

Mathematics Fundamentals

Reid Pryzant

September 3rd, 2015

This section introduces the basic mathematical concepts that are needed to understand deep learning. It represents a bare-bones framework and makes no pretensions about completeness.

1 Algebra

A *function* is a relation between a set of inputs and outputs. Each input has a unique output. You’ve probably seen functions before. $f(x) = x + 2$ is an example of a function. There are a few ways of thinking about functions, however, that you may not have been exposed to.

A function can be thought of as a *mapping* of inputs to outputs. In this way, a function f is written as

$$f : X \rightarrow Y$$

f maps elements $x \in X$ (the \in means “in”) to $y \in Y$. The set of possible inputs (in this case, X) is the *domain* of f and the possible outputs (Y) is the *range* of f .

We are used to functions of the form $f : \mathbb{R} \rightarrow \mathbb{R}$ (functions that map real numbers to real numbers). However, functions can be abstracted to represent almost any kind of mapping. We will be especially interested in functions of the form $f : \mathbb{R} \rightarrow A$ where A is a set of classifications or labels.

2 Linear Algebra

Linear algebra is the mathematics of matrices and vectors. This makes it appropriate for dealing with large systems of equations, something we will encounter frequently during our exploration of deep learning.

A *matrix* is a 2-D array of numbers. Matrices are referred to with bold upper-case letters. The *height* of a matrix is the number of rows in it and the *width* of a matrix is the number of columns. The elements of a matrix have two indices. The first is the row index and the second is the column index. The size of a matrix \mathbf{A} is denoted $|\mathbf{A}|$. Below is a matrix with height 3, length 2, and size 6.

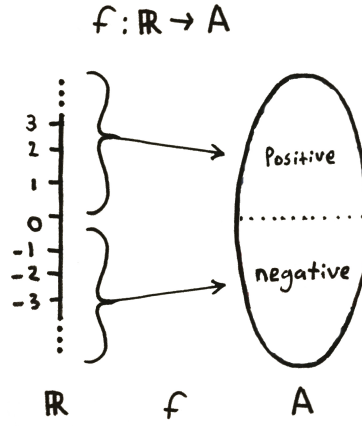


Figure 1: In this case, f is a function that assigns one of two labels to every real number. The labels are representative of sign.

We may also refer to this matrix as a “3 by 2” matrix.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

There are two operations on matrices that you should know:

1. The *transpose* of a matrix is a new matrix whose rows are the columns of the original. The transpose of a matrix \mathbf{A} is denoted \mathbf{A}^\top .

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad \mathbf{A}^\top = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$$

2. *Matrix multiplication* is done in a row-by-column fashion. The rows of the multiplicand are combined with the columns of the multiplier:

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \end{bmatrix}$$

Higher dimensional arrays of numbers are called *tensors*. These have the same properties and notation as matrices, so we will refer to them simply as matrices for the rest of this book.

A *vector* is a 1-D array of numbers that is signified by a bold lower-case letter. Each element in a vector is known as an *attribute*, and can be identified by its index. A vector can also be thought of as a single column of a matrix. This means that vectors share many properties with matrices, like multiplication:

$$\mathbf{x}^\top \mathbf{v} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \sum_{i=1}^n x_i v_i$$

An invaluable property of vectors is their spatial interpretation. A vector can be placed into a coordinate system. In this interpretation, each vector can be visualized as an arrow pointing from the origin to whatever point is specified by the vector's elements. The number of dimensions a vector exists in is equal to its size (the number of attributes).

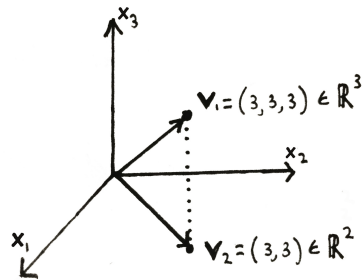


Figure 2: Vector-space visualization of two vectors \mathbf{v}_1 and \mathbf{v}_2 .

Vectors can be added to each other (if they are the same dimension) and multiplied by *scalars*, or numbers. Adding two vectors is akin to attaching two arrows to each other in our coordinate system, while multiplying a vector by a number x is equivalent to scaling the arrow by x .

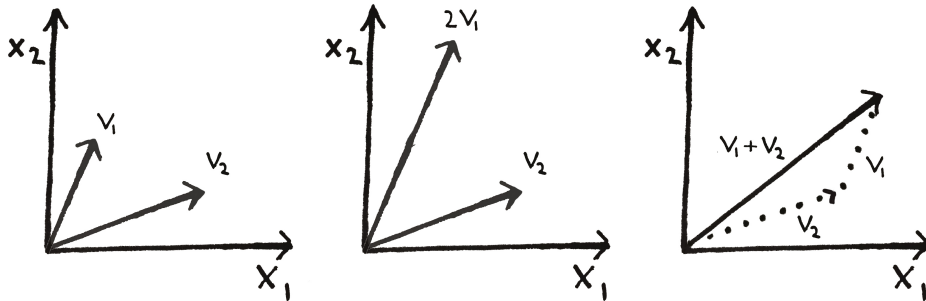


Figure 3: Vectors can be scaled (*middle*) or summed (*right*).

3 Multivariable Calculus

The *derivative* of a function $f(x)$ measures the slope of the line tangent to f . It is written $\frac{df}{dx}$ or $f'(x)$ because you are taking the derivative of f *with respect to* x .

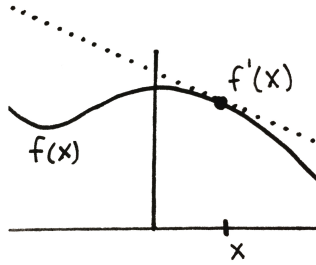


Figure 4: A pointwise derivative of a function f at a location x .

Many functions take several inputs. Some functions can even take a vector or matrix as input. If this is the case and we want to see how the function depends on one of its many inputs, we can take a *partial derivative*. If f takes many inputs such that $f(x_1, x_2, \dots, x_n) = \dots$, then we say that $\frac{\partial f}{\partial x_i}$ is the derivative of f with respect to x_i .

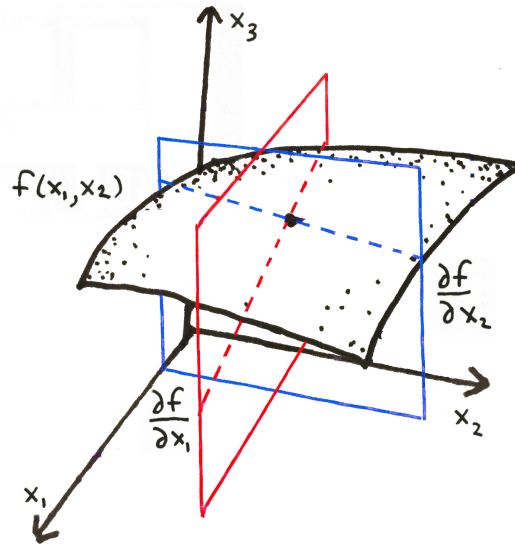


Figure 5: Partial derivatives of a function f with respect to each of its inputs x_1 and x_2 .

The derivative of a vector produces a special matrix called the *Jacobian*. If f is a function such that $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (f takes vectors of length n as input and produces vectors of length m as output), then the Jacobian of f (denoted \mathbf{J}) is a matrix with m rows and n columns, where $j_{ij} = \frac{\partial f_i}{\partial x_j}$.

4 Machine Learning Fundamentals

Deep learning is a branch of machine learning. In order to build an understanding of deep learning, then, we need background on the basic principles of