Independence and the Zeros Vector

Simply put, any vector set that includes the zeros vector is automatically a linearly dependent set. Here’s why: any scalar multiple of the zeros vector is still the zeros vector, so the definition of linear dependence is always satisfied. You can see this in the following equation:

lambda0 * 0 = 0*v1 + 0*v2 + ... + 0*vn

As long as lambda0 ≠ 0, we have a nontrivial solution, and the set fits with the definition of linear dependence.

What About Nonlinear Independence?

“But Mike,” I imagine you protesting, “isn’t life, the universe, and everything nonlinear?” I suppose it would be an interesting exercise to count the total number of linear versus nonlinear interactions in the universe and see which sum is larger. But linear algebra is all about, well, linear operations. If you can express one vector as a nonlinear (but not linear) combination of other vectors, then those vectors still form a linearly independent set. The reason for the linearity constraint is that we want to express transformations as matrix multiplication, which is a linear operation. That’s not to throw shade on nonlinear operations—in my imaginary conversation, you have eloquently articulated that a purely linear universe would be rather dull and predictable. But we don’t need to explain the entire universe using linear algebra; we need linear algebra only for the linear parts. (It’s also worth mentioning that many nonlinear systems can be well approximated using linear functions.)

Subspace and Span

When I introduced linear weighted combinations, I gave examples with specific numerical values for the weights (e.g., lambda1 = 1, lambda3 = -3). A subspace is the same idea but using the infinity of possible ways to linearly combine the vectors in the set. That is, for some (finite) set of vectors, the infinite number of ways to linearly combine them—using the same vectors but different numerical values for the weights—creates a vector subspace. And the mechanism of combining all possible linear weighted combinations is called the span of the vector set. Let’s work through a few examples. We’ll start with a simple example of a vector set containing one vector:

V =
1
3

The span of this vector set is the infinity of vectors that can be created as linear combinations of the vectors in the set. For a set with one vector, that simply means all possible scaled versions of that vector. Figure 3-1 shows the vector and the subspace it spans. Consider that any vector in the gray dashed line can be formed as some scaled version of the vector.

Figure 3-1: A vector (black) and the subspace it spans (gray)

Our next example is a set of two vectors in R^3:

V =
1 0 2
-1 1 2

The vectors are in R^3, so they are graphically represented in a 3D axis. But the subspace that they span is a 2D plane in that 3D space (Figure 3-2). That plane passes through the origin, because scaling both vectors by zero gives the zeros vector.

Figure 3-2: Two vectors (black) and the subspace they span (gray)

The first example had one vector and its span was a 1D subspace, and the second example had two vectors and their span was a 2D subspace. There seems to be a pattern emerging—but looks can be deceiving. Consider the next example:

V =
1 1 1
2 2 2

Two vectors in R^3, but the subspace that they span is still only a 1D subspace—a line (Figure 3-3). Why is that? It's because one vector in the set is already in the span of the other vector. Thus, in terms of span, one of the two vectors is redundant.

Figure 3-3: The 1D subspace (gray) spanned by two vectors (black)

So then, what is the relationship between the dimensionality of the spanned subspace and the number of vectors in the set? You might have guessed that it has something to do with linear independence.

The dimensionality of the subspace spanned by a set of vectors is the smallest number of vectors that forms a linearly independent set. If a vector set is linearly independent, then the dimensionality of the subspace spanned by the vectors in that set equals the number of vectors in that set. If the set is dependent, then the dimensionality of the subspace spanned by those vectors is necessarily less than the number of vectors in that set. Exactly how much smaller is another matter—to know the relationship between the number of vectors in a set and the dimensionality of their spanning subspace, you need to understand matrix rank, which you'll learn about in Chapter 6.

The formal definition of a vector subspace is a subset that is closed under addition and scalar multiplication, and includes the origin of the space. That means that any linear weighted combination of vectors in the subspace must also be in the same subspace.

Basis

How far apart are Amsterdam and Tenerife? Approximately 2,000. What does "2,000" mean? That number makes sense only if we attach a basis unit. A basis is like a ruler for measuring a space.

In this example, the unit is mile. So our basis measurement for Dutch-Spanish distance is 1 mile. We could, of course, use different measurement units, like nanometers or light-years, but I think we can agree that mile is a convenient basis for distance at that scale. What about the length that your fingernail grows in one day—should we still use miles? Technically we can, but I think we can agree that millimeter is a more convenient basis unit. To be clear: the amount that your fingernail has grown in the past 24 hours is the same, regardless of whether you measure it in nanometers, miles, or light-years. But different units are more or less convenient for different problems.

Back to linear algebra: a basis is a set of rulers that you use to describe the information in the matrix (e.g., data). Like with the previous examples, you can describe the same data using different rulers, but some rulers are more convenient than others for solving certain problems.

The most common basis set is the Cartesian axis: the familiar XY plane that you've used since elementary school. We can write out the basis sets for the 2D and 3D Cartesian graphs as follows:

S2 =
1 0
0 1

S3 =
1 0 0
0 1 0
0 0 1

Notice that the Cartesian basis sets comprise vectors that are mutually orthogonal and unit length. Those are great properties to have, and that's why the Cartesian basis sets are so ubiquitous (indeed, they are called the standard basis set).

But those are not the only basis sets. The following set is a different basis set for R^2.

T =
3 1
-3 1

Basis set S2 and T both span the same subspace (all of R^2). Why would you prefer T over S? Imagine we want to describe data points p and q. We can describe those data points as their relationship to the origin—that is, their coordinates—using basis S or basis T.

In basis S, those two coordinates are p = (3, 1) and q = (-6, 2). In linear algebra, we say that the points are expressed as the linear combinations of the basis vectors. In this case, that combination is 3*s1 + 1*s2 for point p, and -6*s1 + 2*s2 for point q.

Now let's describe those points in basis T. As coordinates, we have p = (1, 0) and q = (0, 2). And in terms of basis vectors, we have 1*t1 + 0*t2 for point p and 0*t1 + 2*t2 for point q (in other words, p = t1 and q = 2*t2). Again, the data points p and q are the same regardless of the basis set, but T provided a compact and orthogonal description.

Bases are extremely important in data science and machine learning. In fact, many problems in applied linear algebra can be conceptualized as finding the best set of basis vectors to describe some subspace. You've probably heard of the following terms: dimension reduction, feature extraction, principal components analysis, independent components analysis, factor analysis, singular value decomposition, linear discriminant analysis, image approximation, data compression. Believe it or not, all of those analyses are essentially ways of identifying optimal basis vectors for a specific problem.