Linear algebra refresher for LING 8570

Michael A. Covington
Institute for Artificial Intelligence
The University of Georgia
2008

December 10, 2008

1 Introduction

Vector and matrix algebra is called **linear algebra** because it arose in the context of solving systems of simultaneous linear equations. However, that is not what we are using it for.

For a more detailed review of the concepts covered here, see:

Farin and Hansford, *Practical linear algebra,* chapters 1–2 (especially good for those with little mathematics background).

Shifrin, T., and Adams, M. R. (2002) *Linear algebra: a geometric approach,* chapter 1.

2 Vectors

A **scalar** is a single number.

A **vector** is a sequence of numbers. Traditionally, a vector is represented by a boldface variable and is written either as a list of numbers in parentheses, or as a column in square brackets, like this:

$$\mathbf{v} \quad = \quad (2, 5, 9, 3) \quad = \quad \begin{bmatrix} 2 \\ 5 \\ 9 \\ 3 \end{bmatrix}$$

When doing handwritten work, you can denote boldface by underlining with a wavy line.

It is usual to represent the elements of a vector with ordinary algebraic variables written in italics and identified by subscript numbers:

$$\mathbf{v} \quad = \quad (v_1, v_2, v_3, v_4) \quad = \quad \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

Here v_1, v_2, v_3, v_4 stand for four numbers.

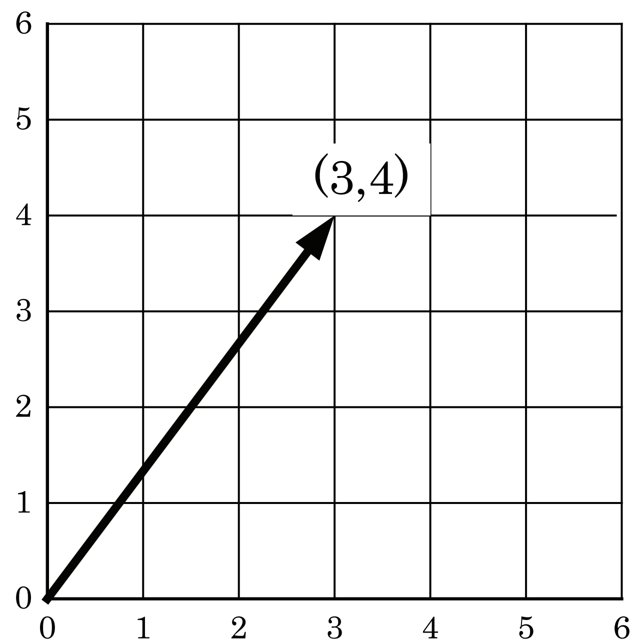

Figure 1: A vector, represented as an arrow drawn from the origin to a point.

3 Vectors and N-dimensional space

A vector of N numbers can be understood as analogous to the coordinates of a point in N-dimensional space. For instance, you need one number to find a point on a line; two numbers to find a point on a plane; three numbers to find a point in 3-dimensional space; and so on. You could understand $(3, 4)$ as "move to the right 3 units and move up 4 units."

Much of the appeal of vectors is that they make it easy to think in as many dimensions as you want, not just two or three.

Vectors are often drawn as arrows from the origin (point 0,0) to the point whose coordinates are given by the numbers in the vector.

Strictly speaking, a vector is not a point. It is a distance and direction. It would be the same vector if it were moved around, so long as its length and orientation didn't change.

4 Length of a vector

The length of a vector is:

$$||\mathbf{v}|| \qquad = \qquad \sqrt{{v_1}^2 + {v_2}^2 + {v_3}^2 + \cdots + {v_n}^2}$$

where $v_1, v_2, v_3, \ldots v_n$ are the elements of the vector $\mathbf{v}$.

That is: Square them all, add them up, and take the square root.

For example, the length of the vector $(3, 4)$ in the picture is $\sqrt{3^2 + 4^2} = 5$, as you can verify.

This is a generalization of the familiar 2-dimensional distance formula. It works in any number of dimensions.

5 The dot product

The dot product is a way of combining two vectors to get one scalar. What you do is multiply the corresponding elements and add the results:

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + v_3 w_3 + \cdots + v_n w_n$$

Consider for example the dot product of $(2, 3, 4)$ and $(4, 0, 7)$:

$$(2, 3, 4) \cdot (4, 0, 7) = 2 \cdot 4 + 3 \cdot 0 + 4 \cdot 7 = 8 + 0 + 28 = 36$$

The dot product is a measure both of how long the vectors are, and of how similar they are. If they are long, the numbers being multiplied will be large. If they are similar, large numbers will be multiplied by large numbers. Both of these things help to make the dot product large.

6 Comparing two vectors

Consider the two vectors:

$$(2, 4, 2) \qquad (20, 43, 21)$$

Obviously, they are different lengths; the second one contains much larger numbers. But the *proportions* between the numbers within them are very similar. In each vector, the third number is the same as the first (or very nearly), and the middle number is twice as big as the first (or very nearly).

How can we measure whether two vectors are similar in this way? One way is to treat them as arrows drawn from the origin, and measure the angle between the arrows.

In the case of the two vectors just given, one is about 10 times as long as the other, but they point in almost exactly the same direction.

Although space precludes proving it here, the angle θ between two vectors can be found by using the following formula:

$$\cos\theta = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\ ||\mathbf{w}||}$$

That is: Take the dot product of the two vectors, and divide it by the product of their lengths.

This formula assumes you are comparing only two vectors, but it works for any two vectors that have the same number of elements, whatever that might be. That is, it can give you an angle in two-, three-, four-, or 100-dimensional space.

The formula gives you the cosine of the angle, not the angle itself. You can use the $\cos^{-1}$ function to get the angle itself, or you can use the cosine as the measure of similarity. The cosine of a small angle is close to 1, and the cosine of a right angle (which is as far apart as vectors can get if they only contain positive numbers) is 0. Thus, 1 is maximum similarity and 0 is minimum similarity.

Exercise. Express the formula

$$\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\ ||\mathbf{w}||}$$

in terms of the vector elements v_1, v_2, etc., without using vector notation. Assume that each of the vectors has 4 elements.

7 Dot product as matrix multiplication

A **matrix** is a square or rectangular arrangement of numbers like this:

$$\begin{bmatrix} 7 & 0 & 6 \\ 2 & 3 & 7 \\ 4 & 2 & 6 \end{bmatrix}$$

(In case you're wondering, an arrangement of numbers in even more dimensions, like for instance a cube or stack of matrices, is called a **tensor**, or more precisely, a vector is a tensor of rank 1, a matrix is a tensor of rank 2, and higher ranks are permitted.)

If you think of a vector as representing a point, then a matrix represents a set of points or relations between points.

Matrix multiplication ("row-by-column multiplication") is outside the scope of this short paper, but if you are familiar with it, you will realize that the dot product of vectors is like what you get when you multiply a 1-row matrix by a 1-column matrix:

$$\begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1 b_1 + a_2 b_2 + a_3 b_3 \end{bmatrix}$$

The only difference is that matrix multiplication yields a 1-element matrix with a number in it, rather than just a number.

We said earlier that we usually treat a vector as a single-column matrix. Accordingly, to get the dot product by matrix multiplication, we must **transpose** one of the vectors to make it into a row instead of a column:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$\mathbf{v}^T = \mathbf{v} \text{ transposed} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$$

Accordingly, the dot product is sometimes written as matrix multiplication by a transposed matrix:

$$\mathbf{v} \cdot \mathbf{w} \quad = \quad \mathbf{v}^T\ \mathbf{w}$$

This notation is particularly popular in machine learning textbooks. *It is slightly sloppy* because it treats a 1-element matrix as a scalar.