8.3 Dominant Eigenvalue and Principal Component Analysis

Dominant Eigenvalue

In Section 8.2 we discussed and illustrated the spectral decomposition of a symmetric matrix. We showed that if A is an $n \times n$ symmetric matrix, then we can express A as a linear combination of matrices of rank one, using the eigenvalues of A and associated eigenvectors as follows. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ a set of associated orthonormal eigenvectors. Then the spectral decomposition of A is given by

$$A = \lambda_1 \mathbf{x}_1 \mathbf{x}_1^T + \lambda_2 \mathbf{x}_2 \mathbf{x}_2^T + \cdots + \lambda_n \mathbf{x}_n \mathbf{x}_n^T. \tag{1}$$

Furthermore, if we label the eigenvalues so that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|,$$

then we can construct approximations to the matrix A, using partial sums of the spectral decomposition. We illustrated such approximations by using matrices of zeros and ones that corresponded to pictures represented by matrices of black and white blocks. As remarked in Section 8.2, the terms using eigenvalues of largest magnitude in the partial sums in (1) contributed a large part of the "information" represented by the matrix A. In this section we investigate two other situations where the largest eigenvalue and its corresponding eigenvector can supply valuable information.

DEFINITION 8.1 If A is a real $n \times n$ matrix with real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then an eigenvalue of largest magnitude is called a **dominant eigenvalue** of A.

EXAMPLE 1 Let

$$A = \begin{bmatrix} 9 & 6 & -14 \\ -2 & 1 & 2 \\ 6 & 6 & -11 \end{bmatrix}.$$

The eigenvalues of A are 3, 1, and -5 (verify). Thus the dominant eigenvalue of A is -5, since $|-5| > 1$ and $|-5| > 3$. ■

Remark Observe that λ_j is a dominant eigenvalue of A, provided that $|\lambda_j| \geq |\lambda_i|$, $i = 1, 2, \ldots, j-1, j+1, \ldots, n$. A matrix can have more than one dominant eigenvalue. For example, the matrix

$$\begin{bmatrix} 4 & 2 & -1 \\ 0 & -2 & 7 \\ 0 & 0 & -4 \end{bmatrix}$$

has both 4 and -4 as dominant eigenvalues.

Let A be a real $n \times n$ matrix with real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, where $|\lambda_1| > |\lambda_i|$, $i = 2, \ldots, n$. Then A has a unique dominant eigenvalue. Furthermore, suppose that A is diagonalizable with associated linearly independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Hence $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a basis for R^n, and every vector $\mathbf{x}$ in R^n is expressible as a linear combination of the vectors in S. Let

$$\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n \quad \text{with } c_1 \neq 0$$

and compute the sequence of vectors $A\mathbf{x}$, $A^2\mathbf{x}$, $A^3\mathbf{x}, \ldots, A^k\mathbf{x}, \ldots$. We obtain the following:

$$\begin{aligned}
A\mathbf{x} &= c_1A\mathbf{x}_1 + c_2A\mathbf{x}_2 + \cdots + c_nA\mathbf{x}_n = c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 + \cdots + c_n\lambda_n\mathbf{x}_n \\
A^2\mathbf{x} &= c_1\lambda_1A\mathbf{x}_1 + c_2\lambda_2A\mathbf{x}_2 + \cdots + c_n\lambda_nA\mathbf{x}_n = c_1\lambda_1^2\mathbf{x}_1 + c_2\lambda_2^2\mathbf{x}_2 + \cdots + c_n\lambda_n^2\mathbf{x}_n \\
A^3\mathbf{x} &= c_1\lambda_1^2A\mathbf{x}_1 + c_2\lambda_2^2A\mathbf{x}_2 + \cdots + c_n\lambda_n^2A\mathbf{x}_n = c_1\lambda_1^3\mathbf{x}_1 + c_2\lambda_2^3\mathbf{x}_2 + \cdots + c_n\lambda_n^3\mathbf{x}_n \\
&\vdots \\
A^k\mathbf{x} &= c_1\lambda_1^k\mathbf{x}_1 + c_2\lambda_2^k\mathbf{x}_2 + \cdots + c_n\lambda_n^k\mathbf{x}_n \\
&\vdots
\end{aligned}$$

We have

$$\begin{aligned}
A^k\mathbf{x} &= \lambda_1^k\left(c_1\mathbf{x}_1 + c_2\frac{\lambda_2^k}{\lambda_1^k}\mathbf{x}_2 + \cdots + c_n\frac{\lambda_n^k}{\lambda_1^k}\mathbf{x}_n\right) \\
&= \lambda_1^k\left(c_1\mathbf{x}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\mathbf{x}_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\mathbf{x}_n\right),
\end{aligned}$$

and since λ_1 is the dominant eigenvalue of A, $\left|\dfrac{\lambda_i}{\lambda_1}\right| < 1$ for $i > 1$. Hence, as $k \to \infty$, it follows that

$$A^k\mathbf{x} \to \lambda_1^k c_1\mathbf{x}_1. \tag{2}$$

Using this result, we can make the following observations: For a real diagonalizable matrix with all real eigenvalues and a unique dominant eigenvalue λ_1, we have

(a) $A^k\mathbf{x}$ approaches the zero vector for any vector $\mathbf{x}$, provided that $|\lambda_1| < 1$.
(b) The sequence of vectors $A^k\mathbf{x}$ does not converge, provided that $|\lambda_1| > 1$.
(c) If $|\lambda_1| = 1$, then the limit of the sequence of vectors $A^k\mathbf{x}$ is an eigenvector associated with λ_1.

In certain **iterative** processes in numerical linear algebra, sequences of vectors of the form $A^k\mathbf{x}$ arise frequently. In order to determine the convergence of the sequence, it is important to determine whether or not the dominant eigenvalue is smaller than 1. Rather than compute the eigenvalues, it is often easier to determine an upper bound on the dominant eigenvalue. We next investigate one such approach.

In the Supplementary Exercises 28–33 in Chapter 5 we defined the **1-norm** of a vector $\mathbf{x}$ in R^n as the sum of the absolute values of its entries; that is,

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|.$$

For an $n \times n$ matrix A, we extend this definition and define the **1-norm** of the matrix A to be the maximum of the 1-norms of its columns. We denote the 1-norm of the matrix A by $\|A\|_1$, and it follows that

$$\|A\|_1 = \max_{j=1,2,\ldots,n} \{\|\text{col}_j(A)\|_1\}.$$

EXAMPLE 2 Let

$$A = \begin{bmatrix} 9 & 6 & -14 \\ -2 & 1 & 2 \\ 6 & 6 & -11 \end{bmatrix}$$

as in Example 1. It follows that $\|A\|_1 = \max\{17, 13, 27\} = 27$. ■

Theorem 8.3 For an $n \times n$ matrix A, the absolute value of the dominant eigenvalue of A is less than or equal to $\|A\|_1$.

Proof

Let $\mathbf{x}$ be any n-vector. Then the product $A\mathbf{x}$ can be expressed as a linear combination of the columns of A in the form

$$A\mathbf{x} = x_1\text{col}_1(A) + x_2\text{col}_2(A) + \cdots + x_n\text{col}_n(A).$$

We proceed by using properties of a norm:

$$\begin{aligned} \|A\mathbf{x}\|_1 &= \|x_1\text{col}_1(A) + x_2\text{col}_2(A) + \cdots + x_n\text{col}_n(A)\|_1 \\ &\quad \text{(Compute the 1-norm of each side.)} \\ &\le \|x_1\text{col}_1(A)\|_1 + \|x_2\text{col}_2(A)\|_1 + \cdots + \|x_n\text{col}_n(A)\|_1 \\ &\quad \text{(Use the triangle inequality of the 1-norm.)} \\ &= |x_1|\|\text{col}_1(A)\|_1 + |x_2|\|\text{col}_2(A)\|_1 + \cdots + |x_n|\|\text{col}_n(A)\|_1 \\ &\quad \text{(Use the scalar multiple of a norm.)} \\ &\le |x_1|\|A\|_1 + |x_2|\|A\|_1 + \cdots + |x_n|\|A\|_1 \\ &\quad \text{(Use } \|\text{col}_j(A)\| \le \|A\|_1\text{.)} \\ &= \|\mathbf{x}\|_1\|A\|_1. \end{aligned}$$

Next, suppose that $\mathbf{x}$ is the eigenvector corresponding to the dominant eigenvalue λ of A and recall that $A\mathbf{x} = \lambda\mathbf{x}$. Then we have

$$\|A\mathbf{x}\|_1 = \|\lambda\mathbf{x}\|_1 = |\lambda|\|\mathbf{x}\|_1 \le \|\mathbf{x}\|_1\|A\|_1,$$

and since $\mathbf{x}$ is an eigenvector, we have $\|\mathbf{x}\|_1 \neq 0$, so

$$|\lambda| \le \|A\|_1.$$

Hence the absolute value of the dominant eigenvalue is "bounded above" by the matrix 1-norm of A. ■

EXAMPLE 3 Let

$$A = \begin{bmatrix} 9 & 6 & -14 \\ -2 & 1 & 2 \\ 6 & 6 & -11 \end{bmatrix}$$

as in Example 1. Then

$$|\text{dominant eigenvalue of } A| \leq \|A\|_1 = \max\{17, 13, 27\} = 27.$$

From Example 1, we know that |dominant eigenvalue of A| = 5. ■

Theorem 8.4 If $\|A\|_1 < 1$, then the sequence of vectors $A^k\mathbf{x}$ approaches the zero vector for any vector $\mathbf{x}$.

Proof

Exercise 9. ■

In Section 8.1 we saw sequences of vectors of the form $T^n\mathbf{x}$ that arise in the analysis of Markov processes. If T is a transition matrix (also called a Markov matrix), then $\|T\|_1 = 1$. Then from Theorem 8.3 we know that the absolute value of the dominant eigenvalue of T is less than or equal to 1. However, we have the following stronger result:

Theorem 8.5 If T is a transition matrix of a Markov process, then the dominant eigenvalue of T is 1.

Proof

Let $\mathbf{x}$ be the n-vector of all ones. Then $T^T\mathbf{x} = \mathbf{x}$ (verify), so 1 is an eigenvalue of T^T. Since a matrix and its transpose have the same eigenvalues, $\lambda = 1$ is also an eigenvalue of T. Now by Theorem 8.3 and the statement in the paragraph preceding this theorem, we conclude that the dominant eigenvalue of the transition matrix T is 1. ■

The preceding results about the dominant eigenvalue were very algebraic in nature. We now turn to a graphical look at the effect of the dominant eigenvalue and an associated eigenvector.

From (2) we see that the sequence of vectors $A\mathbf{x}$, $A^2\mathbf{x}$, $A^3\mathbf{x}, \ldots$ approaches a scalar multiple of an eigenvector associated with the dominant eigenvalue. Geometrically, we can say that the sequence of vectors $A\mathbf{x}$, $A^2\mathbf{x}$, $A^3\mathbf{x}, \ldots$ approaches a line in n-space that is parallel to an eigenvector associated with the dominant eigenvalue. Example 4 illustrates this observation in R^2.

EXAMPLE 4 Let L be a linear transformation from R^2 to R^2 that is represented by the matrix

$$A = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix}$$

with respect to the natural basis for R^2. For a 2-vector $\mathbf{x}$ we compute the terms $A^k\mathbf{x}$, $k = 1, 2, \ldots, 7$. Since we are interested in the direction of this set of vectors in R^2 and for ease in displaying these vectors in the plane R^2, we first scale each of the vectors $A^k\mathbf{x}$ to be a unit vector. Setting

$$\mathbf{x} = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$$

and computing the set of vectors, we get the following:

$\mathbf{x}$	$A\mathbf{x}$	$A^2\mathbf{x}$	$A^3\mathbf{x}$	$A^4\mathbf{x}$	$A^5\mathbf{x}$	$A^6\mathbf{x}$	$A^7\mathbf{x}$
0.2	0.2941	0.0688	−0.7654	−0.9397	−0.8209	−0.7667	−0.7402
0.5	0.9558	0.9976	0.6436	−0.3419	−0.5711	−0.6420	−0.6724

Here, we have shown only four decimal digits. Figure 8.8 shows these vectors in R^2, where $\mathbf{x}$ is labeled with 0 and the vectors $A^k\mathbf{x}$ are labeled with the value of k. ■

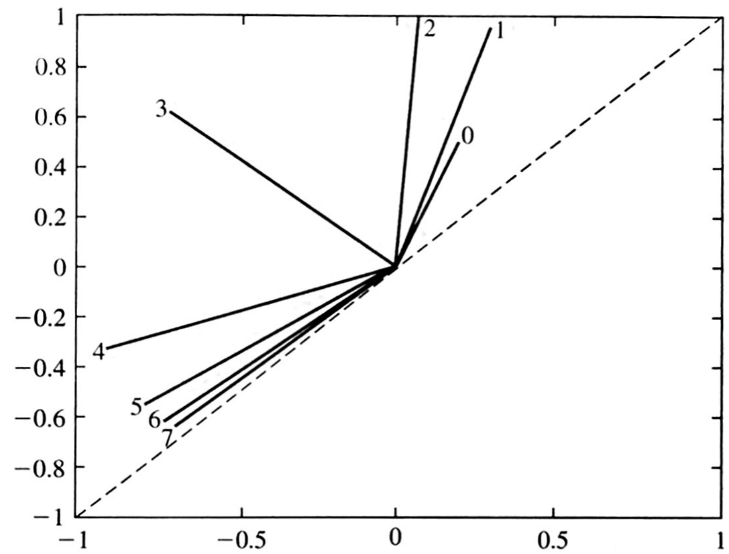

FIGURE 8.8

An eigenvector associated with a dominant eigenvalue is shown as a dashed line segment. The choice of the vector $\mathbf{x}$ is *almost* arbitrary, in the sense that $\mathbf{x}$ cannot be an eigenvector associated with an eigenvalue that is not a dominant eigenvalue, since in that case the sequence $A^k\mathbf{x}$ would always be in the direction of that eigenvector.

EXAMPLE 5

For the linear transformation in Example 4, we compute successive images of the unit circle; that is, $A^k \times$ (unit circle). (See Example 5 of Section 1.7 for a special case.) The first image is an ellipse, and so are the successive images for $k = 2, 3, \ldots$. Figure 8.9 displays five images (where each point displayed in the graphs is the terminal point of a vector that has been scaled to be a unit vector in R^2), and again we see the alignment of the images in the direction of an eigenvector associated with the dominant eigenvalue. This eigenvector is shown as a dashed line segment. ■

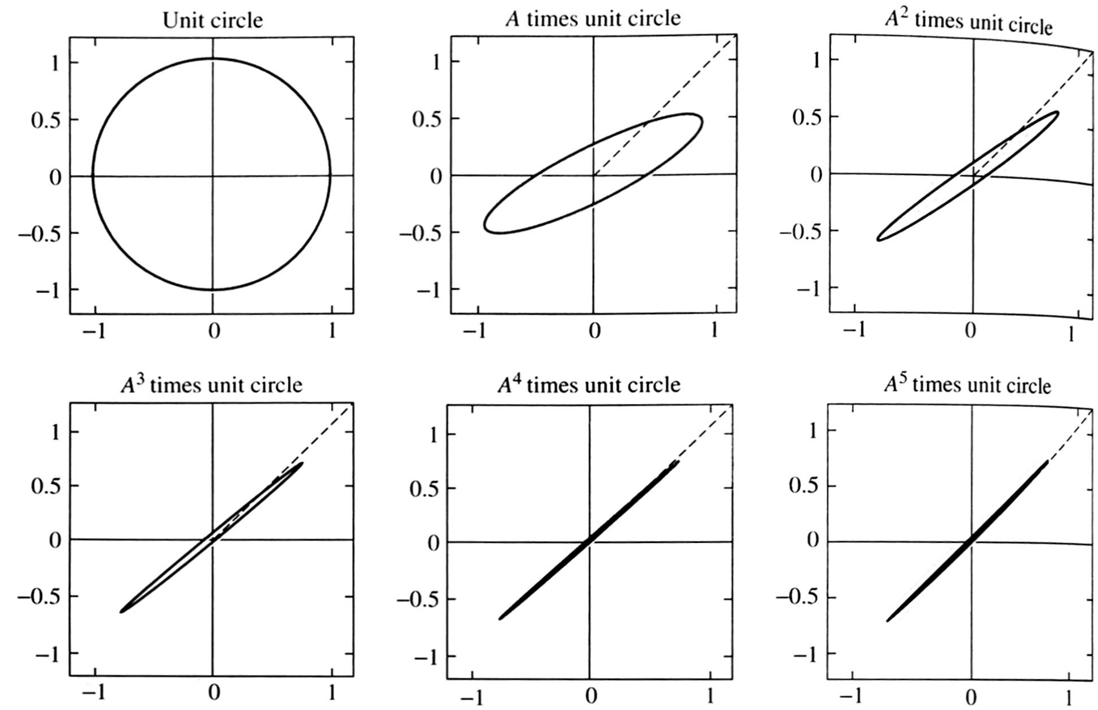

FIGURE 8.9

The sequence of vectors $A\mathbf{x}$, $A^2\mathbf{x}$, $A^3\mathbf{x}$, ... also forms the basis for the numerical method called the **power method** for estimating the dominant eigenvalue of a matrix A. Details of this method can be found in D. R. Hill and B. Kolman, *Modern Matrix Algebra*, Upper Saddle River, NJ: Prentice Hall, 2001, as well as in numerical analysis and numerical linear algebra texts.

■ Principal Component Analysis

The second application that involves the dominant eigenvalue and its eigenvector is taken from applied multivariate statistics and is called **principal component analysis**, often abbreviated **PCA**. To provide a foundation for this topic, we briefly discuss some selected terminology from statistics and state some results that involve a matrix that is useful in statistical analysis.

Multivariate statistics concerns the analysis of data in which several variables are measured on a number of subjects, patients, objects, items, or other entities of interest. The goal of the analysis is to understand the relationships between the variables: how they vary separately, how they vary together, and how to develop an algebraic model that expresses the interrelationships of the variables.

The sets of observations of the variables, the data, are represented by a matrix. Let x_{jk} indicate the particular value of the kth variable that is observed on the jth item. We let n be the number of items being observed and p the number of variables measured. Such data are organized and represented by a rectangular matrix X given by

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix},$$

a **multivariate data matrix**. The matrix X contains all the observations on all of the variables. Each column represents the data for a different variable, and linear combinations of the set of observations are formed by the matrix product $X\mathbf{c}$, where $\mathbf{c}$ is a $p \times 1$ matrix. Useful algebraic models are derived in this way by imposing some optimization criteria for the selection of the entries of the coefficient vector $\mathbf{c}$.

In a single-variable case where the matrix X is $n \times 1$, such as exam scores, the data are often summarized by calculating the arithmetic average, or sample mean, and a measure of spread, or variation. Such summary calculations are referred to as **descriptive statistics**. In this case, for

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

the

$$\textbf{sample mean} = \overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

and the

$$\textbf{sample variance} = s^2 = \frac{1}{n}\sum_{j=1}^{n} (x_j - \overline{x})^2.$$

In addition, the square root of the sample variance is known as the **sample standard deviation**.

EXAMPLE 6 If the matrix

$$X = \begin{bmatrix} 97 & 92 & 90 & 87 & 85 & 83 & 83 & 78 & 72 & 71 & 70 & 65 \end{bmatrix}^T$$

is the set of scores out of 100 for an exam in linear algebra, then the associated descriptive statistics are $\overline{x} \approx 81$, $s^2 \approx 90.4$, and the standard deviation $s \approx 9.5$. ■

These descriptive statistics are also applied to the set of observations of each of the variables in a multivariate data matrix. We next define these, together with statistics that provide a measure of the relationship between pairs of variables:

$$\text{Sample mean for the } k\text{th variable} = \overline{x}_k = \frac{1}{n}\sum_{j=1}^{n} x_{jk}, \quad k = 1, 2, \ldots, p.$$

$$\text{Sample variance for the } k\text{th variable} = s_k^2 = \frac{1}{n}\sum_{j=1}^{n} (x_{jk} - \overline{x}_k)^2, \quad k = 1, 2, \ldots, p.$$

Remark The sample variance is often defined with a divisor of $n-1$ rather than n, for theoretical reasons, especially in the case where n, the number of samples, is small. In many multivariate statistics texts, there is a notational convention employed to distinguish between the two versions. For simplicity in our brief excursion into multivariate statistics, we will use the expression given previously.

Presently, we shall introduce a matrix which contains statistics that relate pairs of variables. For convenience of matrix notation, we shall use the alternative notation s_{kk} for the variance of the kth variable; that is,

$$s_{kk} = s_k^2 = \frac{1}{n}\sum_{j=1}^{n} (x_{jk} - \overline{x}_k)^2, \quad k = 1, 2, \ldots, p.$$

A measure of the linear association between a pair of variables is provided by the notion of **sample covariance**. The measure of association between the ith and kth variables in the multivariate data matrix X is given by

$$\text{Sample covariance} = s_{ik} = \frac{1}{n}\sum_{j=1}^{n} (x_{ji} - \overline{x}_i)(x_{jk} - \overline{x}_k), \quad \begin{array}{l} i = 1, 2, \ldots, p, \\ k = 1, 2, \ldots, p, \end{array}$$

which is the average product of the deviations from their respective sample means. It follows that $s_{ik} = s_{ki}$, for all i and k, and that for $i = k$, the sample covariance is just the variance, $s_k^2 = s_{kk}$.

We next organize the descriptive statistics associated with a multivariate data matrix into matrices:

$$\text{Matrix of sample means} = \overline{\mathbf{x}} = \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_p \end{bmatrix}.$$

$$\text{Matrix of sample variances and covariances} = S_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}.$$

The matrix S_n is a symmetric matrix whose diagonal entries are the sample variances and the subscript n is a notational device to remind us that the divisor n was used to compute the variances and covariances. The matrix S_n is often called the **covariance matrix**, for simplicity.

EXAMPLE 7

A selection of six receipts from a supermarket was collected to investigate the nature of food sales. On each receipt was the cost of the purchases and the number of items purchased. Let the first variable be the cost of the purchases rounded to whole dollars, and the second variable the number of items purchased. The corresponding multivariate data matrix is

$$X = \begin{bmatrix} 39 & 21 \\ 59 & 28 \\ 18 & 10 \\ 21 & 13 \\ 14 & 13 \\ 22 & 10 \end{bmatrix}.$$

Determine the sample statistics given previously, recording numerical values to one decimal place and using this approximation in subsequent calculations.

Solution

We find that the sample means are

$$\overline{x}_1 \approx 28.8 \quad \text{and} \quad \overline{x}_2 \approx 15.8,$$

and thus we take the matrix of sample means as

$$\overline{\mathbf{x}} = \begin{bmatrix} 28.8 \\ 15.8 \end{bmatrix}.$$

The variances are

$$s_{11} \approx 243.1 \quad \text{and} \quad s_{22} \approx 43.1,$$

while the covariances are

$$s_{12} = s_{21} \approx 97.8.$$

Hence we take the covariance matrix as

$$S_n = \begin{bmatrix} 243.1 & 97.8 \\ 97.8 & 43.1 \end{bmatrix}.$$

■

In a more general setting the multivariate data matrix X is a matrix whose entries are random variables. In this setting the matrices of descriptive statistics are computed using probability distributions and expected value. We shall not consider this case, but just note that the vector of means and the covariance matrix can be computed in an analogous fashion. In particular, the covariance matrix is symmetric, as it is for the "sample" case illustrated previously.

We now state several results that indicate how to use information about the covariance matrix to define a set of new variables. These new variables are linear combinations of the original variables represented by the columns of the data matrix X. The technique is called **principal component analysis**, PCA, and is among the oldest and most widely used of multivariate techniques. The new variables are derived in decreasing order of importance so that the first, called the **first principal component**, accounts for as much as possible of the variation in the original data. The second new variable, called the **second principal component**, accounts for another, but smaller, portion of the variation, and so on. For a situation involving p variables, p components are required to account for all the variation, but often, much of the variation can be accounted for by a small number of principal components. Thus, PCA has as its goals the interpretation of the variation and data reduction.

The description of PCA given previously is analogous to the use of the spectral decomposition of a symmetric matrix in the application to symmetric images discussed in Section 8.2. In fact, we use the eigenvalues and associated orthonormal eigenvectors of the covariance matrix S_n to construct the principal components and derive information about them. We have the following result, which we state without proof:

Theorem 8.6 Let S_n be the $p \times p$ covariance matrix associated with the multivariate data matrix X. Let the eigenvalues of S_n be λ_j, $j = 1, 2, \ldots, p$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and let the associated orthonormal eigenvectors be $\mathbf{u}_j$, $j = 1, 2, \ldots, p$. Then the ith principal component $\mathbf{y}_i$ is given by the linear combination of the columns of X, where the coefficients are the entries of the eigenvector $\mathbf{u}_i$; that is,

$$\mathbf{y}_i = i\text{th principal component} = X\mathbf{u}_i.$$

In addition, the variance of $\mathbf{y}_i$ is λ_i, and the covariance of $\mathbf{y}_i$ and $\mathbf{y}_k$, $i \neq k$, is zero. (If some of the eigenvalues are repeated, then the choices of the associated eigenvectors are not unique; hence the principal components are not unique.) ■

Theorem 8.7 Under the hypotheses of Theorem 8.6, the total variance of X given by $\sum_{i=1}^{p} s_{ii}$ is the same as the sum of the eigenvalues of the covariance matrix S_n.

Proof

Exercise 18. ■

This result implies that

$$\begin{pmatrix} \text{Proportion of the} \\ \text{total variance due} \\ \text{to the } k\text{th principal} \\ \text{component} \end{pmatrix} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}, \qquad k = 1, 2, \ldots, p. \tag{3}$$

Thus we see that if $\lambda_1 > \lambda_2$, then λ_1 is the dominant eigenvalue of the covariance matrix. Hence the first principal component is a new variable that "explains," or accounts for, more of the variation than any other principal component. If a large percentage of the total variance for a data matrix with a large number p of columns can be attributed to the first few principal components, then these new variables can replace the original p variables without significant loss of information. Thus we can achieve a significant reduction in data.

EXAMPLE 8 Compute the first principal component for the data matrix X given in Example 7.

Solution

The covariance matrix S_n is computed in Example 7, so we determine its eigenvalues and associated orthonormal eigenvectors. (Here, we record the numerical values to only four decimal places.) We obtain the eigenvalues

$$\lambda_1 = 282.9744 \quad \text{and} \quad \lambda_2 = 3.2256$$

and associated eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 0.9260 \\ 0.3775 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} 0.3775 \\ -0.9260 \end{bmatrix}.$$

Then, using Theorem 8.7, we find that the first principal component is

$$\mathbf{y}_1 = 0.9260\text{col}_1(X) + 0.3775\text{col}_2(X),$$

and it follows that $\mathbf{y}_1$ accounts for the proportion

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} \quad \text{(about 98.9\%)}$$

of the total variance of X (verify). ■

EXAMPLE 9 Suppose that we have a multivariate data matrix X with three columns, which we denote as $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, and the covariance matrix (recording values to only four decimal places) is

$$S_n = \begin{bmatrix} 3.6270 & 2.5440 & 0 \\ 2.5440 & 6.8070 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Determine the principal components $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$.

Solution

We find that the eigenvalues and associated orthonormal eigenvectors are

$$\lambda_1 = 8.2170, \quad \lambda_2 = 2.2170, \quad \text{and} \quad \lambda_3 = 1,$$

$$\mathbf{u}_1 = \begin{bmatrix} 0.4848 \\ 0.8746 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -0.8746 \\ 0.4848 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Thus the principal components are

$$\begin{aligned} \mathbf{y}_1 &= 0.4848\mathbf{x}_1 + 0.8746\mathbf{x}_2 \\ \mathbf{y}_2 &= -0.8746\mathbf{x}_1 + 0.4848\mathbf{x}_2 \\ \mathbf{y}_3 &= \mathbf{x}_3. \end{aligned}$$

Then it follows from (3) that $\mathbf{y}_1$ accounts for 78.61% of the total variance, while $\mathbf{y}_2$ and $\mathbf{y}_3$ account for 19.39% and 8.75%, respectively. ■

There is much more information concerning the influence of and relationship between variables that can be derived from the computations associated with PCA. For more information on PCA, see the references below.

We now make several observations about the geometric nature of PCA. The fact that the covariance matrix S_n is symmetric means that we can find an orthogonal matrix U consisting of eigenvectors of S_n such that $U^T S_n U = D$, a diagonal matrix. The geometric consequence of this result is that the p original variables are rotated to p new orthogonal variables, called the principal components. Moreover, these principal components are linear combinations of the original variables. (The orthogonality follows from the fact that the covariance of $\mathbf{y}_i$ and $\mathbf{y}_k$, $i \neq k$, is zero.) Hence the computation of the principal components amounts to transforming a coordinate system that consists of axes that may not be mutually perpendicular to a new coordinate system with mutually perpendicular axes. The new coordinate system, the principal components, represents the original variables in a more ordered and convenient way. An orthogonal coordinate system makes it possible to easily use projections to derive further information about the relationships between variables. For details, refer to the following references:

REFERENCES

Johnson, Richard A., and Dean W. Wichern. *Applied Multivariate Statisical Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

Jolliffe, I. T. *Principal Component Analysis*. New York: Springer-Verlag, 1986.

Wickens, Thomas D. *The Geometry of Multivariate Statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1995.

For an interesting application to trade routes in geography, see Philip D. Straffin, "Linear Algebra in Geography: Eigenvectors of Networks," *Mathematics Magazine*, vol. 53, no. 5, Nov. 1980, pp. 269–276.

■ Searching with Google: Using the Dominant Eigenvalue

In Section 1.2, after Example 6, we introduced the connectivity matrix A used by the software that drives Google's search engine. Matrix A has entries that are either 0 or 1, with $a_{ij} = 1$ if website j links to website i; otherwise, $a_{ij} = 0$.

EXAMPLE 10

A company with seven employees encourages the use of websites for a variety of business reasons. Each employee has a website, and certain employees include links to coworkers' sites. For this small company, their connectivity matrix is as follows:

$A =$

	E_1	E_2	E_3	E_4	E_5	E_6	E_7
E_1	0	1	0	1	1	0	0
E_2	1	0	1	1	0	1	0
E_3	1	0	0	0	0	0	1
E_4	0	0	1	0	1	1	0
E_5	1	1	0	0	0	1	0
E_6	0	1	1	1	0	0	1
E_7	1	0	0	0	0	0	1

Here, we have assigned the names E_k, $k = 1, 2, \ldots, 7$ to designate the employees. We see from the column labeled E_3 that this employee links to the sites of coworkers E_2, E_4, and E_6. ■

Upon inspecting the connectivity matrix in Example 10, we might try to assign a rank to an employee website by merely counting the number of sites that are linked to it. But this strategy does not take into account the rank of the websites that link to a given site.

There are many applications that use the ranking of objects, teams, or people in associated order of importance. One approach to the ranking strategy is to create a connectivity matrix and compute its dominant eigenvalue and associated eigenvector. For a wide class of such problems, the entries of the eigenvector can be taken to be all positive and scaled so that the sum of their squares is 1. In such cases, if the kth entry is largest, then the kth item that is being ranked is considered the most important; that is, it has the highest rank. The other items are ranked according to the size of the corresponding entry of the eigenvector.

EXAMPLE 11 For the connectivity matrix in Example 10, an eigenvector associated with the dominant eigenvalue is

$$\mathbf{v} = \begin{bmatrix} 0.4261 \\ 0.4746 \\ 0.2137 \\ 0.3596 \\ 0.4416 \\ 0.4214 \\ 0.2137 \end{bmatrix}.$$

It follows that $\max\{v_1, v_2, \ldots, v_7\} = 0.4746$; hence employee number 2 has the highest-ranked website, followed by that of number 5, and then number 1. Notice that the site for employee 6 was referenced more times than that of employee 1 or employee 5, but is considered lower in rank. ■

In carrying out a Google search, the ranking of websites is a salient feature that determines the order of the sites returned to a query. The strategy for ranking uses the basic idea that the rank of a site is higher if other highly ranked sites link to it. In order to implement this strategy for the huge connectivity matrix that is a part of Google's ranking mechanism, a variant of the dominant eigenvalue/eigenvector idea of Example 10 is used. In their algorithm the Google team determines the rank of a site so that it is proportional to the sum of the ranks of all sites that link to it. This approach generates a large eigenvalue/eigenvector problem that uses the connectivity matrix in a more general fashion than that illustrated in Example 11.

REFERENCES

Moler, Cleve. "The World's Largest Matrix Computation: Google's PageRank Is an Eigenvector of a Matrix of Order 2.7 Billion." *MATLAB News and Notes*, October 2002, pp. 12–13.

Wilf, Herbert S. "Searching the Web with Eigenvectors." *The UMAP Journal*, 23(2), 2002, pp. 101–103.

Key Terms

Symmetric matrix
Orthonormal eigenvectors
Dominant eigenvalue
Iterative process
1-norm
Markov process
Power method
Principal component analysis (PCA)
Multivariate data matrix
Descriptive statistics
Sample mean
Sample variance
Sample standard deviation
Covariance matrix
First principal component
Second principal component
Total variance
Google
Connectivity matrix

8.3 Exercises

1. Find the dominant eigenvalue of each of the following matrices:

(a) $\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$ **(b)** $\begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{bmatrix}$

2. Find the dominant eigenvalue of each of the following matrices:

(a) $\begin{bmatrix} 4 & 2 \\ 3 & 3 \end{bmatrix}$ **(b)** $\begin{bmatrix} 2 & 1 & 2 \\ 2 & 2 & -2 \\ 3 & 1 & 1 \end{bmatrix}$

3. Find the 1-norm of each of the following matrices:

(a) $\begin{bmatrix} 3 & -5 \\ 2 & 2 \end{bmatrix}$ **(b)** $\begin{bmatrix} 4 & 1 & 0 \\ 2 & 5 & 0 \\ -4 & -4 & 7 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & -1 & 1 & 0 \\ 4 & -2 & 2 & 3 \\ 1 & 0 & 2 & 6 \\ -3 & 4 & 8 & 1 \end{bmatrix}$

4. Find the 1-norm of each of the following matrices:

(a) $\begin{bmatrix} 4 & -1 \\ 1 & 0 \end{bmatrix}$ **(b)** $\begin{bmatrix} -2 & 0 & 0 \\ -2 & 3 & -1 \\ 3 & -2 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 & -3 & 4 \\ 4 & 1 & 2 & -3 \\ -3 & 4 & 1 & 2 \\ 2 & -3 & 4 & 1 \end{bmatrix}$

5. Determine a bound on the absolute value of the dominant eigenvalue for each of the matrices in Exercise 1.

6. Determine a bound on the absolute value of the dominant eigenvalue for each of the matrices in Exercise 2.

7. Prove that if A is symmetric, then $\|A\|_1 = \|A^T\|_1$.

8. Determine a matrix A for which $\|A\|_1 = \|A^T\|_1$, but A is not symmetric.

9. Prove Theorem 8.4.

10. Explain why $\|A\|_1$ can be greater than 1 and the sequence of vectors $A^k\mathbf{x}$ can still approach the zero vector.

11. Let $X = \begin{bmatrix} 56 & 62 & 59 & 73 & 75 \end{bmatrix}^T$ be the weight in ounces of scoops of birdseed obtained by the same person using the same scoop. Find the sample mean, the variation, and standard deviation of these data.

12. Let $X = \begin{bmatrix} 5400 & 4900 & 6300 & 6700 \end{bmatrix}^T$ be the estimates in dollars for the cost of replacing a roof on the same home. Find the sample mean, the variation, and standard deviation of these data.

13. For the five most populated cities in the United States in 2002, we have the following crime information: For violent offenses known to police per 100,000 residents, the number of robberies appears in column 1 of the data matrix X, and the number of aggravated assaults in column 2. (Values are rounded to the nearest whole number.)

$$X = \begin{bmatrix} 337 & 425 \\ 449 & 847 \\ 631 & 846 \\ 550 & 617 \\ 582 & 647 \end{bmatrix}$$

Determine the vector of sample means and the covariance matrix. (Data taken from TIME Almanac 2006, Information Please LLC, Pearson Education, Boston, MA.)

14. For the five most populated cities in the United States in 2002, we have the following crime information: For property crimes known to police per 100,000 residents, the number of burglaries appears in column 1 of the data matrix X and the number of motor vehicle thefts in column 2. (Values are rounded to the nearest whole number.)

$$X = \begin{bmatrix} 372 & 334 \\ 662 & 891 \\ 869 & 859 \\ 1319 & 1173 \\ 737 & 873 \end{bmatrix}$$

Determine the vector of sample means and the covariance matrix. (Data taken from TIME Almanac 2006, Information Please LLC, Pearson Education, Boston, MA.)

15. For the data in Exercise 13, determine the first principal component.

16. For the data in Exercise 14, determine the first principal component.

17. In Section 5.3 we defined a positive definite matrix as a square symmetric matrix C such that $\mathbf{y}^T C\mathbf{y} > 0$ for every nonzero vector $\mathbf{y}$ in R^n. Prove that any eigenvalue of a positive definite matrix is positive.

18. Let S_n be a covariance matrix satisfying the hypotheses of Theorem 8.6. To prove Theorem 8.7, proceed as follows:
    (a) Show that the trace of S_n is the total variance. (See Section 1.3, Exercise 43, for the definition of trace.)
    (b) Show that there exists an orthogonal matrix P such that $P^T S_n P = D$, a diagonal matrix.
    (c) Show that the trace of S_n is equal to the trace of D.
    (d) Complete the proof.

8.4 Differential Equations

A **differential equation** is an equation that involves an unknown function and its derivatives. An important, simple example of a differential equation is

$$\frac{d}{dt}x(t) = rx(t),$$

where r is a constant. The idea here is to find a function $x(t)$ that will satisfy the given differential equation. This differential equation is discussed further subsequently. Differential equations occur often in all branches of science and engineering; linear algebra is helpful in the formulation and solution of differential equations. In this section we provide only a brief survey of the approach; books on differential equations deal with the subject in much greater detail, and several suggestions for further reading are given at the end of this chapter.

Homogeneous Linear Systems

We consider the **first-order homogeneous linear system** of differential equations,

$$\begin{aligned} x_1'(t) &= a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ x_2'(t) &= a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ &\ \ \vdots \\ x_n'(t) &= a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t), \end{aligned} \tag{1}$$

where the a_{ij} are known constants. We seek functions $x_1(t), x_2(t), \ldots, x_n(t)$ defined and differentiable on the real line and satisfying (1).

We can write (1) in matrix form by letting

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix},$$