Linear Independence, Span, Basis, and Eigendecomposition

QEA

Spring 2016

What is this about?

In this set of exercises, you will learn about some of the fundamental concepts of linear algebra. Some of them can be rather abstract, but we promise that they will come in handy in the very near future, namely when you work on your facial recognition project. There is also a reading and comprehension activity on the PageRank algorithm, which utilizes all of the concepts we have been considering to date, as well as introducing a few more.

Definition and Notation

We begin with a set of definitions and a few remarks connecting these new ideas to ones we have previously considered.

1. A finite set $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, x_m\}$ of vectors in $\mathbf{R}^n$ is said to be *linearly dependent* if there exist scalars $c_1, c_2, \ldots, c_m$, not all zero, such that
$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_m\mathbf{x}_m = \mathbf{0}.$$
Note that $\mathbf{R}^n$ here simply refers to the set of all n dimensional vectors that are made up of real numbers. $\mathbf{R}^n$ is an example of a *vector space* - we will meet different examples of vector spaces in the future. We can also express this equation using a matrix $\mathbf{A}$, whose columns are $\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_m$.
$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_m \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = \mathbf{0}\,. \tag{1}$$
If a non-zero solution exists to $\mathbf{Ac} = \mathbf{0}$ then the set S is linearly dependent. In the case of a square matrix ($n = m$), the vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_m$ are linearly dependent if and only if the $det(\mathbf{A}) = 0$. Otherwise, the only way to satisfy the equation above is if $c_1 = c_2 = \cdots = c_m = 0$. Figure 1 illustrates two examples of three vectors that are in 3D space, but are linearly dependent, since in each case, all three vectors are on a plane.

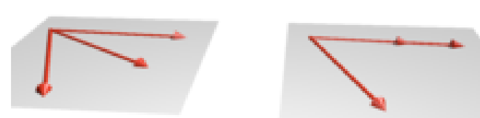

Figure 1: Linearly dependent vectors in $\mathbf{R}^3$. (from Wikimedia Commons).

2. The set S is *linearly independent* if it is not linearly dependent. In other words, the set of vectors S is linearly independent if
$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_m\mathbf{x}_m = \mathbf{0} \tag{2}$$

only when $c_1 = c_2 = \cdots = 0$. In otherwords, if $\mathbf{Ac} = \mathbf{0}$ has only the zero solution then the set if linearly independent. For a square matrix this means the set if linearly independent if and only if $\det(\mathbf{A}) \neq 0$.

3. The *span* of S is the set of all linear combinations of its vectors. In other words, the span of the set S is the set of all possible vectors of the form
$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_m\mathbf{x}_m$$
The *span* is usually denoted by $span(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$.

4. A finite set $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ of vectors is said to form a basis of a vector space V, if the vectors in S are linearly independent, and every point in V can be expressed as a linear combination of the vectors in the set S. Hence, if a set of vectors S is linearly independent those vectors form a basis of the set which is the span of those vectors.

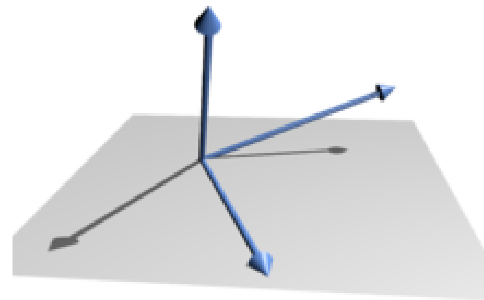

Figure 2: Linearly independent vectors in $\mathbf{R}^3$. (from Wikimedia Commons).

Orthogonality

Two vectors are generally said to be orthogonal if the projection of one vector onto the direction of the other is zero. In our usual spatial coordinate systems, orthogonality is very simple to visualize, but it is much less so when we are dealing with data space which can have a large number of dimensions!

Many of you may have previously encountered the idea of projection and its connection to the dot product, but we'll go through it as a reminder.

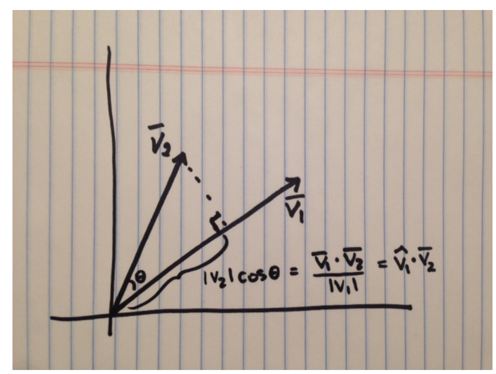

Figure 3: Projection

By simple trigonometry, if we have two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ which have an angle of θ between them, the component of $\mathbf{v}_2$ which lies along the direction of $\mathbf{v}_1$ is $|\mathbf{v}_2|\cos\theta$. Since the dot product of the two vectors can be expressed as $|\mathbf{v}_1||\mathbf{v}_2|\cos\theta$, this component (referred to as the projection) can be written as $\mathbf{v}_1 \cdot \mathbf{v}_2/|\mathbf{v}_1|$. Or, since $\mathbf{v}_1/|\mathbf{v}_1|$ is the unit vector $\hat{\mathbf{v}}_1$ in the direction of vector $\mathbf{v}_1$, we can express the projection of vector $\mathbf{v}_2$ onto the direction of $\mathbf{v}_1$ as $\hat{\mathbf{v}}_1 \cdot \mathbf{v}_2$. If the projection is zero, the vectors are orthogonal, and $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. If the vectors are unit length, in addition to being normal, the vectors are said to be *o*rthonormal. Additionally, if a basis set is made up of orthonormal vectors, it is known as an orthonormal basis.

Decomposition

Suppose we have a set of m basis vectors $\{\mathbf{v}_i\}$ which are normalized ($|\mathbf{v}_i| = 1$), mutually orthogonal ($\mathbf{v}_i \cdot \mathbf{v}_j = 0$ unless $i = j$) and span our space (every point can be written as some linear combination of the vectors $\{\mathbf{v}_i\}$). How do we actually find the linear combination which is equal to a given vector in our space?

Let's say we have a vector $\mathbf{w}$ which we are interested in expressing as a linear combination of our set of orthonormal vectors $\{\mathbf{v}_i\}$. We can write this linear combination as

$$\mathbf{w} = \sum_{i=1}^{m} c_i \mathbf{v}_i \tag{3}$$

and our problem is now to find the coefficients c_i in this expression.

The obvious option is to pack the basis vectors $\mathbf{v}_i$ into the columns of a matrix $\mathbf{A}$, and find solutions of

$$\mathbf{A}\mathbf{c} = \mathbf{w}$$

Since the columns of $\mathbf{A}$ are formed from basis vectors they are linearly independent and a non-zero solution exists and can be determined by the usual methods.

However, our basis vectors form an orthogonal set which permits a more direct calculation. Consider the dot product of a particular vector $\mathbf{v}_k$ in our basis set with our target vector $\mathbf{w}$.

$$\mathbf{v}_k \cdot \mathbf{w} = \mathbf{v}_k \cdot \sum_{i=1}^{m} c_i \mathbf{v}_i \tag{4}$$

Distributing the dot product into the summation we have:

$$\mathbf{v}_k \cdot \mathbf{w} = \sum_{i=1}^{m} c_i \mathbf{v}_k \cdot \mathbf{v}_i \tag{5}$$

But from orthogonality we know that the dot product of any two different vectors in our orthonormal set is zero, so all terms in the sum where $k \neq i$ are zero.

$$\mathbf{v}_k \cdot \mathbf{w} = c_k \mathbf{v}_k \cdot \mathbf{v}_k \tag{6}$$

In addition, since our set of vectors is normalized, we know that $\mathbf{v}_k \cdot \mathbf{v}_k = 1$, leaving us with

$$\mathbf{v}_k \cdot \mathbf{w} = c_k \tag{7}$$

This gives us a very nice, simple way of decomposing a vector into a linear combination of the vectors within our basis set. The dot product of each basis vector with our target vector will result in the coefficient of that term in the linear decomposition.

Exercises

1. Determine if each of the following sets of vectors is linearly independent.

(a) $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$

(d) $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$, where the vectors are all 3 dimensional.

(e) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

2. Which of the following pairs of vectors are orthogonal or orthonormal?

(a) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

(c) $\begin{bmatrix} \frac{2}{\sqrt{13}} \\ \frac{-3}{\sqrt{13}} \end{bmatrix}, \begin{bmatrix} \frac{-2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{bmatrix}$

3. Suppose that you wish to write the vector $\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ as a linear combination of the vectors
$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.
Please write a matrix equation to find the coefficients of the linear combination, and solve for the coefficients using MATLAB if possible.

4. In words, describe the span of the vectors $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ which are all in 3-dimensional Euclidean space. You should use the convention that a vector from the origin to the point (x, y, z) is written as $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

5. In words, describe the span of the vectors $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\-1\end{bmatrix}$.

6. There are many (in general an infinite number) of bases for a given set V. Hence, we can describe elements in the set V as linear combinations of vectors from different bases. Consider the following two basis sets which form bases for 2-dimensional space.

   - $\begin{bmatrix}1\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\end{bmatrix}$ and
   - $\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}$, $\begin{bmatrix}-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}$

   Express the vector $\begin{bmatrix}2\\3\end{bmatrix}$ as a linear combination of the first basis set. Repeat for the second. Writing the coefficients of these linear combinations as vectors, describe the relationship between the two. How does this relate to the basis vectors. Generally, we can write vectors as linear combinations of elements in different basis sets. However, certain basis sets may be more useful than others, as we shall see shortly.

Eigendecomposition

The eigendecomposition is an operation on matrices that is useful in several applications. It can be used to find inverses and powers of matrices, as well as to derive some important results in data analysis. For instance, in a prior in-class exercise, you saw that the eigenvector corresponding to the largest eigenvalue of a covariance matrix was in the direction of greatest variance in your data set. One of the ways one can prove this property is by using the eigenvalue decomposition.

Consider a square matrix $\mathbf{A}$ which has linearly independent eigenvectors. This matrix can be written in the following form

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \tag{8}$$

where the matrix $\mathbf{Q}$ has the i-th eigenvector of $\mathbf{A}$ as its i-th column and $\mathbf{\Lambda}$ is a diagonal matrix with the i-th eigenvalue of $\mathbf{A}$ as its ii-th entry. This is known as the *eigendecomposition* of $\mathbf{A}$.

7. Verify that the expression for the eigendecomposition holds by appealing to the eigenvector equation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$.

8. Consider the following 2×2 matrix $\mathbf{A}$. Compute its eigenvectors and eigenvalues by hand, and then apply it to (8), i.e. determine

the matrices $\mathbf{Q}$, $\mathbf{\Lambda}$, and $\mathbf{Q}^{-1}$.

$$\mathbf{A} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{bmatrix}$$

9. One thing that the eigendecomposition helps us compute is how to raise $\mathbf{A}$ to an integer power. Using the eigendecomposition, prove the following

(a)

$$\mathbf{A}^2 = \mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^{-1} \tag{9}$$

(b)

$$\mathbf{A}^n = \mathbf{Q}\mathbf{\Lambda}^n\mathbf{Q}^{-1} \tag{10}$$

(c)

$$\mathbf{A}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^{-1} \tag{11}$$

Note that for any diagonal matrix $\mathbf{D}$, $\mathbf{D}^k$ is another diagonal matrix whose ii-th entry equals the ii-th entry of $\mathbf{D}$ raised to the k-th power. Hence computing $\mathbf{\Lambda}^n$ is not computationally difficult - you just raise each diagonal entry to the n-th power.

Eigenvalues and eigenvectors of Covariance Matrices

If $\mathbf{A}$ is a covariance matrix, then $\mathbf{A}$ is known to have non-negative eigenvalues, and orthogonal eigenvectors. These properties can be proved as follows. First let $\mathbf{A} = \mathbf{X}^T\mathbf{X}$, where $\mathbf{X}$ is our centralized data matrix (i.e. data matrix with the mean subtracted out). Let $\mathbf{v}_i$ and λ_i be the i-th eigenvector and eigenvalue of $\mathbf{A}$ respectively.

$$\mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i$$

Now we multiply both sides by $\mathbf{v}_i^T$

$$\mathbf{v}_i^T\mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i^T\mathbf{v}_i$$

and use the properties of the transpose

$$(\mathbf{X}\mathbf{v}_i)^T\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i^T\mathbf{v}_i$$

and finally divide both sides by the scalar $\mathbf{v}_i^T\mathbf{v}_i$

$$\frac{(\mathbf{X}\mathbf{v}_i)^T\mathbf{X}\mathbf{v}_i}{\mathbf{v}_i^T\mathbf{v}_i} = \lambda_i$$

Both the numerator and denominator are the result of multiplying the transpose of a column vector with itself. Hence the numerator and denominator are non-negative, which yields the result that the eigenvalues are non-negative, i.e.

$$\lambda_i \geq 0$$

Next we shall show that the eigenvectors are orthogonal. Consider the case that $i \neq j$ and that the matrix $\mathbf{A}$ has distinct eigenvalues

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

Now we multiply both sides by $\mathbf{v}_j^T$

$$\mathbf{v}_j^T \mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_j^T \mathbf{v}_i$$

and use the properties of the transpose

$$(\mathbf{A}^T \mathbf{v}_j)^T \mathbf{v}_i = \lambda_i \mathbf{v}_j^T \mathbf{v}_i$$

Since $\mathbf{A}$ is symmetric, we know that $\mathbf{A}^T = \mathbf{A}$ so that

$$(\mathbf{A}\mathbf{v}_j)^T \mathbf{v}_i = \lambda_i \mathbf{v}_j^T \mathbf{v}_i$$

Recalling that $\mathbf{v}_j$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ_j

$$(\lambda_j \mathbf{v}_j)^T \mathbf{v}_i = \lambda_i \mathbf{v}_j^T \mathbf{v}_i$$

and using properties of the transpose

$$\lambda_j \mathbf{v}_j^T \mathbf{v}_i = \lambda_i \mathbf{v}_j^T \mathbf{v}_i$$

Bringing everything to one side

$$\lambda_j \mathbf{v}_j^T \mathbf{v}_i - \lambda_i \mathbf{v}_j^T \mathbf{v}_i = 0$$

and factoring gives

$$(\lambda_j - \lambda_i)\mathbf{v}_j^T \mathbf{v}_i = 0.$$

Hence, if the eigenvalues are distinct, then it must be the case that $\mathbf{v}_j$ and $\mathbf{v}_i$ are orthogonal for $i \neq j$. Moreover, the eigenvectors can be made orthonormal by scaling each one.

Google PageRank Algorithm

On the website you will find a copy of a review paper on Google's PageRank algorithm. It is written for students like you, and has exercises sprinkled throughout. It connects many of the ideas we have been discussing, especially those involving eigenvalues and eigenvectors. Please start reading it! Make a first pass through the paper, glossing over the details, paying attention to some of the key ideas. Then go back and start reading it more carefully. We will be discussing this algorithm in a future class.