Chapter 9

Neural Network Mathematics: Vectors and Matrices

Objectives

After reading this chapter you will be able to:

- understand the topics that fall under the heading of linear algebra;
- know what a vector is;
- know what a matrix is; and
- perform simple operations, such as addition and multiplication, with vectors and matrices.

9.1 Overview

The preceding chapters began our study of computational methods by introducing DEs and their application to models of the cellular level, models that drew their inspiration from biophysical considerations. One step up from biologically realistic models of single neurons is to consider simplified versions that obey simple rules to determine if they are "on" or "off." We can assemble these simple units into collections. Commonly, collections of simple on–off neurodes are called *neural networks*.

The relatively simple structure of these units and the rules for their combination make large networks computationally tractable, and an attractive setting to assess for properties that emerge with scale. Neural networks are used widely in computational modelling by both neuroscience and psychology. Some of the advances in this area are the result of a productive collaboration with computer science and physics. These disciplines have an expertise for mathematically analyzing large collections of simple elements. We use their tools when we assemble our collections of on–off units. The primary area of mathematics important for neural networks is *linear algebra*. Linear algebra is the mathematics of vectors and matrices. In this chapter I provide a brief fly-over of linear algebra. Then we examine how we can use these tools to develop our own versions of two types of simple neural networks: perceptrons and the Hopfield network.

9.2 Linear Algebra

Linear algebra is the branch of mathematics that deals with vectors and matrices. Like DEs, linear algebra is a broad domain of mathematics with many subdivisions and deep results. However, our needs are modest. A little familiarity with matrices will take our modelling a long way. Further, in many ways matrices are simpler than DEs, because the relevant operations are things like multiplication and addition, things we are already familiar with. We will have to learn new versions of these operations so that we can apply them to our new mathematical objects, but the concepts will be the same.

What Is a Vector?

We can think of vectors in several different ways, and these different ways may help us understand how to use vectors in different situations. A very direct way to think of a vector is as a list of numbers [1,2,3] (the resemblance to a Python list is intentional). However, it is better if we can think more generally. A richer conception of a vector is that it is a geometrical object, an arrow pointing into space. Think of the list of numbers as specifying the coordinates of the arrow's head in some coordinate space. The length of the numbers determines the *dimension* of the space. With this conception, our regular numbers can be seen as a special case of a vector with dimension 1. The points from the Cartesian plane (x, y) that we learned in secondary school are also vectors in a 2-D space.

The dimension of a vector can be arbitrarily large, even infinite, though in applications we stay finite. While it might seem complex to think of a vector as an arrow in a large, say 100-dimensional space, in another sense all vectors are just functions of two things: a direction and a magnitude. It does not matter how high the dimension of the space–a vector is just an arrow of a particular length pointing in a particular direction.

This geometric way to think of vectors as object in space with direction and magnitude helps us look for the qualitative behavior of equations involving matrices and vectors. From this point of view we can see that there are really only two things that we can do to a vector, geometrically speaking: change how long it is or change the direction it is pointing. All our mathematical operations on vectors, which make up the computations of neural networks, can be reconceived from this geometrical perspective.

Because vectors and matrices are the substrate of many neural network applications, we have to learn some basic notation and operations if we are to be able to make sense of the literature in this area. When a variable refers to a vector, instead of a single number (also called a *scalar*), it may be depicted as a lower case letter in a bold font, for example, **v**, or less commonly, with a little arrow on top, like $\vec{v}$.

Row or Column

When vectors are written out as a list of numbers it is conventional to place them between square brackets. It is also the convention that the orientation of this list is columnar (up and down). Row vectors exist, but they arise from *transposing* a column vector. You denote a transposed vector with a little "T" or apostrophe. To transpose means to to exchange rows with columns, and we will see this again with matrices.

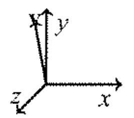

Figure 9.1 This is a visualization of a vector.

For example,

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

and

$$\mathbf{v}^{\mathrm{T}} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

Matrices

If we think of vectors as collections of numbers that specify a point in space, we can think of matrices as collections of vectors. With the convention that vectors are aligned in columns, a matrix can be thought of as having several vectors aligned side by side. Just as we could transpose a single vector (converting a column vector into a row vector) we can also transpose a matrix converting every column into a row. When a variable specifies a matrix it is usually represented as a bold font capital letter, e.g. **M**, with the transpose being written with a superscript "T" ($\mathbf{M}^{\mathrm{T}}$).

If vectors live in a space, what defines the space? Other vectors. If you look at Figure 9.1 you will see that the coordinate axes could be thought of as vectors too. The x, y, and z vectors are orthogonal and we can construct any other vector in this space by adding up a little from each of them. Any collection of vectors that is sufficient to define any other vector in the space is called a *basis*. The x, y, and z axes are a basis for 3-D Cartesian space, but there are other bases, and although we usually think of the elements of a basis as being orthogonal (a higher dimensional version of perpendicular) they do not have to be.

9.3 Elementary Vector and Matrix Operations

There are a few fundamental operations that can be done on vectors and matrices. Since we will be using our spreadsheet programs to implement simple neural networks, the following exercises take us through the basics of how to do this with a spreadsheet as well as their general definition. I present the operations you should try to work through on your own. I present the answer immediately afterwards.

Exercise: Adding Matrices

Open up your spreadsheet program and write the addition of,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ -4 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ -4 & 4 \end{bmatrix}$$

Adding vectors or matrices is just glorified addition. We add each term of one vector/matrix to the number in the same position of the other vector/matrix. This means that not every vector can be added to every other. They have to be the same size, otherwise they will not have numbers in exactly the same position for all possible positions:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ -4 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ -4 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 8 \\ -8 & 10 \end{bmatrix}$$

Multiplication by a Scalar

Use your spreadsheet to calculate the scaled vector

$$4 \times \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

When multiplying a matrix by a scalar, that is, a single number, you multiply each element of the matrix (or vector) by that same number:*

$$4 \times \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 12 \end{bmatrix}$$

There is a spreadsheet shortcut for implementing this. You can use the formula "MMULT." Most spreadsheet programs have some version of this function. It will let you specify one collection of cells in the spreadsheet for a matrix and then another number as the scalar for multiplication.

9.4 Think Geometric

As alluded to above, a geometrical conception of vectors and matrices can help you acquire a great deal of insight into neural networks. Geometrically, vectors have direction and magnitude. To use those quantities we need to be able to calculate them. Another way of thinking of the term "magnitude" is as a vector's length. But before we can talk about how to compute a vector's length, we need to say what we mean by "length."

*This is how you get subtraction. First, multiply by -1 then perform addition.

Something that refers to the size of mathematical objects is a *metric*. Metrics can be thought of as functions. They take a mathematical object as input and spit out a size. For a function to qualify as a metric it must obey a set of rules.

Metric Rules

- $d(x,y) \geq 0$
- $d(x,y) = 0, \Rightarrow x = y$
- $d(x,y) = d(y,x)$
- $d(z,x) \leq d(x,y) + d(y,z)$

Metrics cannot only tell us how big something is, but they can be used as a measure of distance. Therefore, I chose to use the letter d instead of f to emphasize that a metric function is related to distance.

We can repeat these rules in words:

- The function must give a size of at least zero when comparing two items; distance cannot be less than zero.
- If the distance between two things is zero, then they must be the same thing. Only one item per location.
- Whichever direction we go, the distance is the same. Running a lap on the track is the same distance whether I run clockwise or counterclockwise.
- Known as the triangle inequality, it is basically the same ideas as the old saying, "The shortest distance between two points is a straight line."

It is important to see that these rules do not tell you how to calculate the distance of something. The rules just tell you whether a function qualifies as a measure of distance. This abstractness is how higher mathematics gets its power, but it is also one of the reasons that non-mathematicians can feel at sea when dealing with mathematical ideas. The abstractness of the definition of a metric should suggest that there is more than one function that could qualify as a metric, but the lack of a specific example may leave you feeling uncertain about your understanding. When faced with abstract definitions, try to think of a concrete example in order to firm up your understanding. Just do not fall into the trap of thinking your concrete example is the sole example, or even the most important one.

A common distance metric is *Euclidean distance*. We used this in geometry to figure out the length of the hypotenuse of a right triangle. We squared each leg of the triangle, summed them together, and took the square root. We can do this for higher dimensions too:

$$\text{Length of vector } \mathbf{v} = \sqrt{\sum_{i=1}^{N} v_i^2} \tag{9.1}$$

Exercises: Computing Magnitudes

Knowing the formula for Euclidean distance compute the length of

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

If Euclidean distance is not the only metric, what other metrics are there? Try to think of how you would calculate distance on a sphere, like our globe. Does it make sense to use Euclidean distance on a curved surface? And what if you were in Manhattan? Would the distance between your favorite coffee shop and delicatessen be most accurately described by the Euclidean distance (as the crow flies) or by some other metric?

Inner Products

If you have taken any physics, you may have learned of the dot product operation for vectors. This is an example of a more general class of operations called an *inner product*. Like metrics, inner products are a sort of function that takes in objects, like vectors, and produces outputs like a scalar. Also, like metrics, "inner product" is a class where membership means obeying rules. We will not review the rules here, but it is worth knowing that the vector dot product is not the only example of an inner product, just like Euclidean distance is not the only example of a metric.

Vector Dot Product

The vector dot product is

$$\vec{\mathbf{x}}^T \vec{\mathbf{y}} = \sum_{i=1}^{\text{length of } \vec{x}} x_i y_i$$

In words this means that you match up the elements of the two vectors, same position, each with the other, and multiply them together. After this you add up the total. Note that the dot product takes in two vectors and spits out one number. It converts vectors to a scalar.

Exercise: A Vector Dot Product with Itself

Assume that $\vec{\mathbf{x}} = \vec{\mathbf{y}}$ in the definition above. Write out the formula for the dot product as far as you can. What does that formula remind you of?

It should remind you of the formula for the length of a vector (Equation 9.1). Taking the square root of the dot product of a vector with itself gives you the length of the vector.

Matrix Multiplication

As a mnemonic you can think of matrix multiplication as taking a repeated sequence of vector dot products, but with more bookkeeping. Remember that vectors can be columns or rows. When we multiply a vector against a matrix or a matrix against another matrix, the rows of the first one match up against the columns of the second one.

In the example below we take the first row [1,2] from the left matrix and match it up to the first column of the second [1,2]. We then multiply the pairs of numbers (1 × 1, 2 × 2), and sum up to get the new number that goes into spot (1,1) of the matrix. That is, we compute the dot product. The answer then goes in spot (1,1) because we used row 1 of the first matrix and column 1 of the second matrix. When you see an index to a matrix, like (2,3), remember that rows always come first.

If you are looking for a simple project for trying your hand at coding in a conventional computer language, see if you can write a function for computing the dot product between two lists of numbers.

Exercise: Matrix Multiplication in a Spreadsheet

Program your spreadsheet to multiply these two matrices:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ -4 & 6 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 & 3 \\ 2 & 4 & 4 & -2 \end{bmatrix}$$

Did you use the "MMULT" function to make your life easier?

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ -4 & 6 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 & 3 \\ 2 & 4 & 4 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 11 & 10 & -1 \\ 11 & 25 & 22 & 1 \\ 8 & 12 & 16 & -24 \\ 4 & 6 & 8 & -12 \end{bmatrix}$$

Based on this experience try to figure out why it is that you cannot multiply these two matrices together?

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ -4 & 6 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ -4 & 4 \end{pmatrix}$$

If you are not sure, try it.

It is because the matrices are not of compatible sizes. There are not enough columns of the left matrix to match up with the number of rows in the second matrix. This is a repeat of what we saw with addition. Unlike scalars where any one number can be added or multiplied against any other number, for vectors and matrices only some of them can be added or multiplied to some of the others.

Another difference from conventional multiplication is that order matters. That is, you will not necessarily get the same answer if you change the order of matrices in a multiplication. A quick question to test your understanding is to try and guess what shape two matrices must have for the question of the order of multiplication to even be an issue. To prove that the order of multiplication does matter for matrices try, either with pencil and paper or with a spreadsheet, to multiply the following two matrices in both directions **AB** and **BA** where

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix}$$

Situation Normal

In many applications, the length of the vectors which are elements of the space is not pertinent to the questions being asked, and having vectors of different magnitudes may obscure important relations. To eliminate the problem you can make all the vectors the same size. Take a moment to pause and see if you can figure out how you might do this. The name given to this process is *normalization*.

Here I named the vector **u** just so you don't start thinking there is any reason that a vector has to be called **v**.

The formula is

$$\hat{u} = \frac{\vec{\mathbf{u}}}{||\vec{\mathbf{u}}||} = \frac{\vec{\mathbf{u}}}{\sqrt{\sum_{i=1}^{N} u_i^2}}$$

What is the length of a vector after this procedure?

If we follow this procedure for a collection of vectors, they will all be the same size. How do we then measure the distance between them? This is another example of a distance metric. If all the vectors have their feet anchored at the origin of the coordinate system, and if they are all the same length, the only thing that can be different between them is the direction in which they are pointing. How can we calculate the angles they make one to another? If the angle is zero, they are pointing in the same direction. If it is 90 degrees they are perpendicular. We do not need to get out our protractors for this problem, because there is a formula relating the dot product to trigonometry:

$$\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = ||\vec{\mathbf{x}}||\,||\vec{\mathbf{y}}||\,\cos\theta \tag{9.2}$$

θ is the angle between the two vectors. Knowing this relation, can you determine what the angle between two vectors must be if $\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = 0$?

9.5 Functions for Matrices and Vectors

Just like there are functions for numbers, there are things that are like functions for vectors. Remember that our image of a function is as a machine that takes in a number and spits out another. This same metaphor can apply to a vector. A function for a vector would take in one vector and spit out another vector. Functions like this have the name *transformation*. Transformations are also special matrices. If you take a matrix of the proper size and multiply it against a vector you will get back a new vector. Thinking of a matrix as a function can be confusing at first, but it is powerful if you can absorb it. This is because it lets you think about matrices more generally, in terms of what they do, rather than some long, anonymous list of concatenated numbers. To help you develop this insight, figure out what the matrix

$$\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

does when multiplied against any vector of dimension 2.

And what kind of transformation does this matrix produce?

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

As another hint, try setting θ equal to 90 degrees and then sketch the answer.

Remember when considering your answers that a vector has both magnitude and direction.

9.6 Test Your Knowledge

To help you internalize some of these ideas about matrices, and to develop some knowledge of a few additional facts that will prove useful shortly, I have written a scavenger hunt of sorts. Using any old textbook, or even Wikipedia, try to find answers to the following questions:

1. What does it mean for a matrix to be invertible?
2. What is the transpose of a matrix?
3. What is an outer product? Compile a spreadsheet (or computer program) that performs this operation.
4. What does $[\mathbf{AB}]^{\mathrm{T}} =$ equal? Express your answer in terms of matrices **A** and **B**.

9.7 Summary

This chapter has given you a quick glimpse of some essential facts about vectors and matrices that are commonly used in the field of neural networks. In addition, I presented some of the common notation and concepts. In the following chapters, I will develop the powerful idea behind neural networks, that complexity can emerge from collections of simple elements, and then we will develop our own simple neural network models.

Computational Neuroscience and Cognitive Modelling

a student's introduction to methods and procedures

Britt Anderson

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Michael Carmichael
Editorial assistant: Keri Dicken
Production editor: Imogen Roome
Copyeditor: Neville Hankins
Proofreader: Audrey Scriven
Marketing manager: Alison Borg
Cover design: Wendy Scott

First published 2014

Library of Congress Control Number: 2013940255

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 978-1-4462-4929-1
ISBN 978-1-4462-4930-7

Contents