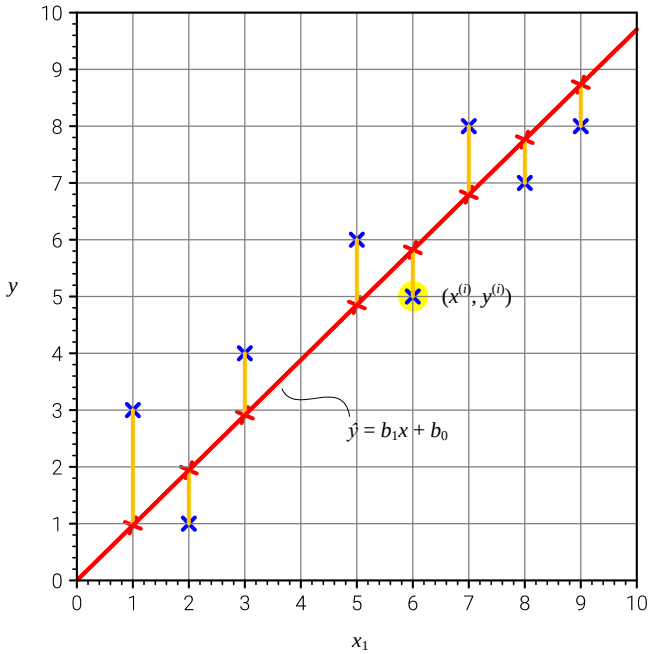

Figure 6. Residuals for all points in univariate linear regression

4.3 Vectors Behind the Line

4.3.1 Turning Data Into Vectors

To simplify computations and connect linear regression with linear algebra, we can express the data and model using vectors. The horizontal coordinates of the scatter points, corresponding to the independent variable x_1, can be written as a column vector $\boldsymbol{x}_1$. Please note that in this book, vectors are represented in lowercase, bold, italic letters, and matrices are represented in uppercase, bold, italic letters.

The vertical coordinates, representing the dependent variable y, form another column vector $\boldsymbol{y}$. The predicted values from the regression line are collected into the prediction vector $\hat{\boldsymbol{y}}$, and the residuals are written as a column vector $\boldsymbol{\varepsilon}$.

The four vector are written as:

$$\boldsymbol{x}_1 = \begin{bmatrix} x_1^{(1)} \\ x_1^{(2)} \\ \vdots \\ x_1^{(i)} \\ \vdots \\ x_1^{(n)} \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(i)} \\ \vdots \\ y^{(n)} \end{bmatrix}, \quad \hat{\boldsymbol{y}} = \begin{bmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(i)} \\ \vdots \\ \hat{y}^{(n)} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \\ \vdots \\ \varepsilon^{(i)} \\ \vdots \\ \varepsilon^{(n)} \end{bmatrix} \tag{3}$$

In Figure 6, the column vectors correspond to the plot as follows: the elements of $\boldsymbol{x}_1$ represent the horizontal coordinates of the red and blue points, the elements of $\boldsymbol{y}$ represent the vertical coordinates of the blue points, the elements of $\hat{\boldsymbol{y}}$ represent the vertical coordinates of the red points on the line, and the elements of $\boldsymbol{\varepsilon}$ correspond to the vertical orange segments connecting the points to the line.

4.3.2 Vector Equation of Linear Regression

Using vector notation allows us to express the red regression line as a simple vector addition:

$$\underset{\hat{\boldsymbol{y}}}{\begin{bmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(i)} \\ \vdots \\ \hat{y}^{(n)} \end{bmatrix}} = b_0 \underset{\boldsymbol{1}}{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}} + b_1 \underset{\boldsymbol{x}_1}{\begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(i)} \\ \vdots \\ x^{(n)} \end{bmatrix}} \tag{4}$$

Thus

$$\hat{\boldsymbol{y}} = b_0 \boldsymbol{1} + b_1 \boldsymbol{x}_1 \tag{5}$$

Here, $\boldsymbol{1}$ is a column vector of ones with the same length as $\boldsymbol{x}_1$.

This compact form makes it easier to handle multiple data points at once and lays the foundation for matrix-based calculations in multiple regression.

The key question now is how to choose the parameters b_0 and b_1 so that the line fits the data as well as possible. This problem is closely related to the concept of orthogonal projection, which will be explored later in this chapter.

4.4 The Geometry of Regression: Orthogonal Projection

4.4.1 Projecting Data onto a Plane

The key to determining the regression coefficients b_0 and b_1 lies hidden in the vector equation above.

(5) is a familiar structure: it is a linear combination. In other words, the predicted values $\hat{\boldsymbol{y}}$ lie within the plane spanned by the vectors $\boldsymbol{1}$ and $\boldsymbol{x}_1$ denoted as span($\boldsymbol{1}$, $\boldsymbol{x}_1$). It is important to note that $\boldsymbol{1}$ and $\boldsymbol{x}_1$ are assumed to be linearly independent, so they indeed define a plane.

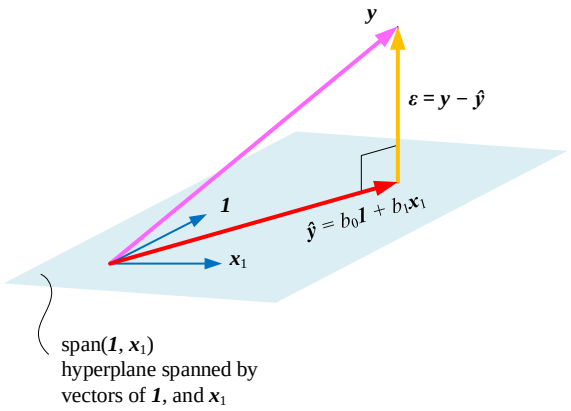

Figure 7. Vector $\boldsymbol{y}$ projected onto the plane spanned by $\boldsymbol{1}$ and $\boldsymbol{x}_1$ in univariate linear regression

4.4.2 Orthogonal Decomposition: Predicted vs. Residual Components

As illustrated in Figure 7, the predicted vector $\hat{\boldsymbol{y}}$ can be expressed as a linear combination of $\boldsymbol{1}$ and $\boldsymbol{x}_1$. The observed data vector $\boldsymbol{y}$, however, generally does not lie exactly in this plane.

To find the best-fitting line, we project $\boldsymbol{y}$ orthogonally onto the plane span($\boldsymbol{1}$, $\boldsymbol{x}_1$). The resulting vector in the plane is $\boldsymbol{y}$, while the component of $\boldsymbol{y}$ perpendicular to the plane is the residual vector $\boldsymbol{\varepsilon}$.

This gives a natural orthogonal decomposition of $\boldsymbol{y}$ into two parts: the predicted vector $\hat{\boldsymbol{y}}$ and the residual vector $\boldsymbol{\varepsilon}$, so that

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{\varepsilon} = b_0\boldsymbol{1} + b_1\boldsymbol{x} + \boldsymbol{\varepsilon} \tag{6}$$

Rewriting this, the residual vector is

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \left(b_0\boldsymbol{1} + b_1\boldsymbol{x}\right) \tag{7}$$

A crucial property is that the residual vector $\boldsymbol{\varepsilon}$ is perpendicular to the plane spanned by $\boldsymbol{1}$ and $\boldsymbol{x}_1$. Thus, $\boldsymbol{\varepsilon}$ is perpendicular to both $\boldsymbol{1}$ and $\boldsymbol{x}_1$.

This orthogonality is the foundation of the least-squares solution and will be used extensively in the derivation of the regression coefficients.

4.5 Matrix Form of Linear Regression

4.5.1 Constructing the Design Matrix

We can further simplify univariate linear regression using matrix notation. Recall that the predicted values $\hat{\boldsymbol{y}}$ can be written as a linear combination of the intercept and the independent variable. This can be compactly expressed in matrix multiplication form:

$$\hat{\boldsymbol{y}} = \underbrace{\begin{bmatrix}\boldsymbol{1} & \boldsymbol{x}\end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix}b_0 \\ b_1\end{bmatrix}}_{\boldsymbol{b}} = \boldsymbol{X}\boldsymbol{b} \tag{8}$$

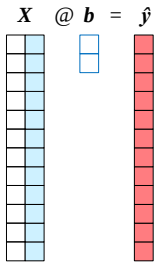

Figure 8. Predicted values $\hat{\boldsymbol{y}}$ computed using matrix multiplication in univariate linear regression

In this expression, $\boldsymbol{X}$ is called the design matrix, and $\boldsymbol{b}$ is the vector of coefficients, here $[b_0, b_1]^T$. The design matrix organizes the data in a convenient form for linear algebra operations. Typically, the first column of $\boldsymbol{X}$ is a column of ones to represent the intercept term, while the remaining columns contain the values of the independent variables. For univariate linear regression with a nonzero intercept, the design matrix $\boldsymbol{X}$ is constructed by placing a column of ones on the left and the column vector $\boldsymbol{x}_1$ on the right.

For the data in Figure 1, the corresponding design matrix is:

$$\boldsymbol{X} = \begin{bmatrix}\boldsymbol{1} & \boldsymbol{x}_1\end{bmatrix} \tag{9}$$

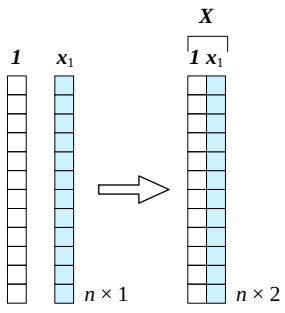

Figure 9. Construction of the design matrix $\boldsymbol{X}$ for univariate linear regression

4.5.2 Writing Regression as Matrix Multiplication

Using this matrix form, the observed values $\boldsymbol{y}$ can be expressed as the sum of the predicted values $\hat{\boldsymbol{y}}$ and the residuals $\boldsymbol{\varepsilon}$:

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{\varepsilon} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{\varepsilon} \tag{10}$$

Here, y y corresponds to the vertical coordinates of the actual data points (blue × in Figure 6), $\hat{\boldsymbol{y}}$ corresponds to the predicted values on the regression line (red ×), and ε represents the vertical residuals (orange line segments).

This matrix formulation not only simplifies calculations but also lays the groundwork for extending linear regression to multiple independent variables, where matrix operations become essential.

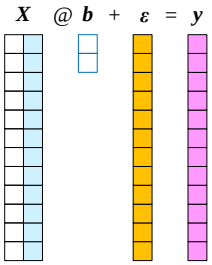

Figure 10. Observed values $\boldsymbol{y}$ as the sum of predicted values $\hat{\boldsymbol{y}}$ and residuals $\boldsymbol{\varepsilon}$ using matrix operations.

It is easy to derive the vector of residuals as

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b} \tag{11}$$

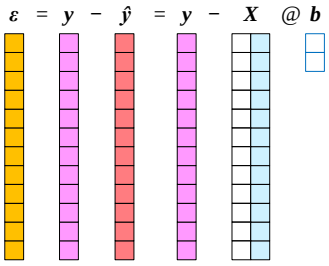

Figure 11. Residual vector $\boldsymbol{\varepsilon}$ computed using matrix operations in univariate linear regression

4.6 Ordinary Least Squares (OLS): Minimizing the Mistakes

4.6.1 From Residuals to the Least-Squares Criterion

An important property of linear regression is that the residual vector $\boldsymbol{\varepsilon}$ is orthogonal to the columns of the design matrix $\boldsymbol{X}$. In other words, the residuals are perpendicular to each vector representing the intercept and the independent variable(s). This orthogonality means that $\boldsymbol{\varepsilon}$ satisfies

$$\begin{cases} \boldsymbol{1}^{\mathrm{T}}\boldsymbol{\varepsilon} = 0 \\ \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\varepsilon} = 0 \end{cases} \tag{12}$$

or equivalently, in matrix form:

$$\begin{bmatrix} \boldsymbol{1}^{\mathrm{T}} \\ \boldsymbol{x}_1^{\mathrm{T}} \end{bmatrix} \boldsymbol{\varepsilon} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{13}$$

Thus

$$\boldsymbol{X}^{\mathrm{T}} \boldsymbol{\varepsilon} = \boldsymbol{0} \tag{14}$$

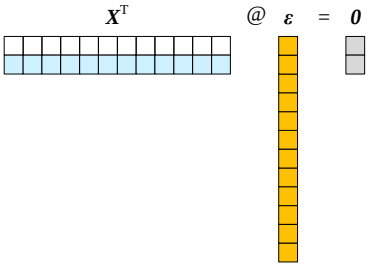

Figure 12. Residual vector ε ε is perpendicular to each column of X X in univariate linear regression

This orthogonality property is crucial because it ensures that the predicted vector $\hat{\boldsymbol{y}}$ is the closest point in the column space of $\boldsymbol{X}$ to the observed vector $\boldsymbol{y}$. In other words, the regression line represents the orthogonal projection of the observed data onto the plane spanned by the intercept and the independent variable. This concept was introduced in the previous chapter and forms the mathematical foundation for solving for the regression coefficients $\boldsymbol{b}$.

4.6.2 The Pseudo-Inverse: A Shortcut to the Solution

Substituting $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{Xb}$ into (14) gives

$$\boldsymbol{X}^{\mathrm{T}} \left(\boldsymbol{y} - \boldsymbol{Xb} \right) = \boldsymbol{0} \tag{15}$$

Rearranging this equation, we obtain:

$$\boldsymbol{X}^{\mathrm{T}} \boldsymbol{Xb} = \boldsymbol{X}^{\mathrm{T}} \boldsymbol{y} \tag{16}$$

Earlier, we assumed that the vectors $\boldsymbol{1}$ and $\boldsymbol{x}_1$ are linearly independent, so the design matrix $\boldsymbol{X} = [\boldsymbol{1}, \boldsymbol{x}_1]$ is full rank. As a result, the Gram matrix $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is invertible, allowing us to derive an analytical solution for the coefficient vector:

$$\boldsymbol{b} = \left(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{y} \tag{17}$$

In the vector space span($\boldsymbol{1}$, $\boldsymbol{x}_1$), the coefficient vector $\boldsymbol{b}$ represents the coordinates of the orthogonal projection of $\boldsymbol{y}$ onto this plane.

In other words, $\boldsymbol{b}$ can be viewed as the scalar projections of $\boldsymbol{y}$ along the vectors $\boldsymbol{1}$ and $\boldsymbol{x}_1$.

As shown in Figure 13, the matrix $\left(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\mathrm{T}}$ used in this calculation is called the pseudo-inverse or generalized inverse of $\boldsymbol{X}$, which provides a convenient way to solve for $\boldsymbol{b}$.

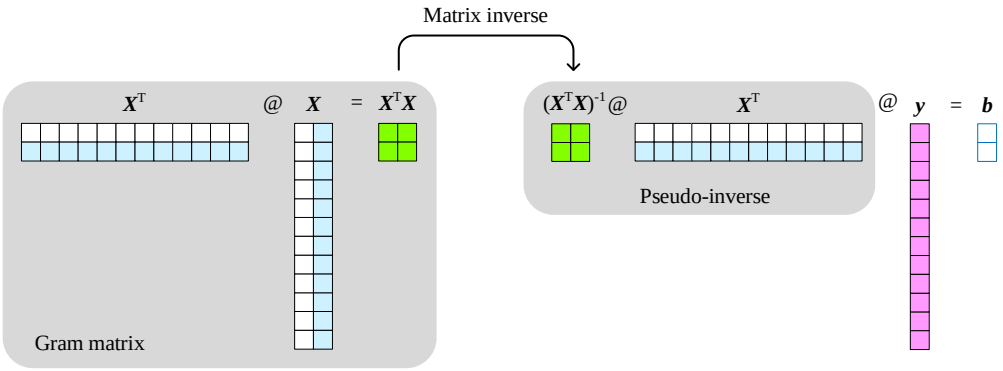

Figure 13. Computing the coefficient vector *b* b using the pseudo-inverse in univariate linear regression

Using actual data values from Figure 2, the regression coefficients for the red line are calculated as:

$$\boldsymbol{b} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} = \left(\begin{bmatrix}1&1\\1&2\\1&3\\1&5\\1&6\\1&7\\1&8\\1&9\end{bmatrix}^{\mathrm{T}} @ \begin{bmatrix}1&1\\1&2\\1&3\\1&5\\1&6\\1&7\\1&8\\1&9\end{bmatrix}\right)^{-1} @ \begin{bmatrix}1&1\\1&2\\1&3\\1&5\\1&6\\1&7\\1&8\\1&9\end{bmatrix}^{\mathrm{T}} @ \begin{bmatrix}3\\1\\4\\6\\5\\8\\7\\8\end{bmatrix} \tag{18}$$

$$= \begin{bmatrix}8&41\\41&269\end{bmatrix}^{-1} @ \begin{bmatrix}42\\261\end{bmatrix} = \begin{bmatrix}0.571&-0.087\\-0.087&0.016\end{bmatrix} @ \begin{bmatrix}42\\261\end{bmatrix} = \begin{bmatrix}1.267\\0.777\end{bmatrix}$$

Thus, $b_0 = 1.267$, and $b_1 = 0.777$.

Thus, the equation of the red regression line in Figure 2 can be written explicitly as

$$\hat{y} = b_0 + b_1 x_1 = 1.267 + 0.777 x_1 \tag{19}$$

This gives us the complete analytical expression for the predicted values along the regression line.

4.6.3 How Projection Produces the Best Linear Fit

Once we have the coefficient vector $\boldsymbol{b}$, we can compute the predicted values $\hat{\boldsymbol{y}}$ using matrix multiplication:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\underbrace{\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}}_{\boldsymbol{b}} \tag{20}$$

The matrix $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}$ in (20) is called the projection matrix. Its role is to project the observed data vector *y* y onto the column space of $\boldsymbol{X}$, which is the plane spanned by the intercept and the independent variable $\boldsymbol{x}_1$. This ensures that the predicted vector $\hat{\boldsymbol{y}}$ lies exactly in that plane and represents the best linear approximation to the observed data in the least-squares sense.

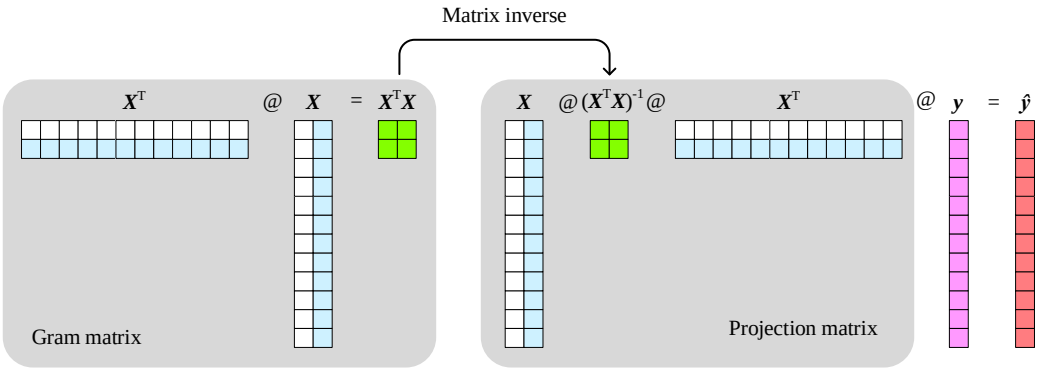

Figure 14. Computing the predicted values $\hat{\boldsymbol{y}}$ using the projection matrix in univariate linear regression

Using the matrix $\boldsymbol{P}$, the predicted vector can be written as

$$\hat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y} \tag{21}$$

Figure 6, the red × points represent the elements of $\hat{\boldsymbol{y}}$. As expected, all these points lie exactly on the red regression line, illustrating how the projection matrix transforms the observed data into its best linear fit.

4.6.4 The Area of Errors: Sum of Squared Residuals

Earlier, we discussed that the goal of linear regression is to find the coefficients b_0 and b_1 that make the residual vector $\boldsymbol{\varepsilon}$ as small as possible. To measure the "size" of a vector, we use a vector norm, and the most common choice is the L2 norm, which corresponds to the Euclidean length of the vector. For the residual vector $\boldsymbol{\varepsilon}$, its L2 norm is given by

$$\left\|\boldsymbol{\varepsilon}\right\| = \left\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\right\| \tag{22}$$

Each element of this vector corresponds to the vertical orange line segments in Figure 6, representing the difference between observed and predicted values. To simplify calculations, we often work with the squared L2 norm, which sums the squares of the residuals:

$$\left\|\boldsymbol{\varepsilon}\right\|_2^2 = \left\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\right\|_2^2 \tag{23}$$

Thus

$$\left\|\boldsymbol{\varepsilon}\right\|_2^2 = \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} = \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\right)^{\mathrm{T}}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\right) \tag{24}$$

Expanding the above gives:

$$\left\|\boldsymbol{\varepsilon}\right\|_2^2 = \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon} = \left\langle\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}\right\rangle = \begin{bmatrix}\varepsilon^{(1)}\\ \varepsilon^{(2)}\\ \vdots\\ \varepsilon^{(i)}\\ \vdots\\ \varepsilon^{(n)}\end{bmatrix}\cdot\begin{bmatrix}\varepsilon^{(1)}\\ \varepsilon^{(2)}\\ \vdots\\ \varepsilon^{(i)}\\ \vdots\\ \varepsilon^{(n)}\end{bmatrix} = \left(\varepsilon^{(1)}\right)^2 + \left(\varepsilon^{(2)}\right)^2 + \cdots + \left(\varepsilon^{(i)}\right)^2 + \cdots + \left(\varepsilon^{(n)}\right)^2 \tag{25}$$

This is the core idea behind ordinary least squares (OLS). In simple terms, OLS finds the line that minimizes the sum of squared differences between the predicted values $\hat{\boldsymbol{y}}$ and the actual observations $\boldsymbol{y}$. Using the data from

Figure 6, the squared L2 norm of $\boldsymbol{\varepsilon}$ corresponds to the sum of the areas of the orange squares in Figure 15, where the side of each square represents a residual.

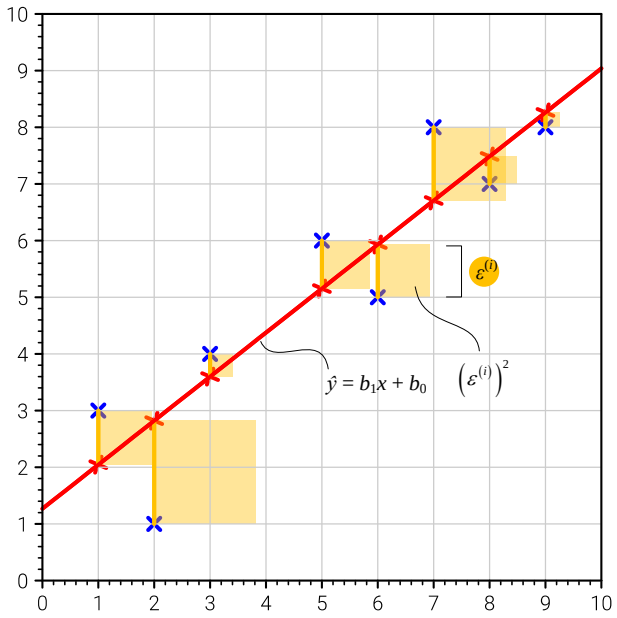

Figure 15. Sum of squared residuals represented as the total area of squares for each data point

It is important to note that obtaining the red regression line is not the end of linear regression analysis—it is just the beginning. To fully understand and trust the model, we must use statistical tools to assess its reliability. This includes evaluating whether the model is statistically significant, checking the significance of each variable, verifying assumptions about residuals such as normality and homoscedasticity, estimating the uncertainty of the parameters, constructing confidence intervals, assessing goodness of fit, and exploring potential multicollinearity among predictors. Only by carefully examining these statistical properties can we make informed judgments about the model's explanatory power, predictive ability, and its practical applicability.

4.7 Conclusions

This chapter introduces simple linear regression, a foundational technique in statistics and machine learning used to model the relationship between a dependent variable and one independent variable. The method assumes a linear and additive effect of the independent variable on the dependent variable, making the model easy to interpret and compute. In univariate linear regression, a straight line is fitted to a set of scattered data points, minimizing the differences between observed and predicted values. These differences, called residuals, can be represented as vectors, allowing the model to be expressed in matrix form. The chapter explains how the predicted values are the orthogonal projection of the observed data onto the plane defined by the intercept and independent variable, and how the regression coefficients can be computed using the pseudo-inverse of the design matrix. Ordinary least squares is used to minimize the sum of squared residuals, producing the best-fitting line, while statistical tools are necessary to evaluate its reliability, significance, and assumptions.