Matrix Mathematics

There are a number of standard mathematical operations on matrices that machine learning programs use repeatedly. We will briefly review some of the most fundamental of these operations.

The matrix transpose is a convenient operation that flips a matrix around its diagonal. Mathematically, suppose A is a matrix; then the transpose matrix A^T is defined by equation $A_{ij}^T = A_{ji}$. For example, the transpose of the rotation matrix R_α is

$$R_\alpha^T = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

Addition of matrices is only defined for matrices of the same shape and is simply performed elementwise. For example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}$$

Similarly, matrices can be multiplied by scalars. In this case, each element of the matrix is simply multiplied elementwise by the scalar in question:

$$2 \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$$

Furthermore, it is sometimes possible to multiply two matrices directly. This notion of matrix multiplication is probably the most important mathematical concept associated with matrices. Note specifically that matrix multiplication is not the same notion as elementwise multiplication of matrices! Rather, suppose we have a matrix A of shape (m, n) with m rows and n columns. Then, A can be multiplied on the right by any matrix B of shape (n, k) (where k is any positive integer) to form matrix AB of shape (m, k). For the actual mathematical description, suppose A is a matrix of shape (m, n) and B is a matrix of shape (n, k). Then AB is defined by

$$(AB)_{ij} = \sum_k A_{ik} B_{kj}$$

We displayed a matrix multiplication equation earlier in brief. Let's expand that example now that we have the formal definition:

$$\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\alpha) \cdot 1 - & \sin(\alpha) \cdot 0 \\ \sin(\alpha) \cdot 1 - & \cos(\alpha) \cdot 0 \end{pmatrix} = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix}$$

The fundamental takeaway is that rows of one matrix are multiplied against columns of the other matrix.

This definition hides a number of subtleties. Note first that matrix multiplication is not commutative. That is, $AB \neq BA$ in general. In fact, AB can exist when BA is not meaningful. Suppose, for example, A is a matrix of shape (2, 3) and B is a matrix of shape (3, 4). Then AB is a matrix of shape (2, 4). However, BA is not defined since the respective dimensions (4 and 2) don't match. As another subtlety, note that, as in the rotation example, a matrix of shape (m, n) can be multiplied on the right by a matrix of shape $(n, 1)$. However, a matrix of shape $(n, 1)$ is simply a column vector. So, it is meaningful to multiply matrices by vectors. Matrix-vector multiplication is one of the fundamental building blocks of common machine learning systems.

One of the nicest properties of standard multiplication is that it is a linear operation. More precisely, a function f is called linear if $f(x + y) = f(x) + f(y)$ and $f(cx) = cf(x)$ where c is a scalar. To demonstrate that scalar multiplication is linear, suppose that a, b, c, d are all real numbers. Then we have

$$a \cdot (b \cdot c) = b \cdot (ac)$$

$$a \cdot (c + d) = ac + ad$$

We make use of the commutative and distributive properties of scalar multiplication here. Now suppose that instead, A, C, D are now matrices where C, D are of the same size and it is meaningful to multiply A on the right with either C or D (b remains a real number). Then matrix multiplication is a linear operator:

$$A(b \cdot C) = b \cdot (AC)$$

$$A(C + D) = AC + AD$$

Put another way, matrix multiplication is distributive and commutes with scalar multiplication. In fact, it can be shown that any linear transformation on vectors corresponds to a matrix multiplication. For a computer science analogy, think of linearity as a property demanded by an abstract method in a superclass. Then standard multiplication and matrix multiplication are concrete implementations of that abstract method for different subclasses (respectively real numbers and matrices).

Tensors

In the previous sections, we introduced the notion of scalars as rank-0 tensors, vectors as rank-1 tensors, and matrices as rank-2 tensors. What then is a rank-3 tensor? Before passing to a general definition, it can help to think about the commonalities