Vectors and Vector spaces

- Many quantities in connectionist models are best represented as vectors (e.g., a group of neurons and weights on the inputs to a given neuron)

- A vector space is a set $V$ of elements, called vectors, with the following properties:
  
  - To every pair, $\mathbf{u}$ and $\mathbf{v}$, of vectors in $V$, there corresponds a vector $\mathbf{u} + \mathbf{v}$ also in $V$, called the sum of $\mathbf{u}$ and $\mathbf{v}$, in such a way that addition is commutative and associative
  
  - For any scalar $c$ and any vector $\mathbf{v}$ in $V$, there is a vector $c \mathbf{v}$ in $V$, called the product of $c$ and $\mathbf{v}$, in such a way that multiplication by scalars is associative and distributive with respect to vector addition
  
  - (and a few other axioms …)
Vectors

- A vector is a useful tool to represent patterns of numbers:
  - For instance, a person’s age (y), height (in), and weight (lb):
    
    \[
    \begin{align*}
    \text{Joe} & \quad \begin{bmatrix} 37 \\ 72 \\ 175 \end{bmatrix} \\
    \text{Mary} & \quad \begin{bmatrix} 10 \\ 30 \\ 61 \end{bmatrix} \\
    \text{Carol} & \quad \begin{bmatrix} 25 \\ 65 \\ 121 \end{bmatrix} \\
    \text{Brad} & \quad \begin{bmatrix} 66 \\ 67 \\ 155 \end{bmatrix}
    \end{align*}
    \]

- Each of these vectors has three components

Visualising vectors

- We can neatly visualise vectors with no more than three components:
  - This will prove helpful in developing a geometrical intuition about vectors (but everything we discuss extends to any number of components)
Scalar multiplication

- A scalar is a single real number, and vectors can be multiplied by scalars:

\[ 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \]

- Multiplying a vector \( \mathbf{v} \) by a positive scalar \( s \) yields a vector \( \mathbf{v}' \) that points in the same direction as \( \mathbf{v} \), but that is longer or shorter by magnitude \( s \)

- Multiplying \( \mathbf{v} \) by a negative scalar, also yields a lengthened or shortened vector \( \mathbf{v}' \), but this time one pointing in the opposite direction of \( \mathbf{v} \)

- Two vectors are said to be collinear, if they are scalar multiples of one another

Addition

- Vectors with an equal number of components can be added:

\[ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \]

- \( \mathbf{v}_1 + \mathbf{v}_2 \) lies in between \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \), and forms the diagonal of a parallelogram with \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \)

- Vector addition is associative: \((\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)\)

- Vector addition is commutative: \(\mathbf{v}_3 + \mathbf{v}_2 + \mathbf{v}_1 = \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\)
Example: Addition and Scalar multiplication

- Using addition and scalar multiplication, we can compute averages:

\[ u = \frac{1}{4} \left( \begin{bmatrix} 37 \\ 72 \\ 175 \end{bmatrix} + \begin{bmatrix} 10 \\ 30 \\ 61 \end{bmatrix} + \begin{bmatrix} 25 \\ 65 \\ 121 \end{bmatrix} + \begin{bmatrix} 66 \\ 67 \\ 155 \end{bmatrix} \right) = \begin{bmatrix} 34.5 \\ 58.5 \\ 128 \end{bmatrix} \]

- In vector notation:

\[ u = \frac{1}{4} \left( v_1 + v_2 + v_3 + v_4 \right) \]

- Vector \( u \) is a linear combination of vectors \( v_1, v_2, v_3, \) and \( v_4 \), and contains the averages of their components

- Scalar multiplication is distributive: \( \frac{1}{4} v_1 + \frac{1}{4} v_2 + \frac{1}{4} v_3 + \frac{1}{4} v_4 = \frac{1}{4} (v_1 + v_2 + v_3 + v_4) \)

Linear combinations

- A vector \( v \) is a linear combination of vectors \( v_1, v_2, \ldots, v_n \) if there are scalars \( c_1, c_2, \ldots, c_n \) such that:

\[ v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n \]

- Example: \( u = \begin{bmatrix} 9 \\ 10 \end{bmatrix} \) is a linear combination of \( v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \) and \( v_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \):

\[ u = 3 \cdot v_1 + 2 \cdot v_2 \]

- We effectively find scalars to adjust \( v_1 \) and \( v_2 \) to form a parallelogram with \( u \)

- Any vector in the shaded area can be constructed in this way using positive scalars
Linear combinations (cont’d)

- The set of all linear combinations of $v_1, v_2, \ldots, v_n$ is said to be the set spanned by $v_1, v_2, \ldots, v_n$

- Example: the vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ span all of three-dimensional space, because any $v = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ can be written as: $v = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

- We call these vectors the standard basis for three-dimensional space

- Q: What about the basis of $n$-dimensional space?

n-dimensional space

- An $n$-dimensional space is the set of vectors spanned by a set of $n$ linearly independent vectors, which we refer to as the basis for that space

  - A set is linearly independent if it does not contain any vector $v_i$ that can be written as a linear combination of other vectors in the set

  - Conversely, a set is linearly dependent if it does contain a vector $v_i$ that can be written as a linear combination of other vectors in the set

- Consequence 1: If a set of $n$ vectors is linearly dependent, it spans less than $n$-dimensional space

- Consequence 2: There can no more than $n$ linearly independent vectors in $n$-dimensional space

- Consequence 3: There is only one way in which a vector can be written as a linear combination of a set of linear independent vectors (i.e., coefficients are unique)
Vectors and Vector spaces

• Lists of numbers, geometrical arrows, n-dimensional space—just what exactly is a vector?

• A vector space is a set \( V \) of elements, called vectors, with the following properties:
  
  • To every pair, \( u \) and \( v \), of vectors in \( V \), there corresponds a vector \( u + v \) also in \( V \), called the sum of \( u \) and \( v \), in such a way that addition is commutative and associative

  • For any scalar \( c \) and any vector \( v \) in \( V \), there is a vector \( cv \) in \( V \), called the product of \( c \) and \( v \), in such a way that multiplication by scalars is associative and distributive with respect to vector addition

  • (and a few other axioms …)

  • … a vector is a rather undefined object; anything obeying these rules is a vector space (e.g., the set of polynomials of order \( n \) is a vector space)

• We use numbers to represent vectors, and we refer to vector components as coordinates in vector space, because these components are unique coefficients for a given basis

Inner products

• If we multiply two vectors \( \mathbf{v} \) and \( \mathbf{w} \) with the same number of components, we obtain their inner product \( \mathbf{v} \cdot \mathbf{w} \):

\[
\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}
\]

\[
\mathbf{v} \cdot \mathbf{w} = (3 \cdot 1) + (-1 \cdot 2) + (2 \cdot 1) = 3.
\]

• The inner product between two vectors is a measure of their similarity:

  • The closer they are in space, the more positive the inner product

  • The more they point in opposite direction, the more negative
Inner products: Length

• We can use the inner product of a vector with itself to measure its length:

\[ v = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad v \cdot v = 3^2 + 4^2 = 25. \]

Hence, following the Pythagorean theorem, we define vector length as:

\[ \|v\| = (v \cdot v)^{1/2} = \sqrt{v \cdot v} \]

• This definition includes our intuitions about length:

\[ \|cv\| = |c| \|v\| \quad \|v_1 + v_2\| \leq \|v_1\| + \|v_2\| \]

(recall the parallelogram)

Inner products: Angle

• We can also use the inner product to measure the angle between two vectors \( v \) and \( w \) (= their inner product adjusted for their lengths):

\[ \cos \theta = \frac{v \cdot w}{\|v\| \|w\|} = \cos \theta = \frac{\sum_{i=1}^{n} v_i w_i}{(\sum_{i=1}^{n} v_i^2)^{1/2}(\sum_{i=1}^{n} w_i^2)^{1/2}} \]

• Example: The angle between \( v_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \) and \( v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \) is found through:

\[ v_1 \cdot v_2 = 1 \quad \|v_1\| = 1 \quad \|v_2\| = \sqrt{2} \quad \cos \theta = \frac{1}{1 \cdot \sqrt{2}} = 0.707 \]

Hence: \( \theta = \cos^{-1} (0.707) = 45^\circ \)

• If \( \cos \theta = 0 \) (=90°), two vectors are orthogonal (at right angles) to one another.
Connectionist Language Processing — Crocker & Brouwer

Connectionist: A single unit

- The activation of unit $u$ computes the inner product of $w$ and $v$: $u = w \cdot v$

- The output of unit $u$ effectively indicates how close an input vector $v$ is to the weight vector $w$ (close $\rightarrow +$; near orthogonal $\rightarrow 0$; opposite $\rightarrow -$)

- A unit thus effectively divides the input space into two parts: a part to which its response is positive and a part where to which its response is negative

Connectionist Language Processing — Crocker & Brouwer

Matrices

- To describe a full layer, we need the concept of a $m \times n$ matrix—an array of real numbers:

  $$ M = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} \quad N = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 10 & -1 \\ -1 & 27 \end{bmatrix} $$

- Matrices, like vectors, can be multiplied by a scalar:

  $$ 3M = 3 \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 12 & 15 \\ 3 & 0 & 3 \end{bmatrix} $$

- Two matrices with the same number of rows and columns can be added:

  $$ M + N = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 2 \\ 4 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 7 \\ 5 & 1 & 0 \end{bmatrix} $$

Connectionist Language Processing — Crocker & Brouwer
Multiplying a Vector by a Matrix

- A \( m \times n \) matrix \( W \) can be multiplied by an \( n \)-component vector \( v \), yielding an \( m \)-component vector \( u \), consisting of the inner products between vector \( v \) and each of the row vectors \( w_i \) of \( W \):

\[
\begin{pmatrix}
  \vdots \\
  v_1 \\
  \vdots \\
  v_n
\end{pmatrix}
\cdot
\begin{pmatrix}
  w_1 \\
  \vdots \\
  w_m
\end{pmatrix}
= v_1 w_1 + v_2 w_2 + \cdots + v_n w_n
\]

\[
Wv = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = [3\cdot1 + 4\cdot0 + 5\cdot2 \\ 1\cdot1 + 0\cdot0 + 1\cdot2] = [13 \\ 3]
\]

- From another perspective, \( u \) is linear combination of the column vectors \( w_j \) of \( W \) with the components of \( v \) as coefficients:

\[
\begin{pmatrix}
  w_1 \\
  \vdots \\
  w_m
\end{pmatrix}
\cdot
\begin{pmatrix}
  v_1 \\
  \vdots \\
  v_n
\end{pmatrix}
= v_1 w_1 + v_2 w_2 + \cdots + v_n w_n
\]

\[
W = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix}
\]

\[
v = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}
\]

\[
Wv = \begin{bmatrix} 3 \cdot 1 + 4 \cdot 0 + 5 \cdot 2 \\ 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 2 \end{bmatrix} = [13 \\ 3]
\]

Vector-Matrix Multiplication as a Function

- The space spanned by the column vectors of a matrix is the column space, and the vector \( u = Wv \) is in the column space of \( W \)

- Matrix \( W \) is thus effectively a function from one set of vectors to another

- That is, if we consider an \( n \)-dimensional vector space \( V \) (the domain) and an \( m \)-dimensional vector space \( U \) (the range), multiplication by a fixed matrix \( W \) is a function from \( V \) to \( U \):
**Connectionism: A single layer**

- Each unit $u_i$ has its own weight vector $w_i$, and the activation of unit $u_i$ is the inner product of $w_i$ and input vector $v$: $u_i = w_i \cdot v$

- If we define a matrix $W$ that has weight vectors $w_i$ as its row vectors, the activation of all units $u_i$ is neatly given as: $u = Wv$

- Each unit $u_i$ matches its weight vector $w_i$ to the input vector $v$

**Connectionism: Multiple layers**

- In a multilayer network, the output vector $u = Mz$ depends on $z = Nv$

- Hence, the output of the network relates to the input through $u = M(Nv)$

- Using matrix multiplication:

$$
\begin{bmatrix}
M \\
N \\
P
\end{bmatrix}
\begin{bmatrix}
n_1 & n_2 & \cdots & n_k \\
M_{n_1} & M_{n_2} & \cdots & M_{n_k}
\end{bmatrix}
= \begin{bmatrix}
3 & 4 & 5 \\
1 & 0 & 1 \\
0 & 1 & 2
\end{bmatrix}
\begin{bmatrix}
1 & 2 \\
0 & 1 \\
-1 & 1
\end{bmatrix}
= \begin{bmatrix}
(3+8-5) & (6+0+5) \\
(1+0-1) & (2+0+1) \\
(0+2-2) & (0+0+2)
\end{bmatrix}
= \begin{bmatrix}
6 & 11 \\
0 & 3 \\
0 & 2
\end{bmatrix}
$$

we can rewrite it as a single layer system $u = M(Nv) = (MN)v = Pv$

- This is why we use a non-linear (e.g., logistic) transformation on the inner products as outputs of $z$ and $u$, thereby effectively modelling decisions
Linear versus Nonlinear systems

- A function $y = f(x)$ describes a linear system, if for any inputs $x_1$ and $x_2$, the following equations hold:
  
  - $f(cx) = cf(x)$
  
  - $f(x_1 + x_2) = f(x_1) + f(x_2)$

- Linear systems are easy to analyse; once we know its responses to a set of inputs forming the basis of the input space, we can compute its response to any other input

- Nonlinear systems are simply all systems for which these equations do not hold, and are therefore more difficult to analyse

In sum …