Chapter 4

Matrix Algebra

4.1 Utility of Matrices

Consider a system of equations with coefficients $a_{11}, a_{12}, \ldots, a_{mn}$, unknown values $x_1, x_2, \ldots, x_n$, and constants $b_1, b_2, \ldots, b_n$.

Each of the m equations describes a hyper-plane slicing through n-dimensional space. The x-values that satisfy all these equations at the same time give coordinates of all the intersection points of the m hyper-planes:

We can express this system more compactly with *matrices*:

There are at least two good ways to interpret $\mathbf{A}\mathbf{x} = \mathbf{b}$. One is that the vectors $\mathbf{x}$ that solve $\mathbf{A}\mathbf{x} = \mathbf{b}$ give the coordinates where the hyperplanes described by $\sum_{j=1}^{n} a_{ij}x_j = b_i$ intersect.

Another is to see that $\mathbf{A}$ *operates* on the vector $\mathbf{x}$ by rotating and stretching or shrinking it so that it becomes the vector $\mathbf{b}$. This is accomplished by using the elements in the $\mathbf{x}$ vector to form a linear combination of the column vectors of $\mathbf{A}$:

An $m \times n$ matrix is a set of numbers arranged in an array of m rows and n columns:

$$\mathbf{A}_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The dimension of the matrix $\mathbf{A}$ is $m \times n$, and is often displayed underneath the matrix name as a reminder. A matrix is square if $m = n$.

A column vector is a matrix with only one column. A row vector is a matrix with only one row.

4.2 Transposes and Basic Operations

The transpose $\mathbf{A}^T$ of the matrix $\mathbf{A}$ is obtained by interchanging the rows and columns of $\mathbf{A}$.

The transpose of an $m \times n$ matrix has dimension $n \times m$.

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are equal iff they have the same dimension and their corresponding elements are equal: $a_{ij} = b_{ij}$ for all i, j.

A matrix that equals its transpose is called *symmetric*. All symmetric matrices are square matrices.

We can add or subtract matrices with the same dimensions. For two $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A}_{m\times n} + \mathbf{B}_{m\times n} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}.$$

Similarly,

$$\mathbf{A}_{m\times n} - \mathbf{B}_{m\times n} = \begin{pmatrix} a_{11}-b_{11} & a_{12}-b_{12} & \cdots & a_{1n}-b_{1n} \\ a_{21}-b_{21} & a_{22}-b_{22} & \cdots & a_{2n}-b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}-b_{m1} & a_{m2}-b_{m2} & \cdots & a_{mn}-b_{mn} \end{pmatrix}.$$

A *scalar* is a number. Given the $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

and a scalar k, the product $k\mathbf{A}$ is defined to be

$$k\mathbf{A} = \begin{pmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{pmatrix}$$

Given matrices $\mathbf{A}_{m\times p}$ and $\mathbf{B}_{p\times n}$, their product $\mathbf{A}_{m\times p}\mathbf{B}_{p\times n}$ is defined by

$$\mathbf{A}_{m\times p}\mathbf{B}_{p\times n} = \begin{pmatrix} \sum_{k=1}^{p} a_{1k}b_{k1} & \sum_{k=1}^{p} a_{1k}b_{k2} & \cdots & \sum_{k=1}^{p} a_{1k}b_{kn} \\ \sum_{k=1}^{p} a_{2k}b_{k1} & \sum_{k=1}^{p} a_{2k}b_{k2} & \cdots & \sum_{k=1}^{p} a_{2k}b_{kn} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{p} a_{mk}b_{k1} & \sum_{k=1}^{p} a_{mk}b_{k2} & \cdots & \sum_{k=1}^{p} a_{mk}b_{kn} \end{pmatrix}$$

When obtaining the product $\mathbf{AB}$, we say "$\mathbf{B}$ is pre-multiplied by $\mathbf{A}$", or "$\mathbf{A}$ is post-multiplied by $\mathbf{B}$". Note the dimension of $\mathbf{A}_{m\times p}\mathbf{B}_{p\times n}$ is $m \times n$. In general, $\mathbf{AB} \neq \mathbf{BA}$.

4.3 Some Special Matrices

A *diagonal matrix* is a square matrix whose only non-zero elements are along the diagonal.

The *identity matrix* $\mathbf{I}$ is a diagonal matrix whose diagonal elements are 1s.

Pre- or post-multiplying a matrix $\mathbf{A}$ by $\mathbf{I}$ gives $\mathbf{A}$ (provided $\mathbf{I}$ has the appropriate dimension for carrying out the multiplication). So identity matrices are multiplicative identity matrices:

$$\mathbf{I}_{m\times m}\mathbf{A}_{m\times n} = \mathbf{A}_{m\times n} = \mathbf{A}_{m\times n}\mathbf{I}_{n\times n}.$$

A *scalar matrix* is a diagonal matrix whose main diagonal elements are the same and all off-diagonal elements are 0s. They can be expressed as the product of a scalar and $\mathbf{I}$:

We use the notation $\mathbf{1}_{m\times 1}$, $\mathbf{J}_{m\times m}$, and $\mathbf{0}_{m\times 1}$ to mean

$$\mathbf{1}_{m\times 1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \qquad \mathbf{J}_{m\times m} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \qquad \mathbf{0}_{m\times 1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Example 27

4.4 Linear Independence and Inverses

Columns of a matrix are *linearly dependent* if one of them is a linear combination of the others. Think of the columns in a matrix as vectors $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_n$. When n scalars $k_1, \ldots, k_n$, not all zero, can be found such that

$$\sum_{i=1}^{n} k_i \mathbf{C}_i = 0, \tag{4.1}$$

the n column vectors are *linearly dependent.* If the only solution to (4.1) is $k_i = 0$ $\forall i$, the n column vectors are *linearly independent.*

The *rank* of a matrix is the maximum number of linearly independent columns. Rank is equivalently defined as the number of of linearly independent rows.

The *inverse* of a square matrix $\mathbf{A}$ is denoted A^{-1} and satisfies

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

when it exists. A matrix without an inverse is *noninvertible* or *singular*. A square matrix with rank less than the number if its columns (or rows) will be singular.

4.5 Summary of Basic Rules

Here are some basic rules for matrices. For the scalar k and matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of appropriate dimension,

$$
\begin{aligned}
\mathbf{A}+\mathbf{B} &= \mathbf{B}+\mathbf{A} && \text{(commutative property of matrix addition)} \\
(\mathbf{A}+\mathbf{B})+\mathbf{C} &= \mathbf{A}+(\mathbf{B}+\mathbf{C}) && \text{(associative property of matrix addition)} \\
(\mathbf{AB})\mathbf{C} &= [\mathbf{A}(\mathbf{BC}) && \text{(associative property of matrix multiplication)} \\
\mathbf{C}(\mathbf{A}+\mathbf{B}) &= \mathbf{CA}+\mathbf{CB} && \text{(distributive law)} \\
k(\mathbf{A}+\mathbf{B}) &= k\mathbf{A}+k\mathbf{B} && \text{(follows from the distributive law)} \\
(\mathbf{A}^T)^T &= \mathbf{A} && \\
(\mathbf{A}+\mathbf{B})^T &= \mathbf{A}^T+\mathbf{B}^T && \\
(\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T && \\
(\mathbf{ABC})^T &= \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T && \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} && \text{(when the inverses exist)} \\
(\mathbf{ABC})^{-1} &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} && \text{(when the inverses exist)} \\
(\mathbf{A}^{-1})^{-1} &= \mathbf{A} && \text{(when } \mathbf{A}^{-1} \text{ exists)} \\
(\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T && \text{(when } \mathbf{A}^{-1} \text{ exists)}
\end{aligned}
$$

4.6 Eigenvectors and Eigenvalues

Suppose you want to solve

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

for the vector x, where λ is a constant. That is, you want to find all the vectors that are rotationally unaffected by hitting them on the left with $\mathbf{A}$ (but we will allow them to get stretched or shrunk). These vectors are called the *eigenvectors* of the matrix $\mathbf{A}$. The corresponding *eigenvalues* are the scalar λ values that determine the amount of stretching or shrinking through multiplication. Here's how you can find them:

4.7 Examples of Matrices in Statistics

If $X_1, X_2, \ldots, X_n$ are random variables, then

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

is a *random vector*. The *expectation* of the random vector $\mathbf{X}$ is defined

The *covariance matrix* of the random vector $\mathbf{X}$ contains the covariances of all pairs of random variables that are elements of $\mathbf{X}$.

If **x** and **y** are column vectors of data values, the sample means, sample variances, and sample covariances can be written as

Furthermore, if $\mathbf{X}_{n\times m}$ is a matrix of data values, and each column contains observations on a different variable, then

Covariance matrices are symmetric and *positive semidefinite*, meaning

We will soon see that in a very general way we can write the vector of least-squares coefficients as

4.8 Homework

1. Solve the following systems of linear equations

   (i) by hand (using back-substitution or whatever method you like);

   (ii) with R by setting $\mathbf{A}$ to the coefficient matrix, $\mathbf{x}$ to the variable vector, and $\mathbf{b}$ to the vector of constants. Use the fact that $\mathbf{A}\mathbf{x} = \mathbf{b}$ implies $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ if $\mathbf{A}$ is nonsingular. What's going on with case (c)?

   (a)

$$\begin{aligned} 2x_1 + 3x_2 &= 4 \\ x_1 - 4x_2 &= -9 \end{aligned}$$

   (b)

$$\begin{aligned} 2x_1 + 2x_2 + 3x_3 &= 7 \\ x_1 - x_2 + 4x_3 &= -3 \\ -2x_1 + 4x_2 + 2x_3 &= 0 \end{aligned}$$

   (c)

$$\begin{aligned} 2x_1 + 2x_2 + 3x_3 &= 7 \\ x_1 - x_2 + 4x_3 &= -3 \\ 3x_1 + x_2 + 7x_3 &= 4 \end{aligned}$$

2. Given the matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 4 \\ 2 & 6 \\ 3 & -1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 2 & 2 \\ 3 & 2 \\ 1 & 7 \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} 3 & 7 & 1 \\ 4 & 7 & 5 \end{pmatrix},$$

   calculate (by hand and with R) $\mathbf{A} + \mathbf{B}$, $\mathbf{A} - \mathbf{B}$, $\mathbf{AC}$, $\mathbf{CA}$, $\mathbf{AB}^T$, $\mathbf{B}^T\mathbf{A}$.

3. Prove that for any 2×2 matrix $\mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

   that $\mathbf{A}^{-1}$ (if it exists) is given by

$$\mathbf{A}^{-1} = \begin{pmatrix} a_{22}/D & -a_{12}/D \\ -a_{21}/D & a_{11}/D \end{pmatrix},$$

   where $D = a_{11}a_{22} - a_{12}a_{21}$.

4. Given the matrices $\mathbf{X}$ and $\mathbf{Y}$

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 6 \\ 1 & 4 \end{pmatrix}, \qquad \mathbf{Y} = \begin{pmatrix} 6 \\ 9 \\ 23 \\ 25 \\ 17 \end{pmatrix}$$

evaluate (using R) the expression $\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. What is the dimension of β? Also, plot the points $(x_{12}, y_1), (x_{22}, y_2), \ldots, (x_{52}, y_5)$ as well as the line $y = \beta_1 x + \beta_0$. What do you observe? Compare this with the output of lm(y $\sim$ x) where the x-values are 2, 3, 5, 6, 4 and the y-values are 6, 9, 23, 25, 17.

5. Find the eigenvalues and eigenvectors of the following matrices with R. Also, what are the ranks of the matrices $\mathbf{I}$, $\mathbf{A}$, and $\mathbf{B}$?

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 0 \\ 2 & 0 & 5 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

6. Suppose X and Y are correlated normal random variables with mean and covariance matrices

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} = \begin{pmatrix} 20 \\ 10 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} \sigma_X^2 & \sigma_{XY}^2 \\ \sigma_{YX}^2 & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} 9 & 12 \\ 12 & 25 \end{pmatrix}.$$

Use R to simulate 1000 x, y-pairs according to this bivariate distribution and make a scatterplot of the points. Make sure the x- and y- ranges are exactly the same (like 0 to 40, and -10 to 30, respectively). *Hint: use the mvrnorm() function from the MASS package.* Next, find the eigenvectors of the covariance matrix with the eigen() command. Note R gives *unit* eigenvectors (their lengths are 1). Draw these eigenvectors on your scatterplot using arrows(). You might want to multiply them by a factor of 10 each to stretch them out. What do you observe? It turns out that eigenvectors of any real, symmetric matrix are orthogonal.

7. Consider the following data:

x	3.04	4.55	7.47	6.33	5.00	5.20	9.66	11.11	8.34	7.10
y	10.48	1.76	-19.38	.50	-.63	0.08	-16.88	-27.66	-11.95	-9.05
z	21.34	15.14	-.10	19.26	15.60	16.14	7.31	-.51	9.81	10.06

Put the x-, y-, and z-values into column vectors in R, and ultimately into a data matrix with three columns and ten rows. Use matrix multiplication to compute the 3 sample covariance matrix. Make a 3D scatterplot of the three variables. Explain how the covariance matrix describes the point pattern in the 3D scatterplot (try using the rgl and scatterplot3d libraries, along with the plot3d() command). Be sure to include a copy of your plot in your homework solutions.

8. Consider the data in the previous problem. Use the matrix form of the least-squares coefficients to build a linear model for z using x and y as predictor variables. To do this, let the data matrix $\mathbf{X}$ have three columns: the first is a column of ten 1s, the second has all the x-values, and the third column has all the y- values. Carry out the process using matrix multiplication in R:
$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$
then verify your coefficients are correct with lm(z $\sim$ x + y).

9. It turns out that positive semidefinite matrices have the property that they do not swing vectors more than 90 degrees. That is, if $A_{n\times n}$ is positive semidefinite, then for any vector $v_{n\times 1}$, the angle between v and Av is no more than 90 degrees. Prove this. *Hint: recall that $x \cdot y = ||x||\,||y|| \cos\theta$, where θ is the angle between x and y.*

10. Prove that covariance matrices are positive semidefinite.