

Chapter 11

Dimensionality Reduction

There are many sources of data that can be viewed as a large matrix. We saw in Chapter 5 how the Web can be represented as a transition matrix. In Chapter 9, the utility matrix was a point of focus. And in Chapter 10 we examined matrices that represent social networks. In many of these matrix applications, the matrix can be summarized by finding “narrower” matrices that in some sense are close to the original. These narrow matrices have only a small number of rows or a small number of columns, and therefore can be used much more efficiently than can the original large matrix. The process of finding these narrow matrices is called *dimensionality reduction*.

We saw a preliminary example of dimensionality reduction in Section 9.4. There, we discussed UV-decomposition of a matrix and gave a simple algorithm for finding this decomposition. Recall that a large matrix M was decomposed into two matrices U and V whose product UV was approximately M . The matrix U had a small number of columns whereas V had a small number of rows, so each was significantly smaller than M , and yet together they represented most of the information in M that was useful in predicting ratings of items by individuals.

In this chapter we shall explore the idea of dimensionality reduction in more detail. We begin with a discussion of eigenvalues and their use in “principal component analysis” (PCA). We cover singular-value decomposition, a more powerful version of UV-decomposition. Finally, because we are always interested in the largest data sizes we can handle, we look at another form of decomposition, called CUR-decomposition, which is a variant of singular-value decomposition that keeps the matrices of the decomposition sparse if the original matrix is sparse.

11.1 Eigenvalues and Eigenvectors of Symmetric Matrices

We shall assume that you are familiar with the basics of matrix algebra: multiplication, transpose, determinants, and solving linear equations for example. In this section, we shall define eigenvalues and eigenvectors of a symmetric matrix and show how to find them. Recall a matrix is symmetric if the element in row i and column j equals the element in row j and column i .

11.1.1 Definitions

Let M be a square matrix. Let λ be a constant and \mathbf{e} a nonzero column vector with the same number of rows as M . Then λ is an *eigenvalue* of M and \mathbf{e} is the corresponding *eigenvector* of M if $M\mathbf{e} = \lambda\mathbf{e}$.

If \mathbf{e} is an eigenvector of M and c is any constant, then it is also true that $c\mathbf{e}$ is an eigenvector of M with the same eigenvalue. Multiplying a vector by a constant changes the length of a vector, but not its direction. Thus, to avoid ambiguity regarding the length, we shall require that every eigenvector be a *unit vector*, meaning that the sum of the squares of the components of the vector is 1. Even that is not quite enough to make the eigenvector unique, since we may still multiply by -1 without changing the sum of squares of the components. Thus, we shall normally require that the first nonzero component of an eigenvector be positive.

Example 11.1: Let M be the matrix

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

One of the eigenvectors of M is

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

and its corresponding eigenvalue is 7. The equation

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = 7 \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

demonstrates the truth of this claim. Note that both sides are equal to

$$\begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix}$$

Also observe that the eigenvector is a unit vector, because $(1/\sqrt{5})^2 + (2/\sqrt{5})^2 = 1/5 + 4/5 = 1$. \square

11.1.2 Computing Eigenvalues and Eigenvectors

We have already seen one approach to finding an *eigenpair* (an eigenvalue and its corresponding eigenvector) for a suitable matrix M in Section 5.1: start with any unit vector \mathbf{v} of the appropriate length and compute $M^i\mathbf{v}$ iteratively until it converges.¹ When M is a stochastic matrix, the limiting vector is the *principal* eigenvector (the eigenvector with the largest eigenvalue), and its corresponding eigenvalue is 1.² This method for finding the principal eigenvector, called *power iteration*, works quite generally, although if the principal eigenvalue (eigenvalue associated with the principal eigenvector) is not 1, then as i grows, the ratio of $M^{i+1}\mathbf{v}$ to $M^i\mathbf{v}$ approaches the principal eigenvalue while $M^i\mathbf{v}$ approaches a vector (probably not a unit vector) with the same direction as the principal eigenvector.

We shall take up the generalization of the power-iteration method to find all eigenpairs in Section 11.1.3. However, there is an $O(n^3)$ -running-time method for computing all the eigenpairs of a symmetric $n \times n$ matrix exactly, and this method will be presented first. There will always be n eigenpairs, although in some cases, some of the eigenvalues will be identical. The method starts by restating the equation that defines eigenpairs, $M\mathbf{e} = \lambda\mathbf{e}$ as $(M - \lambda I)\mathbf{e} = \mathbf{0}$, where

1. I is the $n \times n$ *identity matrix* with 1's along the main diagonal and 0's elsewhere.
2. $\mathbf{0}$ is a vector of all 0's.

A fact of linear algebra is that in order for $(M - \lambda I)\mathbf{e} = \mathbf{0}$ to hold for a vector $\mathbf{e} \neq \mathbf{0}$, the determinant of $M - \lambda I$ must be 0. Notice that $(M - \lambda I)$ looks almost like the matrix M , but if M has c in one of its diagonal elements, then $(M - \lambda I)$ has $c - \lambda$ there. While the determinant of an $n \times n$ matrix has $n!$ terms, it can be computed in various ways in $O(n^3)$ time; an example is the method of “pivotal condensation.”

The determinant of $(M - \lambda I)$ is an n th-degree polynomial in λ , from which we can get the n values of λ that are the eigenvalues of M . For any such value, say c , we can then solve the equation $M\mathbf{e} = c\mathbf{e}$. There are n equations in n unknowns (the n components of \mathbf{e}), but since there is no constant term in any equation, we can only solve for \mathbf{e} to within a constant factor. However, using any solution, we can normalize it so the sum of the squares of the components is 1, thus obtaining the eigenvector that corresponds to eigenvalue c .

Example 11.2: Let us find the eigenpairs for the 2×2 matrix M from Example 11.1. Recall $M =$

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

¹Recall M^i denotes multiplying by the matrix M i times, as discussed in Section 5.1.2.

²Note that a stochastic matrix is not generally symmetric. Symmetric matrices and stochastic matrices are two classes of matrices for which eigenpairs exist and can be exploited. In this chapter, we focus on techniques for symmetric matrices.

Then $M - \lambda I$ is

$$\begin{bmatrix} 3 - \lambda & 2 \\ 2 & 6 - \lambda \end{bmatrix}$$

The determinant of this matrix is $(3 - \lambda)(6 - \lambda) - 4$, which we must set to 0. The equation in λ to solve is thus $\lambda^2 - 9\lambda + 14 = 0$. The roots of this equation are $\lambda = 7$ and $\lambda = 2$; the first is the principal eigenvalue, since it is the larger. Let \mathbf{e} be the vector of unknowns

$$\begin{bmatrix} x \\ y \end{bmatrix}$$

We must solve

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 7 \begin{bmatrix} x \\ y \end{bmatrix}$$

When we multiply the matrix and vector we get two equations

$$\begin{aligned} 3x + 2y &= 7x \\ 2x + 6y &= 7y \end{aligned}$$

Notice that both of these equations really say the same thing: $y = 2x$. Thus, a possible eigenvector is

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

But that vector is not a unit vector, since the sum of the squares of its components is 5, not 1. Thus to get the unit vector in the same direction, we divide each component by $\sqrt{5}$. That is, the principal eigenvector is

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

and its eigenvalue is 7. Note that this was the eigenpair we explored in Example 11.1.

For the second eigenpair, we repeat the above with eigenvalue 2 in place of 7. The equation involving the components of \mathbf{e} is $x = -2y$, and the second eigenvector is

$$\begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$$

Its corresponding eigenvalue is 2, of course. \square

11.1.3 Finding Eigenpairs by Power Iteration

We now examine the generalization of the process we used in Section 5.1 to find the principal eigenvector, which in that section was the PageRank vector – all we needed from among the various eigenvectors of the stochastic matrix of the Web. We start by computing the principal eigenvector by a slight generalization of the approach used in Section 5.1. We then modify the matrix to, in

effect, remove the principal eigenvector. The result is a new matrix whose principal eigenvector is the second eigenvector (eigenvector with the second-largest eigenvalue) of the original matrix. The process proceeds in that manner, removing each eigenvector as we find it, and then using power iteration to find the principal eigenvector of the matrix that remains.

Let M be the matrix whose eigenpairs we would like to find. Start with any nonzero vector \mathbf{x}_0 and then iterate:

$$\mathbf{x}_{k+1} := \frac{M\mathbf{x}_k}{\|M\mathbf{x}_k\|}$$

where $\|N\|$ for a matrix or vector N denotes the *Frobenius norm*; that is, the square root of the sum of the squares of the elements of N . We multiply the current vector \mathbf{x}_k by the matrix M until convergence (i.e., $\|x_k - x_{k+1}\|$ is less than some small, chosen constant). Let \mathbf{x} be \mathbf{x}_k for that value of k at which convergence is obtained. Then \mathbf{x} is (approximately) the principal eigenvector of M . To obtain the corresponding eigenvalue we simply compute $\lambda_1 = \mathbf{x}^T M \mathbf{x}$, which is the equation $M\mathbf{x} = \lambda\mathbf{x}$ solved for λ , since \mathbf{x} is a unit vector.

Example 11.3: Take the matrix from Example 11.2:

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

and let us start with \mathbf{x}_0 a vector with 1 for both components. To compute \mathbf{x}_1 , we multiply $M\mathbf{x}_0$ to get

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

The Frobenius norm of the result is $\sqrt{5^2 + 8^2} = \sqrt{89} = 9.434$. We obtain \mathbf{x}_1 by dividing 5 and 8 by 9.434; that is:

$$\mathbf{x}_1 = \begin{bmatrix} 0.530 \\ 0.848 \end{bmatrix}$$

For the next iteration, we compute

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.530 \\ 0.848 \end{bmatrix} = \begin{bmatrix} 3.286 \\ 6.148 \end{bmatrix}$$

The Frobenius norm of the result is 6.971, so we divide to obtain

$$\mathbf{x}_2 = \begin{bmatrix} 0.471 \\ 0.882 \end{bmatrix}$$

We are converging toward a normal vector whose second component is twice the first. That is, the limiting value of the vector that we obtain by power iteration is the principal eigenvector:

$$\mathbf{x} = \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix}$$

Finally, we compute the principal eigenvalue by

$$\lambda = \mathbf{x}^T M \mathbf{x} = \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} = 6.993$$

Recall from Example 11.2 that the true principal eigenvalue is 7. Power iteration will introduce small errors due either to limited precision, as was the case here, or due to the fact that we stop the iteration before reaching the exact value of the eigenvector. When we computed PageRank, the small inaccuracies did not matter, but when we try to compute all eigenpairs, inaccuracies accumulate if we are not careful. \square

To find the second eigenpair we create a new matrix $M^* = M - \lambda_1 \mathbf{x} \mathbf{x}^T$. Then, use power iteration on M^* to compute its largest eigenvalue. The obtained \mathbf{x}^* and λ^* correspond to the second largest eigenvalue and the corresponding eigenvector of matrix M .

Intuitively, what we have done is eliminate the influence of a given eigenvector by setting its associated eigenvalue to zero. The formal justification is the following two observations. If $M^* = M - \lambda \mathbf{x} \mathbf{x}^T$, where \mathbf{x} and λ are the eigenpair with the largest eigenvalue, then:

1. \mathbf{x} is also an eigenvector of M^* , and its corresponding eigenvalue is 0. In proof, observe that

$$M^* \mathbf{x} = (M - \lambda \mathbf{x} \mathbf{x}^T) \mathbf{x} = M \mathbf{x} - \lambda \mathbf{x} \mathbf{x}^T \mathbf{x} = M \mathbf{x} - \lambda \mathbf{x} = 0$$

At the next-to-last step we use the fact that $\mathbf{x}^T \mathbf{x} = 1$ because \mathbf{x} is a unit vector.

2. Conversely, if \mathbf{v} and λ_v are an eigenpair of a symmetric matrix M other than the first eigenpair (\mathbf{x}, λ) , then they are also an eigenpair of M^* . *Proof:*

$$M^* \mathbf{v} = (M^*)^T \mathbf{v} = (M - \lambda \mathbf{x} \mathbf{x}^T)^T \mathbf{v} = M^T \mathbf{v} - \lambda \mathbf{x} (\mathbf{x}^T \mathbf{v}) = M^T \mathbf{v} = \lambda_v \mathbf{v}$$

This sequence of equalities needs the following justifications:

- (a) If M is symmetric, then $M = M^T$.
- (b) The eigenvectors of a symmetric matrix are *orthogonal*. That is, the dot product of any two distinct eigenvectors of a matrix is 0. We do not prove this statement here.

Example 11.4: Continuing Example 11.3, we compute

$$M^* = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - 6.993 \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} = \\ \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - \begin{bmatrix} 1.397 & 2.795 \\ 2.795 & 5.589 \end{bmatrix} = \begin{bmatrix} 1.603 & -0.795 \\ -0.795 & 0.411 \end{bmatrix}$$

We may find the second eigenpair by processing the matrix above as we did the original matrix M . \square