2 Notation and Definitions

2.1 Notation

Let's start by revisiting the mathematical notation we all learned at school, but some likely forgot right after the prom.

2.1.1 Data Structures

A *scalar* is a simple numerical value, like 15 or -3.25. Variables or constants that take scalar values are denoted by an italic letter, like x or a.

A *vector* is an ordered list of scalar values, called attributes. We denote a vector as a bold character, for example, $\mathbf{x}$ or $\mathbf{w}$. Vectors can be visualized as arrows that point to some directions as well as points in a multi-dimensional space. Illustrations of three two-dimensional vectors, $\mathbf{a} = [2, 3]$, $\mathbf{b} = [-2, 5]$, and $\mathbf{c} = [1, 0]$ are given in Figure 1. We denote an attribute of a vector as an italic value with an index, like this: $w^{(j)}$ or $x^{(j)}$. The index j denotes a specific *dimension* of the vector, the position of an attribute in the list. For instance, in the vector $\mathbf{a}$ shown in red in Figure 1, $a^{(1)} = 2$ and $a^{(2)} = 3$.

The notation $x^{(j)}$ should not be confused with the power operator, like this x^2 (squared) or x^3 (cubed). If we want to apply a power operator, say square, to an indexed attribute of a vector, we write like this: $(x^{(j)})^2$.

A variable can have two or more indices, like this: $x_i^{(j)}$ or like this $x_{i,j}^{(k)}$. For example, in neural networks, we denote as $x_{l,u}^{(j)}$ the input feature j of unit u in layer l.

A **matrix** is a rectangular array of numbers arranged in rows and columns. Below is an example of a matrix with two rows and three columns,

$$\begin{bmatrix} 2 & 4 & -3 \\ 21 & -6 & -1 \end{bmatrix}.$$

Matrices are denoted with bold capital letters, such as $\mathbf{A}$ or $\mathbf{W}$.

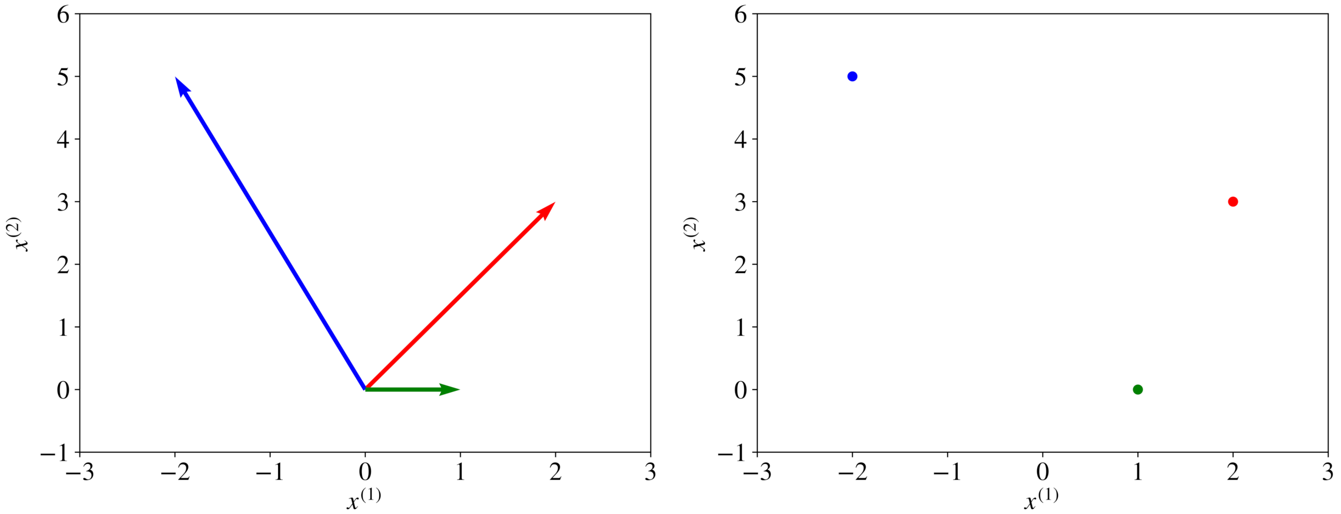

Figure 1: Three vectors visualized as directions and as points.

A *set* is an unordered collection of unique elements. We denote a set as a calligraphic capital character, for example, $\mathcal{S}$. A set of numbers can be finite (include a fixed amount of values). In this case, it is denoted using accolades, for example, $\{1, 3, 18, 23, 235\}$ or $\{x_1, x_2, x_3, x_4, \ldots, x_n\}$. A set can be infinite and include all values in some interval. If a set includes all values between a and b, including a and b, it is denoted using brackets as $[a, b]$. If the set doesn't include the values a and b, such a set is denoted using parentheses like this: (a, b). For example, the set $[0, 1]$ includes such values as 0, 0.0001, 0.25, 0.784, 0.9995, and 1.0. A special set denoted $\mathbb{R}$ includes all numbers from minus infinity to plus infinity.

When an element x belongs to a set $\mathcal{S}$, we write $x \in \mathcal{S}$. We can obtain a new set $\mathcal{S}_3$ as an *intersection* of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$. In this case, we write $\mathcal{S}_3 \leftarrow \mathcal{S}_1 \cap \mathcal{S}_2$. For example $\{1, 3, 5, 8\} \cap \{1, 8, 4\}$ gives the new set $\{1, 8\}$.

We can obtain a new set $\mathcal{S}_3$ as a *union* of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$. In this case, we write $\mathcal{S}_3 \leftarrow \mathcal{S}_1 \cup \mathcal{S}_2$. For example $\{1, 3, 5, 8\} \cup \{1, 8, 4\}$ gives the new set $\{1, 3, 4, 5, 8\}$.

2.1.2 Capital Sigma Notation

The summation over a collection $X = \{x_1, x_2, \ldots, x_{n-1}, x_n\}$ or over the attributes of a vector $\mathbf{x} = [x^{(1)}, x^{(2)}, \ldots, x^{(m-1)}, x^{(m)}]$ is denoted like this:

$$\sum_{i=1}^{n} x_i \stackrel{\text{def}}{=} x_1 + x_2 + \ldots + x_{n-1} + x_n, \text{ or else: } \sum_{j=1}^{m} x^{(j)} \stackrel{\text{def}}{=} x^{(1)} + x^{(2)} + \ldots + x^{(m-1)} + x^{(m)}.$$

The notation $\stackrel{\text{def}}{=}$ means "is defined as".

2.1.3 Capital Pi Notation

A notation analogous to capital sigma is the *capital pi notation*. It denotes a product of elements in a collection or attributes of a vector:

$$\prod_{i=1}^{n} x_i \stackrel{\text{def}}{=} x_1 \cdot x_2 \cdot \ldots \cdot x_{n-1} \cdot x_n,$$

where $a \cdot b$ means a multiplied by b. Where possible, we omit $\cdot$ to simplify the notation, so ab also means a multiplied by b.

2.1.4 Operations on Sets

A derived set creation operator looks like this: $\mathcal{S}' \leftarrow \{x^2 \mid x \in \mathcal{S}, x > 3\}$. This notation means that we create a new set $\mathcal{S}'$ by putting into it x squared such that x is in $\mathcal{S}$, and x is greater than 3.

The cardinality operator $|\mathcal{S}|$ returns the number of elements in set $\mathcal{S}$.

2.1.5 Operations on Vectors

The sum of two vectors $\mathbf{x} + \mathbf{z}$ is defined as the vector $[x^{(1)} + z^{(1)}, x^{(2)} + z^{(2)}, \ldots, x^{(m)} + z^{(m)}]$.

The difference of two vectors $\mathbf{x} - \mathbf{z}$ is defined as $[x^{(1)} - z^{(1)}, x^{(2)} - z^{(2)}, \ldots, x^{(m)} - z^{(m)}]$.

A vector multiplied by a scalar is a vector. For example $\mathbf{x}c \stackrel{\text{def}}{=} [cx^{(1)}, cx^{(2)}, \ldots, cx^{(m)}]$.

A *dot-product* of two vectors is a scalar. For example, $\mathbf{wx} \stackrel{\text{def}}{=} \sum_{i=1}^{m} w^{(i)} x^{(i)}$. In some books, the dot-product is denoted as $\mathbf{w} \cdot \mathbf{x}$. The two vectors must be of the same dimensionality. Otherwise, the dot-product is undefined.

The multiplication of a matrix $\mathbf{W}$ by a vector $\mathbf{x}$ results in another vector. Let our matrix be,

$$\mathbf{W} = \begin{bmatrix} w^{(1,1)} & w^{(1,2)} & w^{(1,3)} \\ w^{(2,1)} & w^{(2,2)} & w^{(2,3)} \end{bmatrix}.$$

When vectors participate in operations on matrices, a vector is by default represented as a matrix with one column. When the vector is on the right of the matrix, it remains a column vector. We can only multiply a matrix by vector if the vector has the same number of rows