Lecture 22

Multiple Linear Regression I

An Introduction to Matrices

A matrix is an array such as the following

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ or } \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} \text{ or } \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

The entries are indexed by their row and column number. In the matrix,

$$\mathbf{A} = \begin{bmatrix} 12 & 5 \\ 8 & 2 \\ 1.2 & 1.1 \end{bmatrix}$$

$A_{1,1} = 12$, $A_{2,2} = 2$, and $A_{3,2} = 1.1$.

Matrices are added entry-wise:

$$\begin{bmatrix} 12 & 5 \\ 8 & 2 \\ 1.2 & 1.1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 13 & 5 \\ 9 & 2 \\ 2.2 & 2.1 \end{bmatrix}$$

Thus, to be added together, two matrices must have the same dimensions.

Matrix multiplication is useful but a little more complicated. To be multiplied as $\mathbf{A} \times \mathbf{B}$ the number of columns of $\mathbf{A}$ must equal the number of rows of $\mathbf{B}$. If $\mathbf{A}$ has dimensions $n \times k$ and $\mathbf{B}$ has dimensions $k \times m$, then their product has dimensions $n \times m$. The entries in the matrix product are the sum of the products of the row entries in $\mathbf{A}$ and column entries in $\mathbf{B}$ as in the following examples:

$$\begin{bmatrix} 1 & 5 \\ 8 & 2 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + 5 \times 4 & 1 \times 2 + 5 \times 5 & 1 \times 3 + 5 \times 6 \\ 8 \times 1 + 2 \times 4 & 8 \times 2 + 2 \times 5 & 8 \times 3 + 2 \times 6 \\ 1 \times 1 + 1 \times 4 & 1 \times 2 + 1 \times 5 & 1 \times 3 + 1 \times 6 \end{bmatrix} = \begin{bmatrix} 21 & 27 & 33 \\ 16 & 26 & 36 \\ 5 & 7 & 9 \end{bmatrix}$$

In the previous example a 3×2 and a 2×3 matrix were multiplied to give a 3×3 product.

This is useful when thinking about regression because regression can be put into a matrix format as follows.

Matrix Formulation of Simple Linear Regression

For simple regression, we have paired data (X_1, Y_1), (X_2, Y_2), $\ldots$, (X_n, Y_n). A matrix formulation of regression is the following:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

How do we solve for a and b? Just like with regular multiplication, we need to be able to invert (divide). The inverse of a number is what you multiply the number by to get 1. In the case of matrices, the "1" is the identity matrix

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

It is a matrix with 1's down the diagonal and zeros elsewhere. If $\mathbf{A}$ has an inverse, it is denotes $\mathbf{A}^{(-1)}$ and it is the matrix so that

$$\mathbf{A} \times \mathbf{A}^{(-1)} = \mathbf{A}^{(-1)} \times \mathbf{A} = \mathbf{I}.$$

Among other things, this forces $\mathbf{A}$ to be square (to have the same number of rows as it has columns).

So, to solve for a and b above, we need to invert, but there is no square matrix around. To get a square matrix, we consider the transpose of a matrix which is the matrix you get by flipping rows and columns:

$$\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix}$$

Multiplying the regression equation through by this matrix we get

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \times \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \times \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \times \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or, carrying out the matrix multiplication

$$\begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} \sum \epsilon_i \\ \sum \epsilon_i X_i \end{bmatrix}$$

Now, we have a square matrix and we can invert it. There is an easy formula for the inverse of a 2×2 matrix. You can try it out and see that it works:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{(-1)} = \begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$$

Thus, in the regression equation, we see that

$$\begin{bmatrix} \frac{\sum X_i^2}{n\sum X_i^2 - (\sum X_i)^2} & \frac{-\sum X_i}{n\sum X_i^2 - (\sum X_i)^2} \\ \frac{-\sum X_i}{n\sum X_i^2 - (\sum X_i)^2} & \frac{n}{n\sum X_i^2 - (\sum X_i)^2} \end{bmatrix} \times \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} \frac{\sum X_i^2}{n\sum X_i^2 - (\sum X_i)^2} & \frac{-\sum X_i}{n\sum X_i^2 - (\sum X_i)^2} \\ \frac{-\sum X_i}{n\sum X_i^2 - (\sum X_i)^2} & \frac{n}{n\sum X_i^2 - (\sum X_i)^2} \end{bmatrix} \times \begin{bmatrix} \sum \epsilon_i \\ \sum \epsilon_i X_i \end{bmatrix}$$

Since the residuals sum to zero and are independent of the X's, the best guess for the coefficients is

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \frac{\sum X_i^2}{n\sum X_i^2 - (\sum X_i)^2} & \frac{-\sum X_i}{n\sum X_i^2 - (\sum X_i)^2} \\ \frac{-\sum X_i}{n\sum X_i^2 - (\sum X_i)^2} & \frac{n}{n\sum X_i^2 - (\sum X_i)^2} \end{bmatrix} \times \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2} \\ \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \end{bmatrix}$$

These are the same formulas as before.

The point here is that for more than just one predictor variables, regression is best thought of in terms of matrices. Further, the fact that the coefficients are found by inverting matrices will explain some of the difficulties one sometimes encounters with regressions in practice; difficulties that we will discuss in a later section.

Matrix Formulation of Multiple Linear Regression

For more than one predictor, we need to use other notation for intercepts and slopes. If we have k predictors, then we write $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k$ and the matrix formulation of our data, which is now of the form $(X_{11}, X_{21}, \ldots, X_{k1}, Y_1)$, $(X_{12}, X_{22}, \ldots, X_{k2}, Y_2)$, $\ldots$, $(X_{1n}, X_{2n}, \ldots, X_{kn}, Y_n)$, is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Using bold letters for the corresponding matrices, we write the above equation simply as

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

Then our best guess for the coefficients is

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

which is how the computer finds the estimates of the coefficients for you. The estimate of the variance-covariance matrix for the coefficients is

$$\text{var}(\beta) = (\mathbf{X}^T\mathbf{X})^{(-1)}\sigma^2$$

Interpretation of Regression Coefficients in Multiple Linear Regression

The slope coefficient for a particular predictor measure how much the response changes to a one-unit increase in that predictor IF ALL ELSE REMAINS THE SAME. That is, after controlling for all other predictor variables and fixing their values, the slope of a particular predictor tells how much influence that predictor has on the response.

That is, for $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k$, the slope β_1 is how much the response Y increases with one unit of change in X_1 if all other predictor variables stay the same.

This is how we know smoking causes lung cancer; study after study controlling for all kinds of predictors such as age, occupation, socioeconomic status, etc.., still give a significant impact of smoking on lung cancer.

Exercises for Lecture 22

1. –

2. –