Lecture 18 – Systems of Linear Equations in Regression and Least Squares Solutions

The vectors and matrices and their use in determining the least squares fitted model discussed in this lecture can be extended with few alterations to multiple linear regression, but for the sake of simplicity only simple linear regression will be considered.

The Model: $\boldsymbol{Y} = \beta_0 + \beta_1 x + \boldsymbol{\varepsilon}$

The Fitted Line: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

Example: Consider the bivariate date below for response variable $\boldsymbol{Y}$ and Predictor $\boldsymbol{X}$.

x	1	2	4	5
y	8	4	6	2

System of Linear Equations: The problem of fitting a line through the four points in the table above leads to the system of four linear equations in the two unknowns $\hat{\beta}_0$ and $\hat{\beta}_1$,

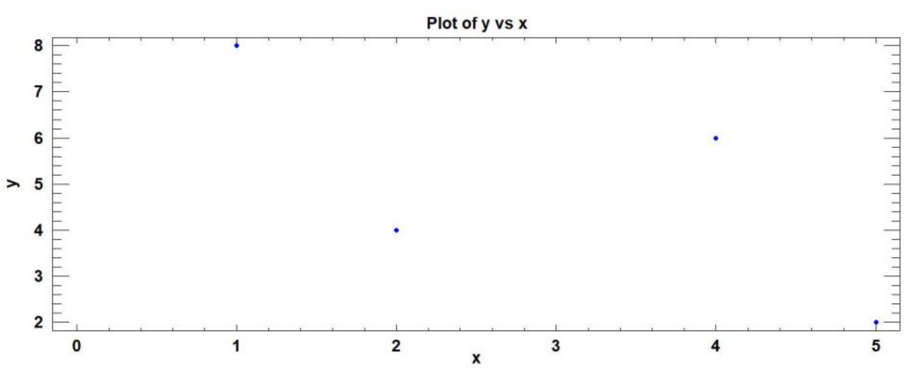
$$\begin{aligned} 8 &= \hat{\beta}_0 + \hat{\beta}_1 \\ 4 &= \hat{\beta}_0 + 2\hat{\beta}_1 \\ 6 &= \hat{\beta}_0 + 4\hat{\beta}_1 \\ 2 &= \hat{\beta}_0 + 5\hat{\beta}_1 \end{aligned}$$

Rewriting this in matrix form,

$$\underset{\mathrm{Y}}{\begin{bmatrix} 8 \\ 4 \\ 6 \\ 2 \end{bmatrix}} = \underset{\mathbf{X}}{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}} \underset{\hat{\beta}}{\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}}$$

, or more compactly as $\mathbf{X}\hat{\beta} = \mathrm{Y}$, where Y is the vector of observations, $\hat{\beta}$ is the vector of coefficients, and $\mathbf{X}$ is called the *Design Matrix*.

Ideally, we would proceed to solve the system for $\hat{\beta}_0$ and $\hat{\beta}_1$, but we already know that a solution doesn't exist. (Systems with more equations then unknowns usually have no solution. Additionally, each equation is a requirement that the solution, the line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, passes through a point, and we know by looking at the scatterplot of the data below that there is no choice of $\hat{\beta}_0$ and $\hat{\beta}_1$ leading to a line passing through all four points.)



We therefore proceed as follows. Multiplying the original matrix equation by $\mathbf{X}^T$, $\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathrm{Y}$, creates two *new* equations in $\hat{\beta}_0$ and $\hat{\beta}_1$ where the 2×2 matrix $\left(\mathbf{X}^T\mathbf{X}\right)$ is invertible. Then, multiplying both sides of the new equation by $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, we arrive at the solution to the new equations:

$$\boxed{\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{Y}}$$

This is a good time to remind ourselves that we haven't solve the *original* set of equations, which, in fact, weren't solvable. We have, however, found the intercept and slope of the least-squares regression line. (The proof of this was not done in class, but is done in the class notes and may be done in a later lecture.)

Example - continued: For the bivariate data in our example, the calculations proceed as follows:

- $\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 5 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix}$
- $\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \dfrac{1}{(4)(46)-(12)(12)}\begin{bmatrix} 46 & -12 \\ -12 & 4 \end{bmatrix} = \dfrac{1}{20}\begin{bmatrix} 23 & -6 \\ -6 & 2 \end{bmatrix}$
- $\mathbf{X}^T\mathrm{Y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 5 \end{bmatrix}\begin{bmatrix} 8 \\ 4 \\ 6 \\ 2 \end{bmatrix} = \begin{bmatrix} 20 \\ 50 \end{bmatrix}$
- $\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{Y} = \dfrac{1}{20}\begin{bmatrix} 23 & -6 \\ -6 & 2 \end{bmatrix}\begin{bmatrix} 20 \\ 50 \end{bmatrix} = \begin{bmatrix} 8 \\ -1 \end{bmatrix}$, so $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 8 \\ -1 \end{bmatrix}$ and $\hat{y} = 8 - x$ is the fitted line.

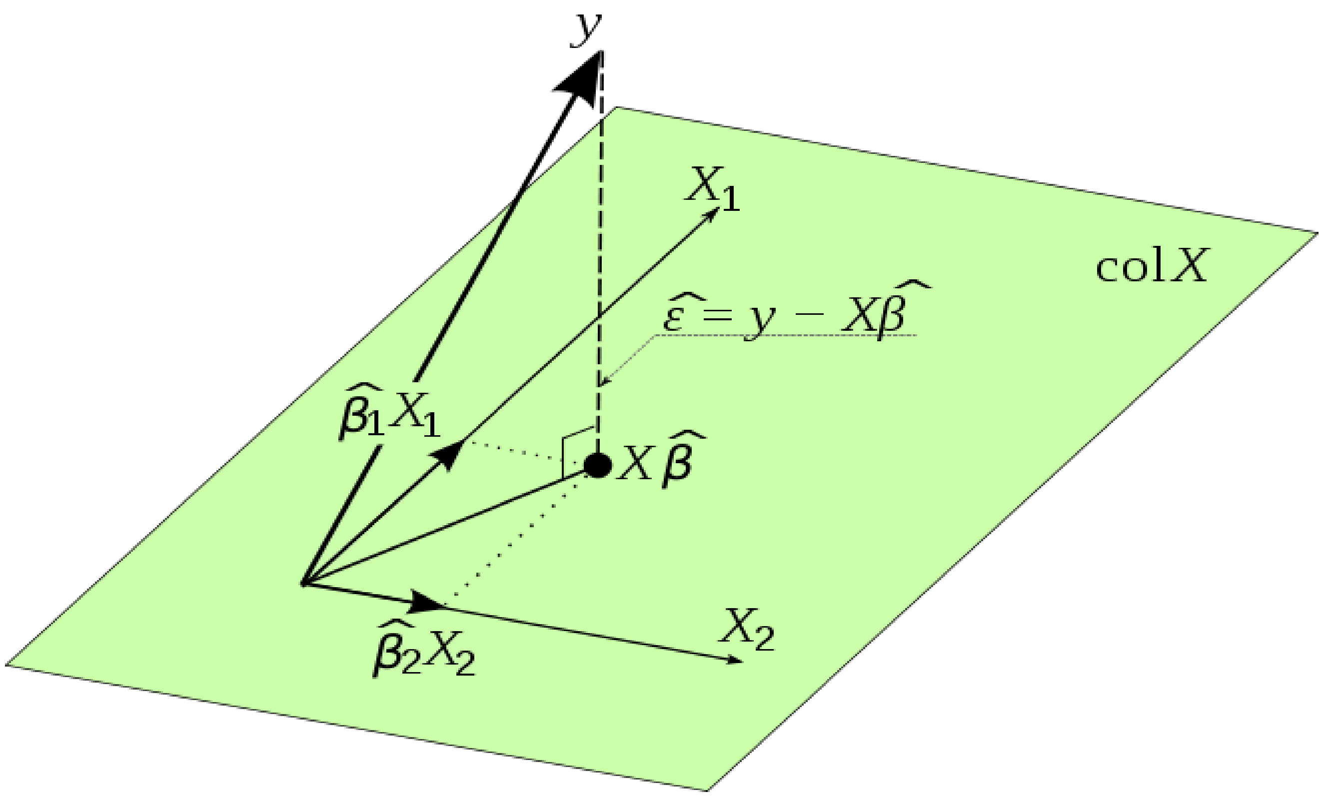
Finally, multiplying both sides of the last equation by the design matrix,

$$\mathbf{X}\hat{\beta} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{Y} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} 8 \\ -1 \end{bmatrix} = \begin{array}{c} \hat{\mathrm{Y}} \\ \begin{bmatrix} 7 \\ 6 \\ 4 \\ 3 \end{bmatrix} \end{array}$$

, where $\hat{\mathrm{Y}}$ is the vector of fitted values.

$\boxed{\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T}$ is called the **Hat Matrix** because it puts a hat on Y, $\mathbf{H}\mathrm{Y} = \hat{\mathrm{Y}}$.

The picture below, which we will keep returning to as we discuss the role of vectors and matrices in regression, helps to explain what we've done. The original system of four equations in two unknowns didn't have a solution because the vector of observations Y didn't lie in the plane defined by the columns of the design matrix $\mathbf{X}$, called the column space of $\mathbf{X}$ or $Col(\mathbf{X})$. The hat matrix $\mathbf{H}$ projects Y onto $Col(\mathbf{X})$, and the resulting projection $\hat{\mathrm{Y}}$ is the vector of values that the least-squares line fits to the data.



y
X_1
$\mathrm{col}\, X$
$\widehat{\varepsilon} = y - X\widehat{\beta}$
$\widehat{\beta}_1 X_1$
$X\widehat{\beta}$
X_2
$\widehat{\beta}_2 X_2$