Simple Linear Regression using Matrices

Math 158, Spring 2009
Jo Hardin
Simple Linear Regression
with Matrices

Everything we've done so far can be written in matrix form. Though it might seem no more efficient to use matrices with simple linear regression, it will become clear that with multiple linear regression, matrices can be very powerful. Chapter 5 contains a lot of matrix theory; the main take away points from the chapter have to do with the matrix theory *applied to the regression setting*. Please make sure that you read the chapters / examples having to do with the regression examples.

Special Matrices

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{J} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \mathbf{0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Regression Model

Matrix Addition

$$
\begin{aligned}
Y_i &= E[Y_i] + \epsilon_i \\
\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} &= \begin{pmatrix} E[Y_1] \\ E[Y_2] \\ \vdots \\ E[Y_n] \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \\
\underline{Y} &= E[\underline{Y}] + \underline{\epsilon}
\end{aligned}
$$

Matrix Multiplication

Example

Consider multiplying an $r \times c$ matrix with a $c \times s$ matrix. The *interior* dimensions must always be the same. The resulting matrix will always be $r \times s$. The element in the i^{th} row and j^{th} column is given by:

$$\sum_{k=1}^{c} a_{ik} b_{kj}$$

$$
\begin{aligned}
AB &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix} \\
AB &= \begin{pmatrix} 3 & -1 & 0 \\ 0 & 1 & 1 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 2 \\ 0 & -1 \end{pmatrix} \\
&= \begin{pmatrix} 4 & 1 \\ -1 & 1 \\ -2 & -3 \end{pmatrix}
\end{aligned}
$$

and in the linear regression context...

$$
\begin{aligned}
E[Y_i] &= \beta_0 + \beta_1 X_i \\
\begin{pmatrix} E[Y_1] \\ E[Y_2] \\ \vdots \\ E[Y_n] \end{pmatrix} &= \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \\
E[\underline{Y}] &= X\underline{\beta}
\end{aligned}
$$

$$
\begin{aligned}
\underline{Y}^t\underline{Y} &= \begin{pmatrix} Y_1 & Y_2 & \cdots & Y_n \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \sum_{i=1}^{n} Y_i^2 \\
X^t X &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \\
X^t\underline{Y} &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}
\end{aligned}
$$

Matrix Inverses in the Regression Context

Matrix Inverses

An $n \times n$ matrix A is called *invertible* if there exists an $n \times n$ matrix B such that

$$AB = BA = I_n$$

B is called the *inverse* of A and is typically denoted by $B = A^{-1}$. (Note, inverses only exist for square matrices with non-zero determinants.)

Example

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad A^{-1} = \begin{pmatrix} d/(ad-bc) & -b/(ad-bc) \\ -c/(ad-bc) & a/(ad-bc) \end{pmatrix}$$

where the determinant is given by $D = ad - bc$.

$$\begin{aligned} A^{-1}A &= \begin{pmatrix} d/(ad-bc) & -b/(ad-bc) \\ -c/(ad-bc) & a/(ad-bc) \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} (ad-bc)/(ad-bc) & (bd-bd)/(ad-bc) \\ (-ac+ac)/(ad-bc) & (ad-bc)/(ad-bc) \end{pmatrix} \\ &= I_2 \end{aligned}$$

$$A = \begin{pmatrix} 3 & 2 \\ 1 & 6 \end{pmatrix} \quad A^{-1} = \begin{pmatrix} 6/16 & -2/16 \\ -1/16 & 3/16 \end{pmatrix}$$

$$\begin{aligned} A^{-1}A &= \begin{pmatrix} 6/16 & -2/16 \\ -1/16 & 3/16 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 1 & 6 \end{pmatrix} \\ &= \begin{pmatrix} (18-2)/16 & (12-12)/16 \\ (-3+3)/16 & (-2+18)/16 \end{pmatrix} \\ &= I_2 \end{aligned}$$

Variance of Coefficients

$$\begin{aligned} \text{Recall: } X^tX &= \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \\ \text{So, } D &= n\sum X_i^2 - (\sum X_i)^2 = n\sum (X_i - \overline{X})^2 \\ (X^tX)^{-1} &= \begin{pmatrix} \frac{\sum X_i^2}{n\sum(X_i-\overline{X})^2} & \frac{-\sum_{i=1}^n X_i}{n\sum(X_i-\overline{X})^2} \\ \frac{-\sum_{i=1}^n X_i}{n\sum(X_i-\overline{X})^2} & \frac{n}{n\sum(X_i-\overline{X})^2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i-\overline{X})^2} & \frac{-\overline{X}}{\sum(X_i-\overline{X})^2} \\ \frac{-\overline{X}}{\sum(X_i-\overline{X})^2} & \frac{1}{\sum(X_i-\overline{X})^2} \end{pmatrix} \\ \sigma^2\{\underline{b}\} &= \sigma^2 \cdot (X^tX)^{-1} \end{aligned}$$

Estimating Coefficients

Recall the normal equations that come from differentiating the sum of squared residuals with respect to both β_0 and β_1:

$$\begin{aligned}
nb_0 + b_1 \sum X_i &= \sum Y_i \\
b_0 \sum X_i + b_1 \sum X_i^2 &= \sum X_i Y_i \\
\begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} &= \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix} \\
(X^t X)\underline{b} &= X^t \underline{Y} \\
\underline{b} &= (X^t X)^{-1}(X^t \underline{Y})
\end{aligned}$$

$$\begin{aligned}
\text{checking:} & \\
(X^t X)^{-1}(X^t \underline{Y}) &= \begin{pmatrix} \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} & \frac{-\overline{X}}{\sum (X_i - \overline{X})^2} \\ \frac{-\overline{X}}{\sum (X_i - \overline{X})^2} & \frac{1}{\sum (X_i - \overline{X})^2} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} \\
&= \begin{pmatrix} \frac{\sum Y_i}{n} + \frac{\sum Y_i \overline{X}^2}{\sum (X_i - \overline{X})^2} + \frac{-\sum X_i Y_i (\overline{X})}{\sum (X_i - \overline{X})^2} \\ \frac{-\sum Y_i (\overline{X})}{\sum (X_i - \overline{X})^2} + \frac{\sum X_i Y_i}{\sum (X_i - \overline{X})^2} \end{pmatrix} \\
&= \begin{pmatrix} \overline{Y} - b_1 \overline{X} \\ \frac{\sum Y_i (X_i - \overline{X})}{\sum (X_i - \overline{X})^2} \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}
\end{aligned}$$

Fitted Values

$$\hat{Y}_i = b_0 + b_1 X_i$$

$$\begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

$$\begin{aligned}
\underline{\hat{Y}} &= X\underline{b} \\
&= X(X^t X)^{-1}(X^t \underline{Y}) \\
&= H\underline{Y} \\
\text{“hat” matrix: } H &= X(X^t X)^{-1} X^t
\end{aligned}$$

Note that the predicted values are simply a *linear combinations* of the response variable (Y) with *coefficients* of the explanatory variables (X).

Residuals

$$
\begin{aligned}
e_i &= Y_i - \hat{Y}_i \\
\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} &= \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} - \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} \\
\underline{e} &= \underline{Y} - \hat{\underline{Y}} \\
&= \underline{Y} - X\underline{b} \\
&= \underline{Y} - H\underline{Y} \\
&= (I - H)\underline{Y}
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2\{\underline{e}\} &= \sigma^2\{(I - H)\underline{Y}\} \\
&= (I - H)\sigma^2\{\underline{Y}\} \\
&= (I - H)\sigma^2 \\
s^2(\underline{e}) &= MSE \cdot (I - H)
\end{aligned}
$$

ANalysis Of VAriance

$$
\begin{aligned}
SSTO &= \underline{Y}^t\underline{Y} - \left(\frac{1}{n}\right)\underline{Y}^t J \underline{Y} \\
SSE &= \underline{Y}^t\underline{Y} - \underline{b}^t X^t \underline{Y} \\
SSR &= \underline{b}^t X^t \underline{Y} - \left(\frac{1}{n}\right)\underline{Y}^t J \underline{Y}
\end{aligned}
$$

Prediction of New Observations

$$
\underline{X}_h = \begin{pmatrix} 1 \\ X_h \end{pmatrix} \quad \underline{X}_h^t = \begin{pmatrix} 1 & X_h \end{pmatrix}
$$

$$
\begin{aligned}
\hat{Y}_h &= \underline{X}_h^t \underline{b} \\
\sigma^2\{\hat{Y}_h\} &= \underline{X}_h^t \sigma^2\{\underline{b}\} \underline{X}_h \\
&= \sigma^2 \cdot \underline{X}_h^t (X^t X)^{-1} \underline{X}_h \\
s^2\{\hat{Y}_h\} &= MSE \cdot \underline{X}_h^t (X^t X)^{-1} \underline{X}_h
\end{aligned}
$$

$$
s^2\{\hat{Y}_{h(new)}\} = MSE \cdot (1 + \underline{X}_h^t (X^t X)^{-1} \underline{X}_h)
$$

R Code for Dealing with Matrices

Addition

```
> matrix1 <- matrix(c(1:12),ncol=4, byrow=T)
```

1	2	3	4
5	6	7	8
9	10	11	12

```
> matrix2 <- matrix(seq(2,24,by=2),ncol=4, byrow=T)
```

2	4	6	8
10	12	14	16
18	20	22	24

```
> matrix1 + matrix2
```

gives exactly what you'd expect, element by element addition.

3	6	9	12
15	18	21	24
27	30	33	36

Multiplication

```
> matrix1 * matrix2
```

gives element by element multiplication.

2	8	18	32
50	72	98	128
162	200	242	288

```
> matrix1 %*% matrix2
```

gives an error because you can't multiply a 3×4 by a 3×4 .

```
> matrix1 %*% t(matrix2)
```

gives matrix multiplication of the first matrix times the transpose of the second (3×3).

60	140	220
140	348	556
220	556	892

```
> t(matrix1) %*% matrix2
```

gives matrix multiplication of the transpose of the first matrix times the second (4×4).

214	244	274	304
244	280	316	353
274	316	358	400
304	353	400	448

Note that the products are only symmetric because matrix1 = 2 matrix2.

Taking Inverses

The function for inverting matrices in R is `solve`. Remember that solve only works on square matrices with non-zero determinants.

```
> matrix3 <- matrix(c(5,7,1,4,3,6,2,0,8), ncol=3, byrow=T)
```

5	7	1
4	3	6
2	0	8

```
> solve(matrix3)
```

gives the inverse of matrix3.

-0.923	2.154	-1.5
0.769	-1.462	1.0
0.231	-0.538	0.5

```
> solve(matrix3) %*% matrix3
```

gives the identity matrix, I_3.

1	$8.9e-16$	$-1.78e-15$
0	$1.0e+00$	$1.78e-15$
0	$0.0e+00$	$1.0e+00$