Lecture 24 – Vectors and Matrices in Regression

This lecture is a follow-up to Lecture 23. In the previous lecture the original system of n linear equations in $\hat{\beta}_0$ and $\hat{\beta}_1$, $\mathbf{X}\hat{\beta} = \mathrm{Y}$, was manipulated into to a simpler system of two equations in the two unknowns $\hat{\beta}_0$ and $\hat{\beta}_1$, $\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathrm{Y}$. This last system has solution $\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{Y}$. It was claimed in Lecture 23 that this solution is the solution obtained with Calculus subject to the least-squares criterion. We now demonstrate this for the simple regression case.

Notation and Terminology:

- $\mathrm{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$ is the vector of observations on the response variable ***Y***.
- $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$ is the vector of regression coefficients.
- $\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ is the design matrix, where $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ and $\mathrm{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ are the column vectors of $\mathbf{X}$.

We now derive the necessary results:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_{\boldsymbol{n}} \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$\mathbf{X}^T\mathrm{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{\boldsymbol{n}} \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Then the simplified system of equations $\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$ becomes $\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$ and leads

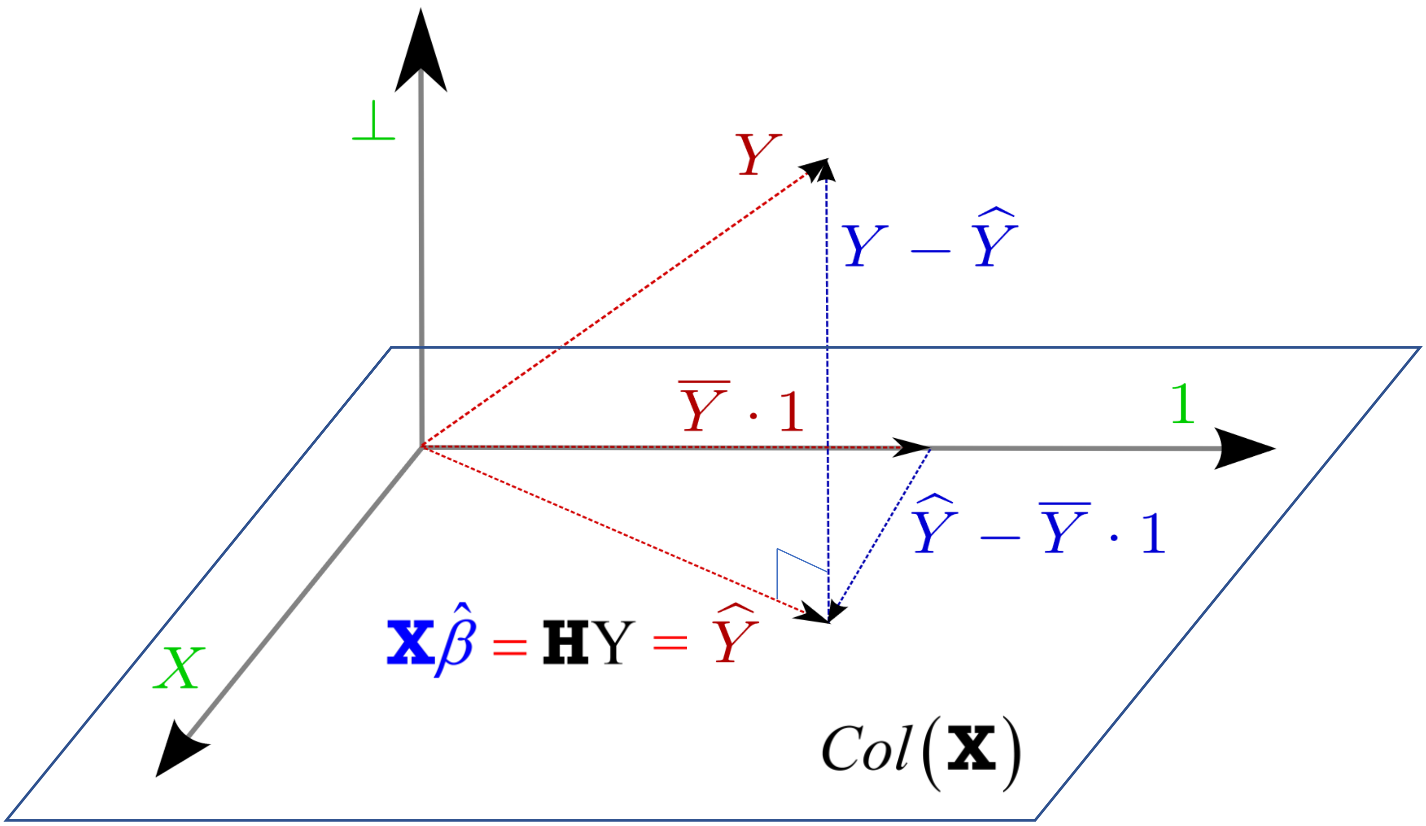
to
$$\begin{bmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \\ \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \quad \text{or} \quad \boxed{\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i &= \sum Y_i \\ \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 &= \sum X_i Y_i \end{aligned}}$$

The final (boxed) system of two equations in $\hat{\beta}_0$ and $\hat{\beta}_1$ is the same system we solved previously when minimizing the sum of squared residuals (the least squares criterion) using calculus. This shows that the matrix approach used in Lecture 23, which uses the Hat matrix $\mathbf{H}$ to project the vector of observations Y onto the column space of the design matrix $\mathbf{X}$, produces the least-squares fitted line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

Example: Continuing our running example from previous lectures with data

x	1	2	4	5
y	8	4	6	2

$$\overset{\mathbf{X}}{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}} \overset{\hat{\beta}}{\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}} = \overset{\mathrm{Y}}{\begin{bmatrix} 8 \\ 4 \\ 6 \\ 2 \end{bmatrix}} \quad \text{and} \quad \mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix}, \quad \mathbf{X}^T\mathrm{Y} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} 20 \\ 50 \end{bmatrix}$$



As pictured above, the Hat Matrix $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$ projects the vector of observations Y onto the Column Space of the design matrix $\mathbf{X}$, also called the Solution Space because the original system of equations $\mathbf{X}\hat{\beta} = \mathrm{Y}$ has a solution *if and only if* Y lies in $Col(\mathbf{X})$. The projected vector $\hat{\mathrm{Y}} = \mathbf{H}\mathrm{Y} = \mathbf{X}\hat{\beta}$ is the vector of fitted values. (**Note:** By construction, $\hat{\mathrm{Y}} = \mathbf{H}\mathrm{Y}$ always lies in the solution space.).