5 Multivariate Linear Regression

5.1 From a Line to a Plane: Extending Linear Regression to Two Variables

5.1.1 The Geometry of the Regression Plane

In the previous section, we explored simple linear regression using the concept of orthogonal projection, focusing on the relationship between one independent variable and a dependent variable. In this section, we expand our view to consider multiple variables, starting with two independent variables. It is highly recommended to study this section alongside the previous one, as the concepts build directly on what we have already learned.

Building on the idea of simple linear regression, two-variable linear regression models the relationship between a dependent variable y y and two independent variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Intuitively, we are now trying to fit a plane, rather than a line, through the data points in three-dimensional space. The model can be expressed as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon \tag{1}$$

Here, b_0 is the intercept, b_1 and b_2 are the coefficients for the independent variables, and ε represents the residual error — the part of y not explained by the model (see Figure 1).

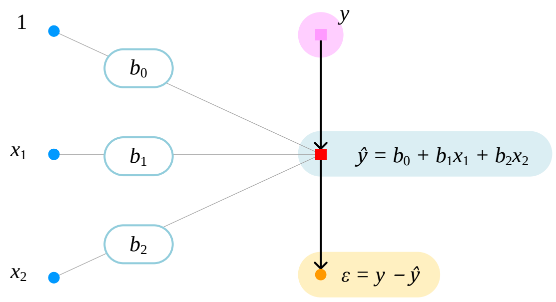

Figure 1. Relationship between dependent and two independent variables in two-variable linear regression

From a geometric perspective, two-variable linear regression produces a plane in three-dimensional space. Figure 2 illustrates how the data points relate to this regression plane, showing the goal of fitting the plane as closely as possible to the scatter of points.

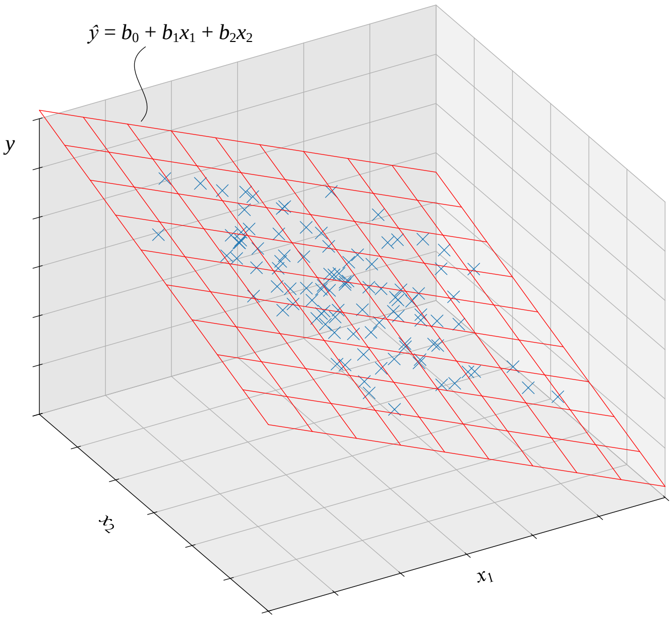

Figure 2. The regression plane in two-variable linear regression showing data points and fitted surface. Figure generated by Ch05_01_Bivariate_Linear_Regression.ipynb.

5.1.2 Building Intuition: The Regression Plane as a Sum of Simpler Surfaces

The parameters b_0, b_1, and b_2 each affect the plane in different ways.

The intercept b_0 controls the overall height of the plane. Increasing or decreasing b_0 shifts the plane up or down without changing its slope or orientation (see Figure 3).

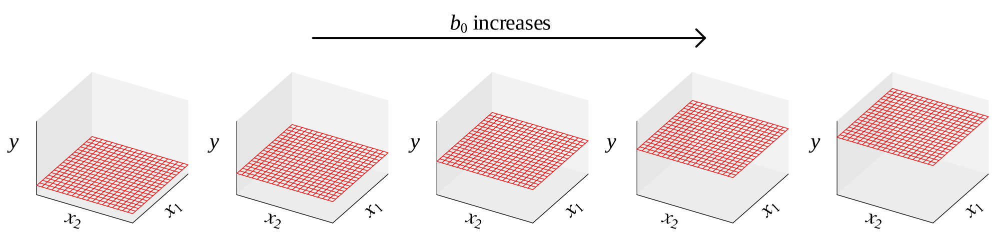

Figure 3. Intercept b_0 shifts the plane vertically without tilting

The coefficients b_1 and b_2 determine the slope of the plane along the x_1 and x_2 directions, respectively.

As illustrated in Figure 4, he magnitude of b_1 controls how steeply the plane rises or falls along x_1, while its sign decides whether the plane ascends or descends as x_1 increases.

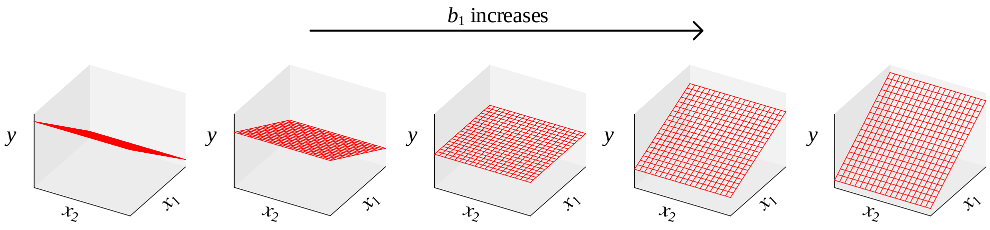

Figure 4. Coefficient b_1 controls the slope along the x_1 direction, b_0 and b_2 set to zero

Similarly, b_2 controls the slope along x_2, with its sign determining the direction of tilt (see Figure 5). Together, b_1 and b_2 define the overall tilt and orientation of the plane, while b_0 sets its vertical position in space.

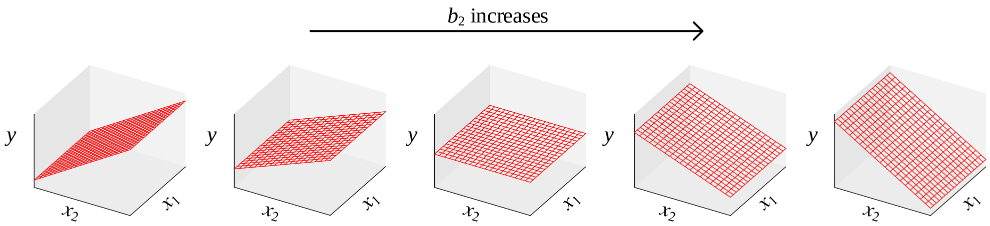

Figure 5. Coefficient b_2 controls the slope along the x_1 direction, b_0 and b_1 set to zero

Intuitively, we can also think of $\hat{y} = b_0 + b_1x_1 + b_2x_2$ as the sum of three simpler planes (see Figure 6), each contributing a different aspect of the overall shape. Using the orthogonal projection perspective introduced in the previous section, two-variable regression can be seen as finding the plane that best projects the data points onto this three-dimensional space.

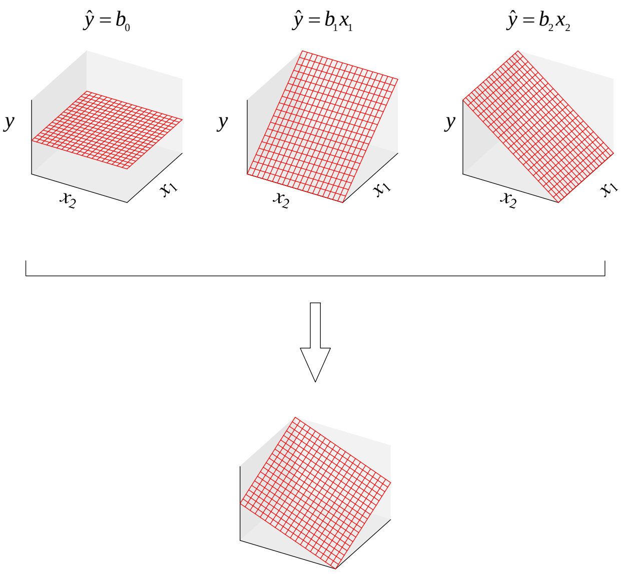

Figure 6. The regression plane as the sum of three component planes corresponding to $\hat{y} = b_0 + b_1x_1 + b_2x_2$

5.1.3 Orthogonal Projection View of Two-Variable Regression

In two-variable linear regression, the vector of observed values y y can be expressed as a combination of the intercept and the two independent variables. Geometrically, the predicted values $\hat{\boldsymbol{y}}$ represent the orthogonal projection of $\boldsymbol{y}$ onto the plane spanned by the vectors $\boldsymbol{1}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$, as illustrated in Figure 7.

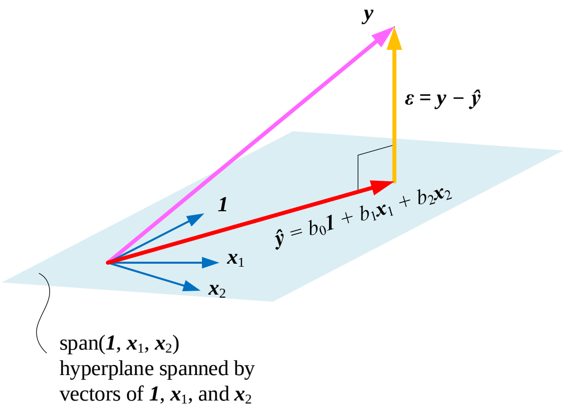

Figure 7. Orthogonal projection of $\boldsymbol{y}$ onto the plane spanned by $\boldsymbol{1}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$

Mathematically, $\hat{\boldsymbol{y}}$ is a linear combination of the vectors $\boldsymbol{1}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$:

$$\hat{\boldsymbol{y}} = b_0\boldsymbol{1} + b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 \tag{2}$$

Thus

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{\varepsilon} = b_0\boldsymbol{1} + b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \boldsymbol{\varepsilon} \tag{3}$$

Figure 8 shows the shape of the design matrix $\boldsymbol{X}$ for two-variable regression.

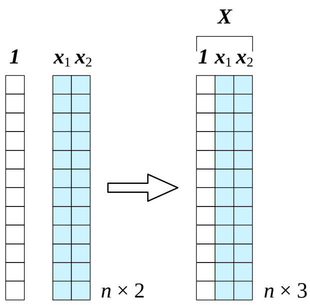

Figure 8. Design matrix $\boldsymbol{X}$ in two-variable linear regression

Since there are two independent variables, $\boldsymbol{X}$ has three column vectors corresponding to $\boldsymbol{1}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$. Here we assume these vectors are linearly independent, so the design matrix $\boldsymbol{X}$ is full column rank, the Gram matrix $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is invertible, and the regression coefficients $\boldsymbol{b}$ can be computed directly. In practice, the columns of $\boldsymbol{X}$ may be linearly dependent, which requires special handling such as regularization.

We can write the linear combination in matrix form (see Figure 9):

$$\hat{\boldsymbol{y}} = \underbrace{\begin{bmatrix} \boldsymbol{1} & \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}}_{\boldsymbol{b}} = \boldsymbol{X}\boldsymbol{b} \tag{4}$$

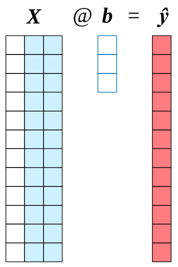

Figure 9. Matrix multiplication to obtain $\hat{\boldsymbol{y}}$

Figure 10 illustrates the matrix multiplication for computing vector $\boldsymbol{b}$.

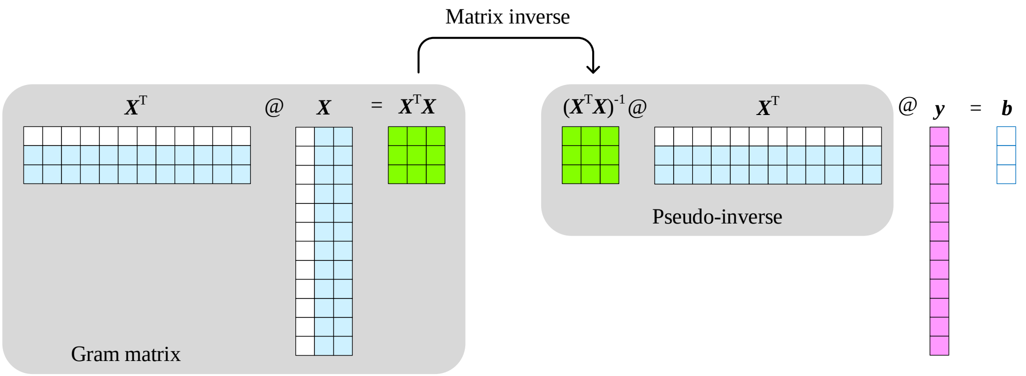

Figure 10. Calculating the coefficient vector $\boldsymbol{b}$ using the normal equation

Once $\boldsymbol{b}$ is obtained, the predicted values $\hat{\boldsymbol{y}}$ can be calculated as shown in Figure 11. This matrix-based perspective connects the geometric idea of projection with the algebraic computation of regression coefficients.

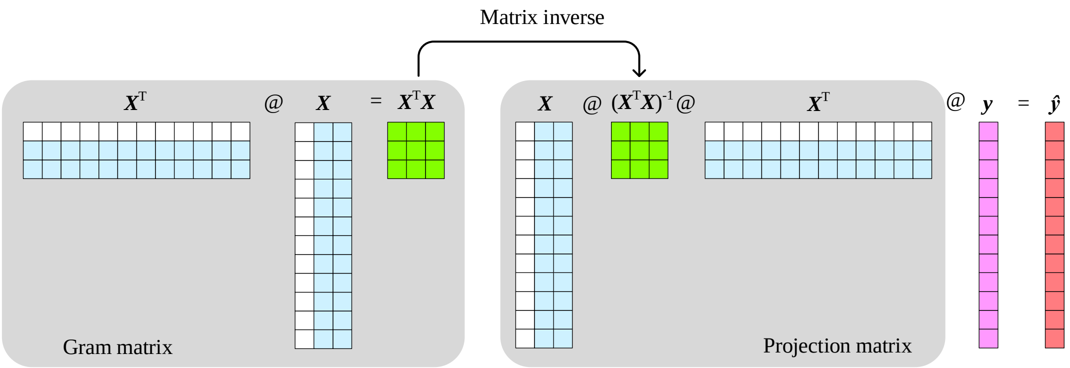

Figure 11. Computing predicted values $\hat{y}$ from the regression coefficients

5.2 Climbing Dimensions: From Planes to Hyperplanes

5.2.1 Geometric Perspective: Fitting a Hyperplane

Continuing our exploration into higher dimensions, we now consider multivariable linear regression. This technique models the linear relationship between a dependent variable $\boldsymbol{y}$ and multiple independent variables x_1, x_2, … , x_D. The general form of the model is:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_D x_D + \varepsilon \tag{5}$$

Again, b_0 is the intercept, b_1, b_2, … , b_D are the coefficients for each independent variable, ε is the residual error, and D denotes the number of independent variables.

Geometrically, multivariable linear regression produces a hyperplane in a $(D + 1)$-dimensional space. A hyperplane is a generalization of a plane to higher dimensions, representing the best linear approximation of the data in this multidimensional space (Figure 12).

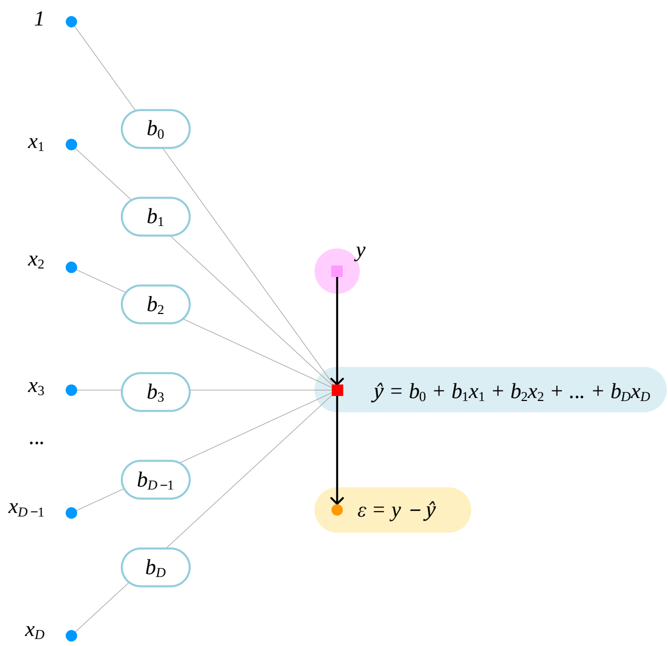

Figure 12. Relationship between the dependent variable and multiple independent variables in multivariable linear regression

5.2.2 Vector Space View: Orthogonal Projection in Higher Dimensions

From the vector space perspective, the vectors $\boldsymbol{1}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, … , $\boldsymbol{x}_D$ span a subspace span $(\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$. The predicted values $\hat{\boldsymbol{y}}$ are the orthogonal projection of $\boldsymbol{y}$ onto this subspace, which means $\hat{\boldsymbol{y}}$ is the closest point within the hyperplane to the actual data points (Figure 13).

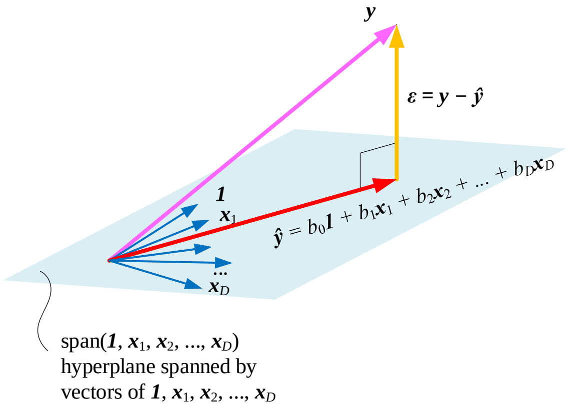

Figure 13. Orthogonal projection of $\boldsymbol{y}$ onto the hyperplane spanned by $\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D$

In other words, $\hat{\boldsymbol{y}}$ can be expressed as a linear combination of 1 , x 1 , x 2 , … , x D 1,x 1 ,x 2 ,…,x D :

$$\hat{\boldsymbol{y}} = b_1 \boldsymbol{x}_1 + b_2 \boldsymbol{x}_2 + \cdots + b_D \boldsymbol{x}_D \tag{6}$$

The residual vector $\boldsymbol{\varepsilon}$ is orthogonal to the subspace spanned by the independent variables. It also means that $\boldsymbol{\varepsilon}$ is perpendicular to each of the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D$:

$$\begin{cases} \boldsymbol{x}_1^{\mathrm{T}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0 \\ \boldsymbol{x}_2^{\mathrm{T}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0 \\ \qquad \vdots \\ \boldsymbol{x}_D^{\mathrm{T}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = 0 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_D^{\mathrm{T}} \end{bmatrix}}_{X^{\mathrm{T}}} (\boldsymbol{y} - \hat{\boldsymbol{y}}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{7}$$

In terms of vector spaces, span $(\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$ and span $(\boldsymbol{\varepsilon})$ form orthogonal complements.

The design matrix of this regression is

$$\boldsymbol{X}_{n\times(D+1)} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D \end{bmatrix} \tag{8}$$

Figure 14 shows the design matrix $\boldsymbol{X}$ for multivariable regression, where each column represents one of the vectors $\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D$.

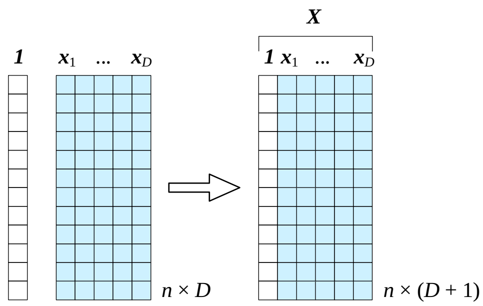

Figure 14. Design matrix X X in multivariable linear regression

5.3 The Algebra Behind Geometry: Matrix Formulation of Regression

5.3.1 The Design Matrix and the Normal Equation

In multivariable linear regression, the predicted values $\hat{\boldsymbol{y}}$ can be conveniently written as a column vector using matrix multiplication:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{b} \tag{9}$$

Here, $\boldsymbol{X}$ is the design matrix containing all independent variables and a column of ones for the intercept, and $\boldsymbol{b}$ is the coefficient vector. This matrix calculation is illustrated in Figure 15.

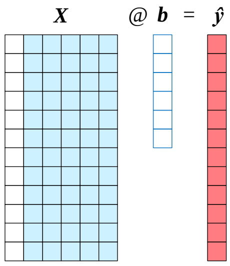

Figure 15. Computing predicted values $\hat{\boldsymbol{y}}$ using matrix multiplication in multivariable regression

If the Gram matrix $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is invertible, the coefficient vector $\boldsymbol{b}$ can be obtained directly using the normal equation:

$$\boldsymbol{b} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{10}$$

Figure 16 shows this calculation. Essentially, finding $\boldsymbol{b}$ is equivalent to solving an overdetermined system of linear equations, where the number of equations (data points) exceeds the number of unknowns (coefficients).

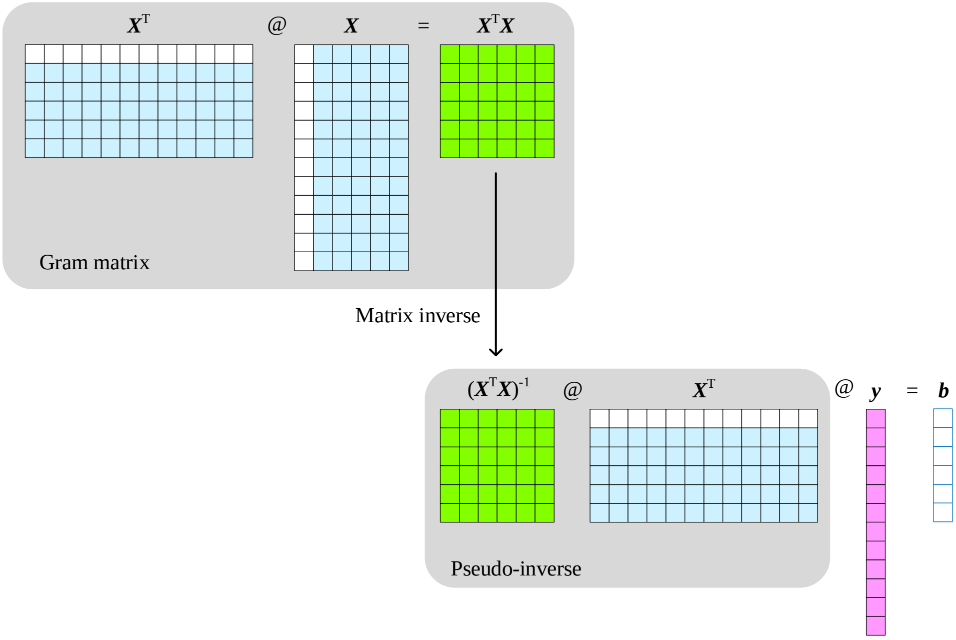

Figure 16. Solving for the coefficient vector b b using the normal equation

5.3.2 Computing Predictions with Matrices

The fitted values $\hat{\boldsymbol{y}}$ can also be expressed using a projection matrix:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{11}$$

This is illustrated in Figure 17. From a geometric perspective, this matrix projects the vector y y onto the subspace spanned by the columns of $\boldsymbol{X}$. In regression terminology, $\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}$ is called the hat matrix because it maps $\boldsymbol{y}$ to $\hat{\boldsymbol{y}}$.

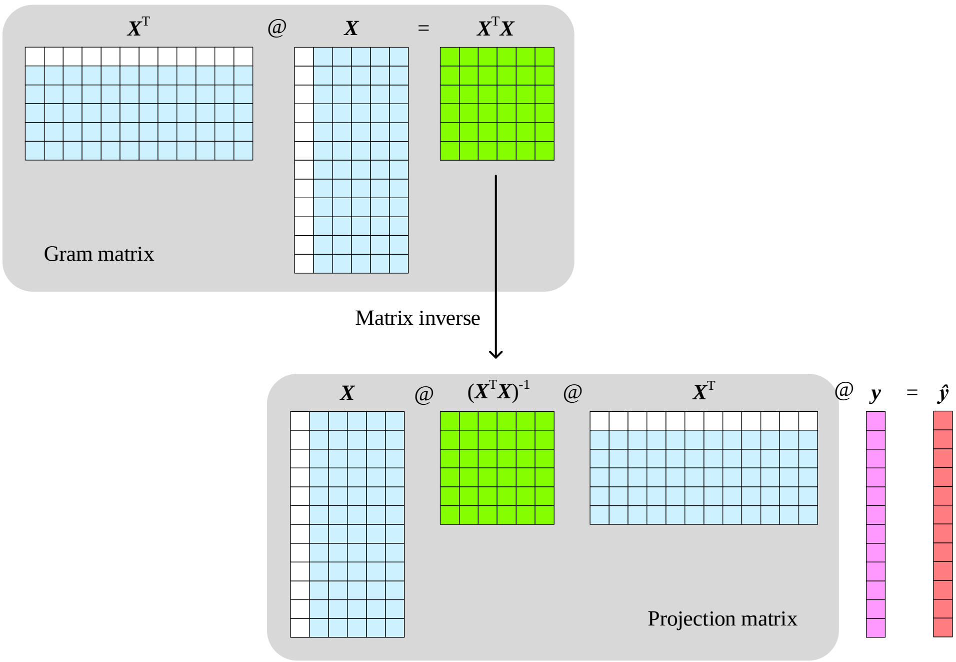

Figure 17. Computing predicted values y ^ y ^ using matrix multiplication in multivariable regression

5.3.3 The Residual Space: Orthogonal Complement of the Model

The residual vector $\boldsymbol{\varepsilon}$ can also be written in terms of a projection matrix:

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} = \left(\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\right)\boldsymbol{y} \tag{12}$$

Geometrically, this projects $\boldsymbol{y}$ onto the orthogonal complement of the column space of $\boldsymbol{X}$. In other words, the residual vector $\boldsymbol{\varepsilon}$ is perpendicular to the hyperplane defined by $\boldsymbol{X}$, making it the normal vector of that hyperplane.

Finally, we can express y y as the sum of its projection onto the hyperplane and its projection onto the residual space:

$$\hat{\boldsymbol{y}} + \boldsymbol{\varepsilon} = \underbrace{\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}}_{\text{Projection matrix}}\boldsymbol{y} + \underbrace{\left(\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\right)}_{\text{Projection matrix}}\boldsymbol{y} = \boldsymbol{I}\boldsymbol{y} = \boldsymbol{y} \tag{13}$$

This decomposition provides a clear geometric view of linear regression: $\boldsymbol{y}$ is orthogonally split into the fitted component $\hat{\boldsymbol{y}}$ lying in the hyperplane spanned by the independent variables, and the residual component $\boldsymbol{\varepsilon}$ lying in the orthogonal complement (see Figure 13).

5.4 Conclusion

This chapter extends linear regression from one variable to multiple variables, showing how the model generalizes from a line to a plane and eventually to a hyperplane in higher dimensions. It explains that in two-variable regression, the goal is to fit a plane that best represents the relationship between one dependent variable and two independent variables. Each coefficient controls how the plane tilts or shifts in space.

The chapter then introduces the concept of orthogonal projection, describing the fitted values as the projection of the observed data onto the subspace defined by the independent variables. Using matrix notation, it connects the geometric interpretation with the algebraic computation through the design matrix, normal equation, and projection matrices.

Finally, it highlights how the response vector is decomposed into two parts — the fitted values and the residuals — which are orthogonal to each other, forming the foundation of multivariable regression analysis.