A Matrix Algebra Primer

Matrices, Vectors and Scalar Multiplication

The matrix, **D**, represents data organized into rows and columns where the rows represent one variable, e.g. time, and the columns represent a second variable, e.g. mass. Each intersection of a row and column has a numeric value called an element, $d_{r,c}$. Each element represents a parameter whose value depends upon the row and column variables, e.g. counts. A column vector, **v**, can be thought of as a one-column matrix. Unless otherwise noted, all vectors are assumed to be column vectors. Multiplication by a scalar is straightforward.

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 4 & 2 \\ 8 & 4 \\ 4 & 8 \\ 2 & 4 \\ 1 & 2 \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 8 \\ 4 \\ 2 \\ 1 \end{pmatrix} \qquad 2\mathbf{v} = \begin{pmatrix} 2 \\ 4 \\ 8 \\ 16 \\ 8 \\ 4 \\ 2 \end{pmatrix}$$

Transpose

The transpose, **E**, of the matrix, **D**, is obtained by converting each row into a corresponding column. The transpose, **w**, of a column vector, **v**, is a row vector. To save space in written material a vector, **v**, might be defined using the notation, $\mathbf{v}^T =$. This allows it to be written horizontally, as shown below, instead of consuming vertical space.

$$\mathbf{E} = \mathbf{D}^T = \begin{pmatrix} 1 & 2 & 4 & 8 & 4 & 2 & 1 \\ 0 & 1 & 2 & 4 & 8 & 4 & 2 \end{pmatrix} \qquad \mathbf{w} = \mathbf{v}^T = \begin{pmatrix} 1 & 2 & 4 & 8 & 4 & 2 & 1 \end{pmatrix}$$

Vector Dot Product

The dot product of two vectors is the sum of the element-by-element products. The result is a scalar.

$$\mathbf{v1} = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 8 \\ 4 \\ 2 \\ 1 \end{pmatrix} \qquad \mathbf{v2} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 4 \\ 8 \\ 4 \\ 2 \end{pmatrix} \qquad s = \mathbf{v1} \bullet \mathbf{v2} = 1 \times 0 + 2 \times 1 + 4 \times 2 + \cdots + 1 \times 2 = 84$$

Matrix Multiplication

Matrix multiplication depends upon the ordering of the two matrices. That is, **DE** does not ordinarily give the same result as **ED**. Additionally, the number of **columns of the left multiplicand** has to be equal to the **number of rows of the right multiplicand**. The number of **rows of the product matrix** is equal to the **number of rows of the left multiplicand**. The **number of columns of the product matrix** is equal to the **number of columns of the right multiplicand**. A useful matrix nomenclature provides the number of rows and columns below the matrix symbols, (rows, columns). Note that this can be used to check row and column restrictions.

$$\underset{(2,2)}{\mathbf{P1}} = \underset{(2,7)}{\mathbf{E}}\;\underset{(7,2)}{\mathbf{D}} \qquad\qquad \underset{(7,7)}{\mathbf{P2}} = \underset{(7,2)}{\mathbf{D}}\;\underset{(2,7)}{\mathbf{E}}$$

Using the above matrices, the easiest multiplication to perform by hand is **P2**.

$$\begin{pmatrix} 1 & 2 & 4 & 8 & 4 & 2 & 1 \\ 0 & 1 & 2 & 4 & 8 & 4 & 2 \end{pmatrix}$$

$$\mathbf{P2} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 4 & 2 \\ 8 & 4 \\ 4 & 8 \\ 2 & 4 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 & 8 & 4 & 2 & 1 \\ 2 & 5 & 10 & 20 & 16 & 8 & 4 \\ 4 & 10 & 20 & 40 & 32 & 16 & 8 \\ 8 & 20 & 40 & 80 & 64 & 32 & 16 \\ 4 & 16 & 32 & 64 & 80 & 40 & 20 \\ 2 & 8 & 16 & 32 & 40 & 20 & 10 \\ 1 & 4 & 8 & 16 & 20 & 10 & 5 \end{pmatrix}$$

Thus, $p2_{1,1}$, is the dot product of the first row of **D** and the first column of **E**. Likewise, $p2_{1,2}$ is the dot product of the first row of **D** and the second column of **E**, and $p2_{2,1}$ is the dot product of the second row of **D** and the first column of **E**. It takes a lot more math to compute **P1**.

$$\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 4 & 2 \\ 8 & 4 \\ 4 & 8 \\ 2 & 4 \\ 1 & 2 \end{pmatrix}$$

$$\mathbf{P1} = \begin{pmatrix} 1 & 2 & 4 & 8 & 4 & 2 & 1 \\ 0 & 1 & 2 & 4 & 8 & 4 & 2 \end{pmatrix} \begin{pmatrix} 106 & 84 \\ 84 & 105 \end{pmatrix}$$

Matrix Inversion

There is no such thing as matrix division. Instead one computes the inverse of a matrix and multiplies by the inverse. The procedure used to invert a matrix is complicated and at the "Primer" level not worth worrying about. The inverse, **I1**, of **P1**, is given by the following. Note that **I1×P1** = **P1×I1** = 1. (The symbol × as used here does not denote a cross-product.)

$$\mathbf{I1} = \begin{pmatrix} 0.0258 & -0.0206 \\ -0.0206 & 0.0260 \end{pmatrix} \qquad \mathbf{P1} = \begin{pmatrix} 106 & 84 \\ 84 & 105 \end{pmatrix}$$

$$\mathbf{I1} \times \mathbf{P1} = \mathbf{P1} \times \mathbf{I1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1$$

An inverse is only defined for a square matrix. Not all square matrices have an inverse.

Diagonal Matrix

A diagonal matrix has zero elements except along the matrix diagonal. An identity matrix has a diagonal of all ones.

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 10 \end{pmatrix} \qquad \mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Least-Squares Matrix Formalism

Suppose that a set of data are fit to a quadratic equation, $y = a_0 + a_1x + a_2x^2$. If the x-values chosen were 0, 1, 2, 3 and 4, a predictor variable matrix, **X**, can be defined. Let the data vector be denoted **y** and a coefficient vector by **a**.

$$\mathbf{y} = \begin{pmatrix} 0.1 \\ 1.05 \\ 3.88 \\ 9.16 \\ 15.8 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix} \qquad \mathbf{a} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

The least-squares relationship can then be written using matrix notation.

$$\underset{(5,1)}{\mathbf{y}} = \underset{(5,3)}{\mathbf{X}} \; \underset{(3,1)}{\mathbf{a}}$$

Since **X** is not square, the matrix equation has to be solved using the following steps.

$$\underset{(5,1)}{\mathbf{y}} = \underset{(5,3)}{\mathbf{X}}\,\underset{(3,1)}{\mathbf{a}}$$

$$\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,1)}{\mathbf{y}} = \left(\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,3)}{\mathbf{X}}\right)\underset{(3,1)}{\mathbf{a}}$$

$$\left(\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,3)}{\mathbf{X}}\right)^{-1}\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,1)}{\mathbf{y}} = \left(\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,3)}{\mathbf{X}}\right)^{-1}\left(\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,3)}{\mathbf{X}}\right)\underset{(3,1)}{\mathbf{a}}$$

$$\underset{(3,1)}{\mathbf{a}} = \left(\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,3)}{\mathbf{X}}\right)^{-1}\underset{(3,5)}{\mathbf{X}^T}\,\underset{(5,1)}{\mathbf{y}}$$

If you want to "follow along" with the least-squares calculation here are the associated numeric values.

$$\mathbf{X}^T\mathbf{y} = \begin{pmatrix} 29.99 \\ 99.49 \\ 351.81 \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 5 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{pmatrix}$$

$$\mathbf{C} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 0.88571 & -0.77143 & 0.14286 \\ -0.77143 & 1.24286 & -0.28571 \\ 0.14286 & -0.28571 & 0.071429 \end{pmatrix}$$

$$a = \begin{pmatrix} 0.0717 \\ -0.000429 \\ 0.988 \end{pmatrix}$$

The computed y-values, $\hat{\mathbf{y}}$, and the residuals, $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$, are given below.

$$\hat{\mathbf{y}}^T = \begin{pmatrix} 0.0717 & 1.059 & 4.022 & 8.961 & 15.878 \end{pmatrix}$$

$$\mathbf{r}^T = \begin{pmatrix} 0.0283 & -0.00914 & -0.142 & 0.199 & -0.0757 \end{pmatrix}$$

C is called the variance-covariance matrix,. If this matrix is multiplied by the variance of the fit, σ^2, the diagonal elements, $c_{i,i}$ are the variance of the coefficients and the off-diagonal elements, $c_{i,j}$, are the covariance of the coefficients. Note that $c_{i,j} = c_{j,i}$, that is, **C** is symmetric about the diagonal thus cov(a_0,a_1) = cov(a_1,a_0).

$$\sigma_{a_0}^2 = \sigma^2 c_{1,1} \quad \sigma_{a_1}^2 = \sigma^2 c_{2,2} \quad \sigma_{a_2}^2 = \sigma^2 c_{3,3}$$

$$\operatorname{cov}(a_0, a_1) = \sigma^2 c_{1,2} \quad etc.$$

The standard deviation of the fit (in R the residual standard error) is obtained from the squared residuals and the degrees of freedom (number of data values minus the number of computed parameters).

$$s_{fit} = \sqrt{\frac{\mathbf{r} \bullet \mathbf{r}}{5-3}} = \sqrt{\frac{\sum_{i=1}^{5}\left(y_i - \hat{y}_i\right)^2}{2}} = 0.182$$

As noted above, the diagonal of **C** is used to compute coefficient errors.

$$\begin{aligned} s_{a_0} &= s_{fit}\sqrt{c_{1,1}} = 0.182 \times 0.941 = 0.171 \\ s_{a_1} &= s_{fit}\sqrt{c_{2,2}} = 0.182 \times 1.115 = 0.203 \\ s_{a_2} &= s_{fit}\sqrt{c_{3,3}} = 0.182 \times 0.267 = 0.0487 \end{aligned}$$

The Hat Matrix

The hat matrix, H, is given by,

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$$

It is given this name since it transforms the experimental y-values into the computed least-squares y-values.

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$$

The hat matrix is useful for recognizing outliers in linear regression.

Associated R Code

See the R file, “Linear Algebra Primer.R” to see how all of the above computations are achieved in R.