Appendix A

Matrix Algebra

A.1 Definitions

An $n \times m$ **matrix** is a rectangular array of numbers with n rows and m columns. For example,

$$\begin{pmatrix} 5 & -4 & 0 \\ -1 & 2 & 1 \end{pmatrix}$$

is a 2×3 matrix. The numbers n and m are the **dimensions** of the matrix. A matrix such that the dimensions n and m are equal is called a **square** matrix. A **column vector** is an $n \times 1$ matrix and a **row vector** is a $1 \times m$ matrix.

Two matrices of the same size can be added by the addition of their corresponding elements.

Example A.1 *The sum of two matrices.*

The sum of the 2×3 matrices

$$\begin{pmatrix} 5 & -4 & 0 \\ -1 & 2 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -2 & 4 & 3 \\ 3 & 0 & -5 \end{pmatrix}$$

is

$$\begin{pmatrix} 5 & -4 & 0 \\ -1 & 2 & 1 \end{pmatrix} + \begin{pmatrix} -2 & 4 & 3 \\ 3 & 0 & -5 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 3 \\ 2 & 2 & -4 \end{pmatrix}.$$

Only matrices that have exactly the same dimensions can be added.

The **scalar product** of a constant with a matrix is formed by multiplying every entry in the matrix by the constant.

Example A.2 *The scalar product of a constant and a matrix.*

The scalar product of the constant 3 with the matrix

$$\begin{pmatrix} 3 & 0 & 3 \\ 2 & 2 & -4 \end{pmatrix}$$

is

$$3\begin{pmatrix} 3 & 0 & 3 \\ 2 & 2 & -4 \end{pmatrix} = \begin{pmatrix} 9 & 0 & 9 \\ 6 & 6 & -12 \end{pmatrix}.$$

The **dot product** of the row vector

$$(x_1 \ldots x_n)$$

with the column vector

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

is the number $x_1y_1 + \ldots + x_ny_n$.

The **product** of an $n \times m$ matrix with an $m \times r$ matrix is an $n \times r$ matrix, where the entry in row i and column j is defined to be the dot product of row i in the first matrix with column j in the second matrix.

Example A.3 *The product of two matrices.*

The product of the 2×3 matrix

$$\begin{pmatrix} 3 & 0 & 3 \\ 2 & 2 & -4 \end{pmatrix}$$

with the 3×2 matrix

$$\begin{pmatrix} 1 & -2 \\ 3 & 0 \\ 2 & -1 \end{pmatrix}$$

is the 2×2 matrix

$$\begin{pmatrix} 3 & 0 & 3 \\ 2 & 2 & -4 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 3 & 0 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 9 & -9 \\ 0 & 0 \end{pmatrix}.$$

The entries are obtained as follows:

Row	Column	Entry
1	1	$(3)(1)+(0)(3)+(3)(2)=9$
1	2	$(3)(-2)+(0)(0)+(3)(-1)=-9$
2	1	$(2)(1)+(2)(3)+(-4)(2)=0$
2	2	$(2)(-2)+(2)(0)(-4)(-1)=0$

Unlike ordinary multiplication, matrix multiplication is usually not commutative; for example,

$$\begin{pmatrix} 3 & 0 & 3 \\ 2 & 2 & -4 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 3 & 0 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 9 & -9 \\ 0 & 0 \end{pmatrix},$$

but

$$\begin{pmatrix} 1 & -2 \\ 3 & 0 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} 3 & 0 & 3 \\ 2 & 2 & -4 \end{pmatrix} = \begin{pmatrix} -1 & -4 & 11 \\ 9 & 0 & 9 \\ 4 & -2 & 10 \end{pmatrix}.$$

Matrices must have compatible sizes to be multiplied; the number of columns in the first matrix must be the same as the number of rows in the second matrix. The product has the number of rows of the first matrix and the number of columns of the second matrix.

The **transpose** of a matrix is obtained by interchanging rows and columns.

Example A.4 *The transpose of a matrix.*

The transpose of the matrix

$$A = \begin{pmatrix} 4 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}.$$

is the matrix

$$A^T = \begin{pmatrix} 4 & 2 \\ -1 & 1 \\ 0 & 3 \end{pmatrix}.$$

The **determinant** of an $n \times n$ matrix is a number that can be computed from the matrix. The determinant of a 2×2 matrix is computed by the formula

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

The determinant of a 3×3 matrix is given by

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - afh - bdi - ceg.$$

Example A.5 *The determinant of a* 2×2 *matrix.*

$$\det \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} = (2)(-1) - (1)(3) = -2 - 3 = -5.$$

Example A.6 *The determinant of a* 3×3 *matrix.*

$$\det \begin{pmatrix} 2 & 1 & 0 \\ 2 & 4 & -1 \\ 3 & -1 & 2 \end{pmatrix}$$

$$\begin{aligned} &= 2 \cdot 4 \cdot 2 + (1)(-1)(3) + 0 \cdot 2(-1) - (2)(-1)(-1) - 1 \cdot 2 \cdot 2 - 0 \cdot 4 \cdot 3 \\ &= 16 - 3 + 0 - 2 - 4 + 0 = 7. \end{aligned}$$

Although similar formulas exist for computing the determinants for matrices which are larger than 2×2 or 3×3, they are quite complicated and we will not discuss them here. (The formulas for determinants of 4×4 and 5×5 matrices have 24 and 120 terms, respectively.) It is usually most convenient to use computer software to find the determinant of a large matrix.

A special type of matrix is the identity matrix. An **identity matrix** is a square matrix that contains 1s on its diagonal and 0 everywhere else. If I is an identity matrix and A is an arbitrary matrix, then

$$AI = A = IA$$

as long as the products AI and IA are defined.

Example A.7 *Identity matrices.*

The matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

are identity matrices.

A matrix A^{-1} is defined to be the **inverse** of the $n \times n$ matrix A if

$$AA^{-1} = A^{-1}A = I$$

where I is the $n \times n$ identity matrix. For computing the inverse of a 2×2 matrix, the formula

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

is used. The formula

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}^{-1} = \frac{1}{D} \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix}$$

where

$$D = aei + bfg + cdh - afh - bdi - ceg$$

is used for computing the inverse of a 3×3 matrix.

Example A.8 *The inverse of a 2 × 2 matrix.*

$$\begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}^{-1} = \frac{1}{(2)(1)-(1)(1)} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

Example A.9 *The inverse of a 3 × 3 matrix.*

$$\begin{pmatrix} 3 & 0 & 0 \\ 1 & 1 & 2 \\ -2 & 0 & 1 \end{pmatrix}^{-1} = \frac{1}{D} \begin{pmatrix} 1\cdot 1 - 2\cdot 0 & 0\cdot 0 - 0\cdot 1 & 0\cdot 2 - 0\cdot 1 \\ 2(-2) - 1\cdot 1 & 3\cdot 1 - 0(-2) & 0\cdot 1 - 3\cdot 2 \\ 1\cdot 0 - 1(-2) & 0(-2) - 3\cdot 0 & 3\cdot 1 - 0\cdot 1 \end{pmatrix}$$

where

$$D = 3\cdot 1\cdot 1 + 0\cdot 2\cdot(-2) + 0\cdot 1\cdot 0 - 3\cdot 2\cdot 0 + 0\cdot 1\cdot 1 + 0\cdot 1(-2)$$

so

$$\begin{pmatrix} 3 & 0 & 0 \\ 1 & 1 & 2 \\ -2 & 0 & 1 \end{pmatrix}^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ -5 & 3 & -6 \\ 2 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 1/3 & 0 & 0 \\ -5/3 & 1 & -2 \\ 2/3 & 0 & 1 \end{pmatrix}.$$

As in the case of determinants, we will not compute the inverses of matrices that are larger than 3 × 3. Computer software such as SAS (see Appendix B) can be used to compute the inverses of larger matrices. However, there is a special case where it is easy to compute the inverses of matrices of larger dimension. A matrix such that all entries off the main diagonal are zero is called a **diagonal matrix**. The inverse of a diagonal matrix is given by the following formula:

$$\begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}^{-1} = \begin{pmatrix} 1/a_1 & 0 & \cdots & 0 \\ 0 & 1/a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/a_n \end{pmatrix}.$$

A.2 Eigenvalues and Eigenvectors

In many areas of mathematics and science, one wishes to find all the numbers λ and all vectors v that satisfy the equation

$$Av = \lambda v \tag{A.1}$$

where A is a square matrix. Any number λ satisfying (A.1) is called an **eigenvalue** of A. A vector v satisfying (A.1) is called an **eigenvector** of A. The eigenvalues λ are found by solving the equation

$$\det(A - \lambda I) = 0. \tag{A.2}$$

Then for every λ that is found as a solution of (A.2), the corresponding set of eigenvectors v is found from (A.1).

Example A.10 *Finding eigenvalues and eigenvectors.*

Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}.$$

Solving (A.2) gives us

$$\begin{aligned} \det\left[\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] &= \det\begin{pmatrix} 5-\lambda & 2 \\ 2 & 2-\lambda \end{pmatrix} \\ &= (5-\lambda)(2-\lambda) - (2)(2) \\ &= \lambda^2 - 7\lambda + 6 = 0. \end{aligned}$$

Solving this equation for λ gives $\lambda_1 = 6$ and $\lambda_2 = 1$ for eigenvalues.

Substituting $\lambda_1 = 6$ into (A.1) gives

$$\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 6\begin{pmatrix} x \\ y \end{pmatrix},$$

where

$$v = \begin{pmatrix} x \\ y \end{pmatrix}.$$

This leads to the system of equations

$$\begin{aligned} 5x + 2y &= 6x \\ 2x + 2y &= 6y \end{aligned}$$

$$\begin{aligned} -x+2y &= 0 \\ 2x-4y &= 0, \end{aligned}$$

which implies that $x = 2y$. Thus the eigenvectors corresponding to $\lambda_1 = 6$ are

$$v_1 = \begin{pmatrix} 2y \\ y \end{pmatrix},$$

for arbitrary y, which lie on a line through the origin of slope 1/2.

Substituting $\lambda_2 = 1$ into (A.1) gives

$$\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1 \begin{pmatrix} x \\ y \end{pmatrix},$$

where

$$v = \begin{pmatrix} x \\ y \end{pmatrix}.$$

This leads to the system of equations

$$\begin{aligned} 5x+2y &= x \\ 2x+2y &= y \end{aligned}$$

or

$$\begin{aligned} 4x+2y &= 0 \\ 2x+\ y &= 0, \end{aligned}$$

which implies that $y = -2x$. Therefore the eigenvectors corresponding to $\lambda_2 = 1$ are

$$v_2 = \begin{pmatrix} x \\ -2x \end{pmatrix},$$

which lie on a line through the origin of slope -2.

In Example A.10, the lines with slopes 1/2 and -2 are perpendicular to each other. It can be shown in general that any two eigenvectors that correspond to unique eigenvalues are perpendicular. Computer software that finds eigenvalues and eigenvectors usually gives a set of mutually perpendicular eigenvectors—*one corresponding to each*

...such that $v^T v = 1$ for each eigenvector v. Corresponding to the eigenvalue ... SAS (Statistical Software System) software package gives the eigenvector

$$\begin{pmatrix} 0.894 \\ 0.447 \end{pmatrix}.$$

It gives

$$\begin{pmatrix} -0.447 \\ 0.894 \end{pmatrix}$$

as the eigenvector corresponding to the eigenvalue 1. See Section B.5 for more details.

Example A.11 *Repeated eigenvalues.*

Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 9 & 6 & 3 \\ 6 & 4 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Equation (A.2) becomes

$$\det\left[\begin{pmatrix} 9 & 6 & 3 \\ 6 & 4 & 2 \\ 3 & 2 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right] = \det\begin{pmatrix} 9-\lambda & 6 & 3 \\ 6 & 4-\lambda & 2 \\ 3 & 2 & 1-\lambda \end{pmatrix} = 0,$$

which is

$$(9-\lambda)(4-\lambda)(1-\lambda)+6\cdot2\cdot3+3\cdot6\cdot2-4(9-\lambda)-36(1-\lambda)-9(4-\lambda) = (14-\lambda)\lambda^2 = 0.$$

The solutions to this equation are 14, 0, and 0, which are the eigenvalues.

Substituting $\lambda_1 = 14$ into (A.1) gives the system of equations

$$\begin{aligned} 9x + 6y + 3z &= 14x \\ 6x + 4y + 2z &= 14y \\ 3x + 2y + 1z &= 14z \end{aligned}$$

or

$$\begin{aligned} -5x + 6y + 3z &= 0 \\ 6x - 10y + 2z &= 0 \\ 3x + 2y - 13z &= 0 \end{aligned}$$

which has as its solution $x = 3z$ and $y = 2z$ where z is arbitrary. Thus the eigenvectors corresponding to $\lambda_1 = 14$ are of the form

$$v_1 = \begin{pmatrix} 3z \\ 2z \\ z \end{pmatrix}.$$

When $\lambda = 0$, (A.1) becomes

$$\begin{aligned} 9x + 6y + 3z &= 0 \\ 6x + 4y + 2z &= 0 \\ 3x + 2y + 1z &= 0, \end{aligned}$$

$x = -(2/3)y - (1/3)z$ where y and z are arbitrary, or

$$v = \begin{pmatrix} -(2/3)y - (1/3)z \\ y \\ z \end{pmatrix}.$$

Whereas the set of eigenvectors corresponding to $\lambda_1 = 14$ is a line through the origin, the set of eigenvectors corresponding to $\lambda_2 = \lambda_3 = 0$ is a plane through the origin. This is because 0 is a solution of multiplicity 2 of (A.2). In general, a solution of multiplicity k of (A.2) produces a k-dimensional hyperplane of eigenvectors through the origin.

SAS gives the eigenvector

$$v_1 = \begin{pmatrix} 0.801 \\ 0.535 \\ 0.267 \end{pmatrix}$$

corresponding to $\lambda = 14$ and the eigenvectors

$$v_2 = \begin{pmatrix} -0.365 \\ 0.084 \\ 0.927 \end{pmatrix} \quad \text{and} \quad v_3 = \begin{pmatrix} -0.473 \\ 0.841 \\ -0.262 \end{pmatrix}$$

corresponding to $\lambda_2 = \lambda_3 = 0$. The perpendicular eigenvectors v_2 and v_3 determine a plane. See Appendix B for details of using SAS to compute eigenvalues.

An important application of eigenvectors in pattern recognition is to find the directions of greatest variance of a covariance matrix. The eigenvector corresponding to the largest eigenvalue of the covariance matrix Σ is the direction of greatest variance. The largest eigenvalue is the variance in that direction. The eigenvector corresponding to the second largest eigenvector is the direction of greatest variation perpendicular to the first eigenvector. In general, the kth eigenvector is the direction of greatest variation perpendicular to the first through $(k-1)$st eigenvectors. The eigenvalues of Σ, also called **principal components**, allow us to identify the directions of greatest variation in a class. They can help identify new features that are useful for classification.