Connectionist Language Processing

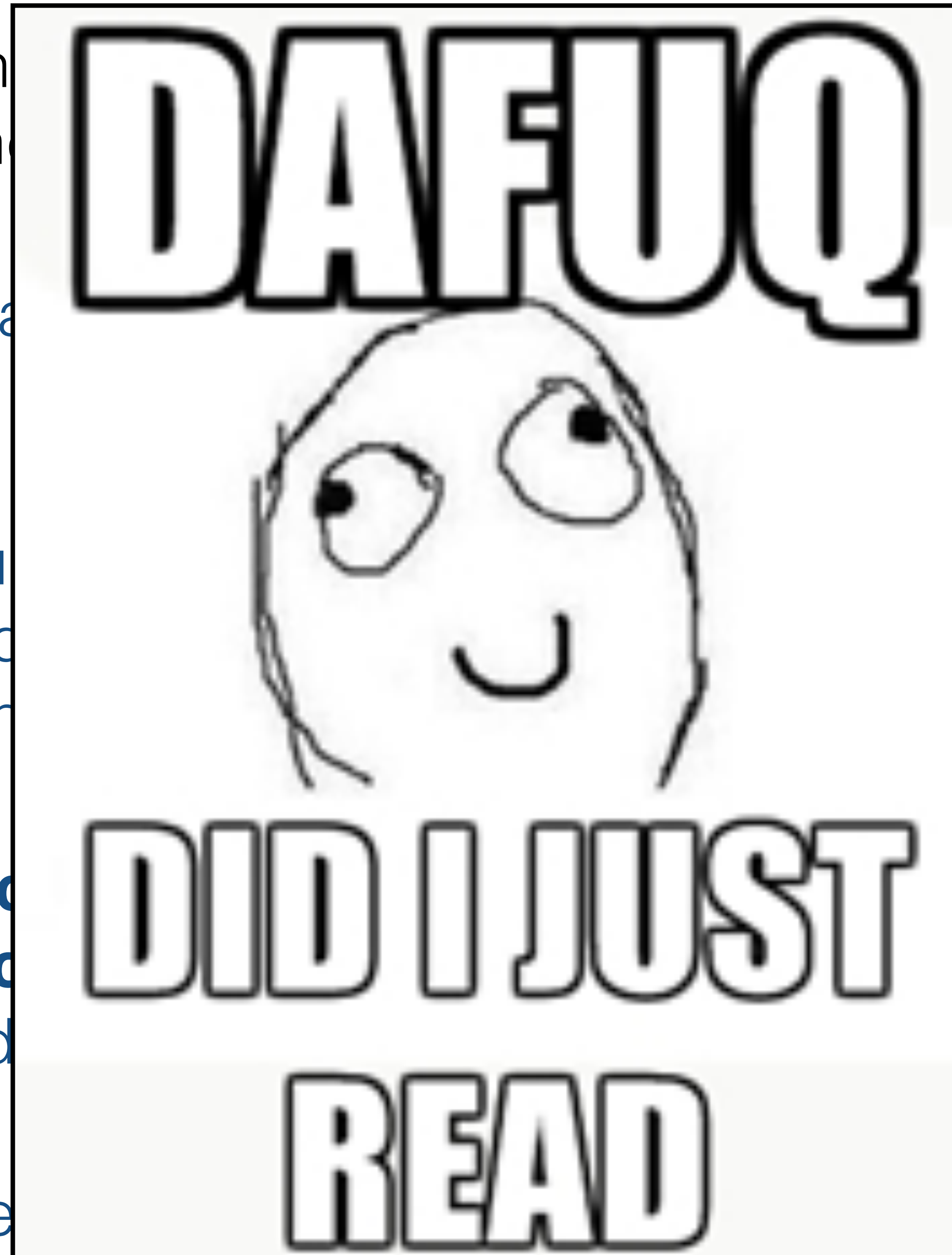
Lecture 2: **A Primer on Linear Algebra**

Based on: Jordan, M. I. (1986). An introduction to linear algebra in parallel distributed processing. *Parallel distributed processing*, 1, 365-422.

Matthew W. Crocker
crocker@coli.uni-sb.de
Harm Brouwer
brouwer@coli.uni-sb.de

Vectors and Vector spaces

- Many quantities in … presented as **vectors** (e.g., a group of n… to a given neuron)
- A **vector space** is a… , with the following properties:
  - To every pair, **u**… esponds a vector **u + v** also in ***V***, called… y that addition is **commutative** an…
  - For any scalar **c**… vector **cv** in ***V***, called the product of **c**… cation by scalars is **associative** and… addition
  - (and a few othe…



Vectors

- A **vector** is a useful tool to represent patterns of numbers:
    - For instance, a person's **age** (y), **height** (in), and **weight** (lb):

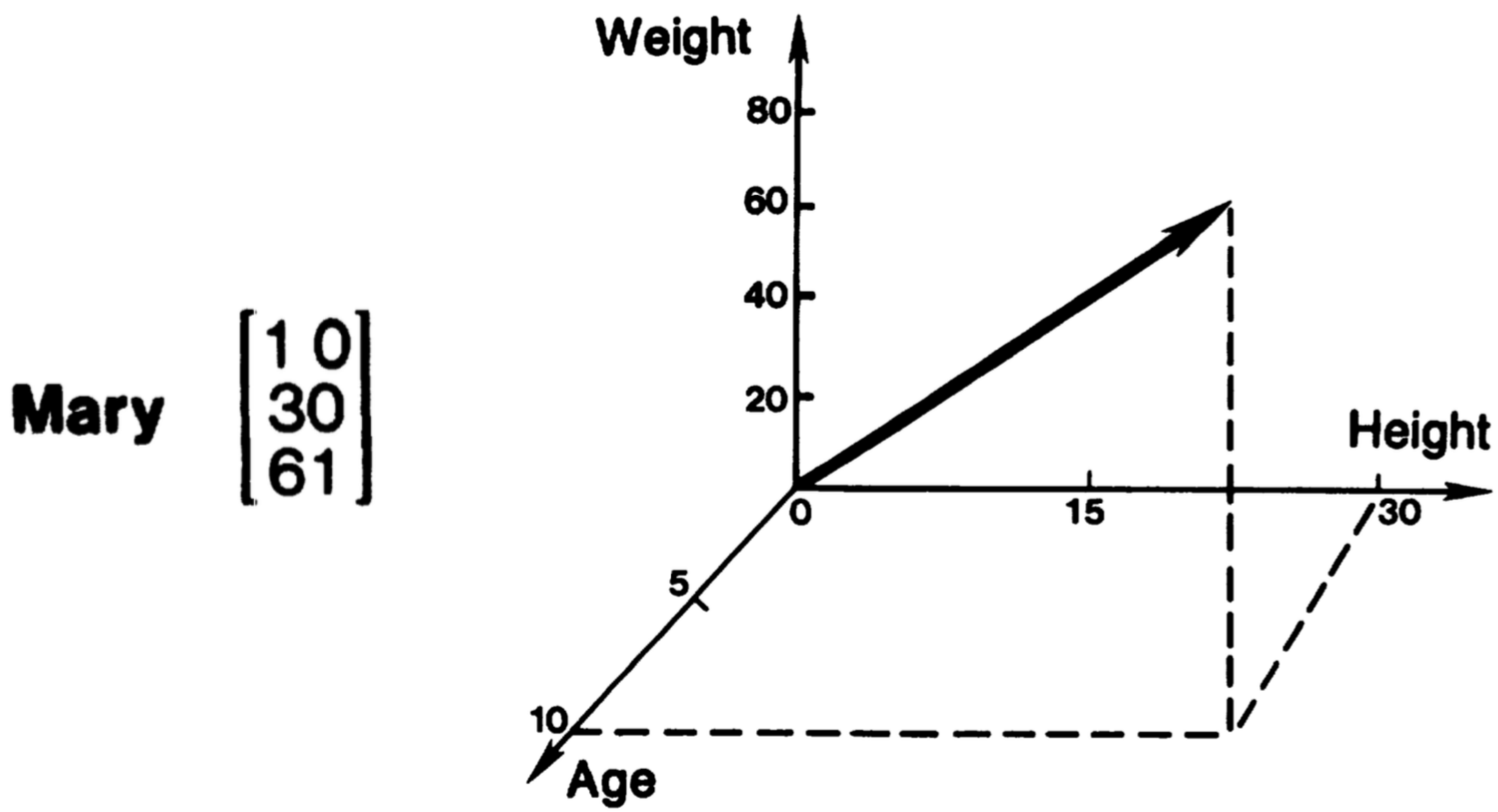
$$\textbf{Joe} \begin{bmatrix} 37 \\ 72 \\ 175 \end{bmatrix} \qquad \textbf{Mary} \begin{bmatrix} 10 \\ 30 \\ 61 \end{bmatrix}$$

$$\textbf{Carol} \begin{bmatrix} 25 \\ 65 \\ 121 \end{bmatrix} \qquad \textbf{Brad} \begin{bmatrix} 66 \\ 67 \\ 155 \end{bmatrix}$$

- Each of these vectors has three **components**

Visualising vectors

- We can neatly visualise vectors with no more than three components:

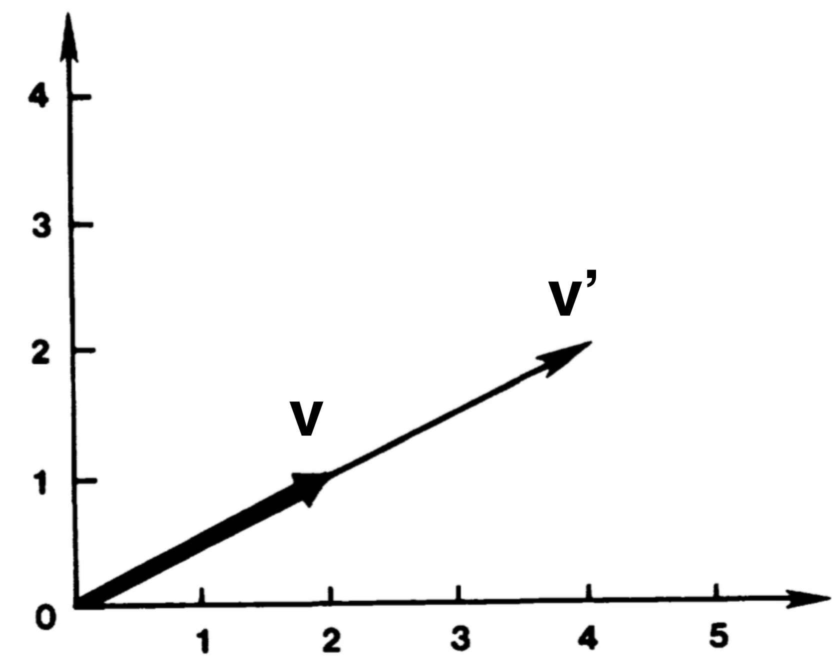


- This will prove helpful in developing a **geometrical intuition** about vectors (but everything we discuss extends to any number of components)

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 **v** by a positive scalar **s** yields a vector **v'** that points in the same direction as **v**, but that is longer or shorter by magnitude **s**

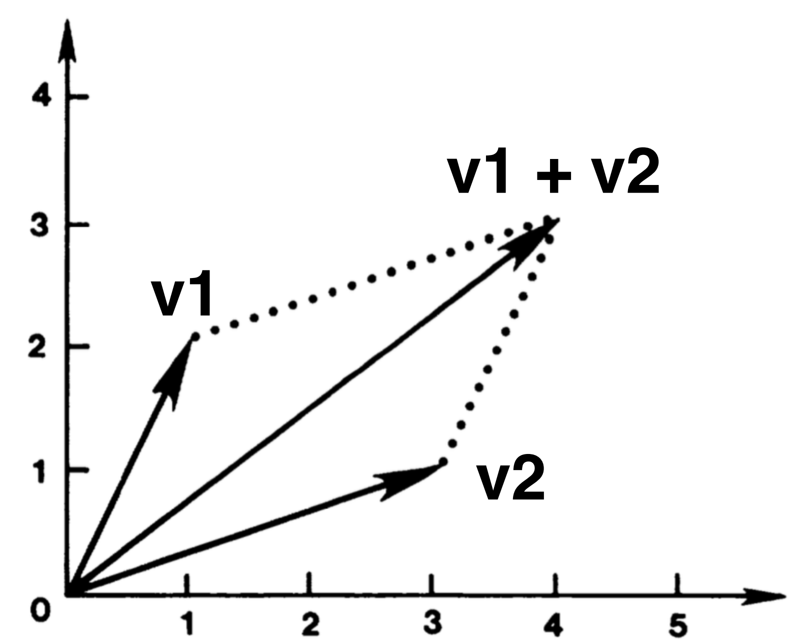
  - Multiplying **v** by a negative scalar, also yields a lengthened or shortened vector **v'**, but this time one pointing in the opposite direction of **v**

- 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**:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$



  - **v1 + v2** lies in between **v1** and **v2**, and forms the diagonal of a **parallelogram** with **v1** and **v2**

- Vector addition is **associative**: **(v1 + v2) + v3 = v1 + (v2 + v3)**

- Vector addition is **commutative**: **v3 + v2 + v1 = v1 + v2 + v3**

Example: Addition and Scalar multiplication

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

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

- In vector notation:

$$\mathbf{u} = \frac{1}{4}\left(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4\right)$$

  - Vector **u** is a **linear combination** of vectors **v1**, **v2**, **v3**, and **v4**, and contains the averages of their components

  - Scalar multiplication is **distributive: ¼v1 + ¼v2 + ¼v3 + ¼v4 = ¼(v1 + v2 + v3 + v4)**

Linear combinations

- A vector **v** is a linear combination of vectors **v1**,**v2**,…,**v*n*** if there are scalars **c1**,**c2**,…,**c*n*** such that:

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

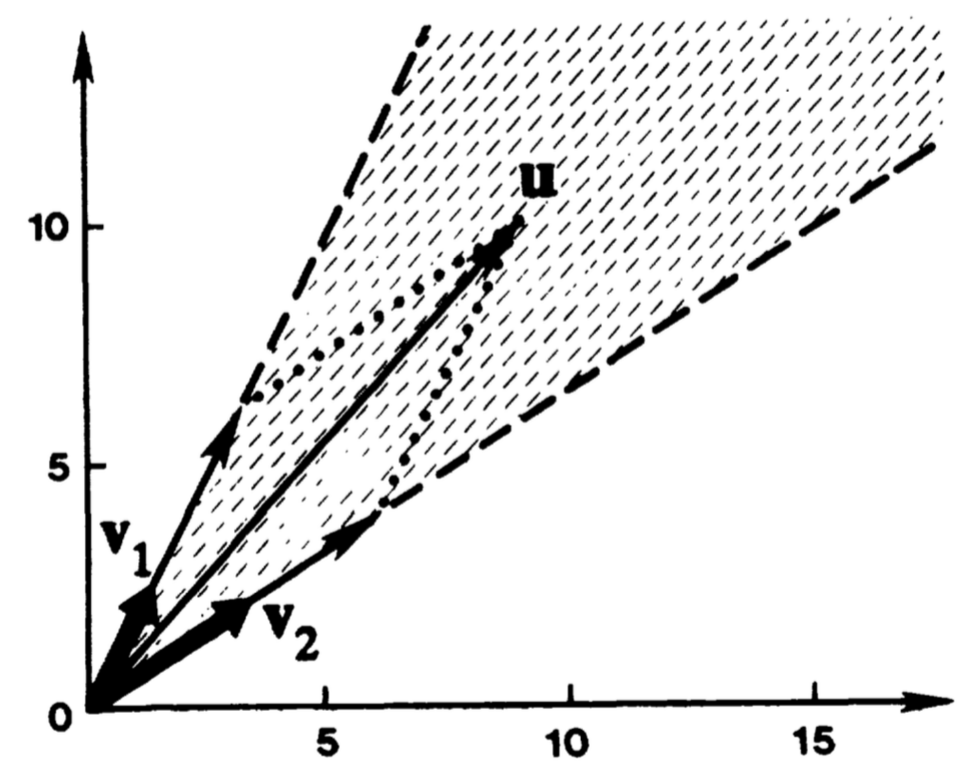
- Example: $\mathbf{u} = \begin{bmatrix} 9 \\ 10 \end{bmatrix}$ is a linear combination of $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

u = 3 * v1 + 2 * v2

- We effectively find scalars to adjust **v1** and **v2** to form a parallelogram with **u**

- Using positive scalars, any vector in the shaded area can be constructed.

- Using positive and negative scalars, any vector in the plane can be constructed



Linear combinations (cont'd)

- The set of all linear combinations of **v1**,**v2**,…,**v*n*** is said to be the set **spanned** by **v1**,**v2**,…,**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 $\mathbf{v} = \begin{bmatrix}a\\b\\c\end{bmatrix}$ can be written as: $\mathbf{v} = a\begin{bmatrix}1\\0\\0\end{bmatrix} + b\begin{bmatrix}0\\1\\0\end{bmatrix} + c\begin{bmatrix}0\\0\\1\end{bmatrix}$

  - 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 $\mathbf{v}_i$ 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 $\mathbf{v}_i$ 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 **v** and **w** with the same number of components, we obtain their **inner product v · w**:

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

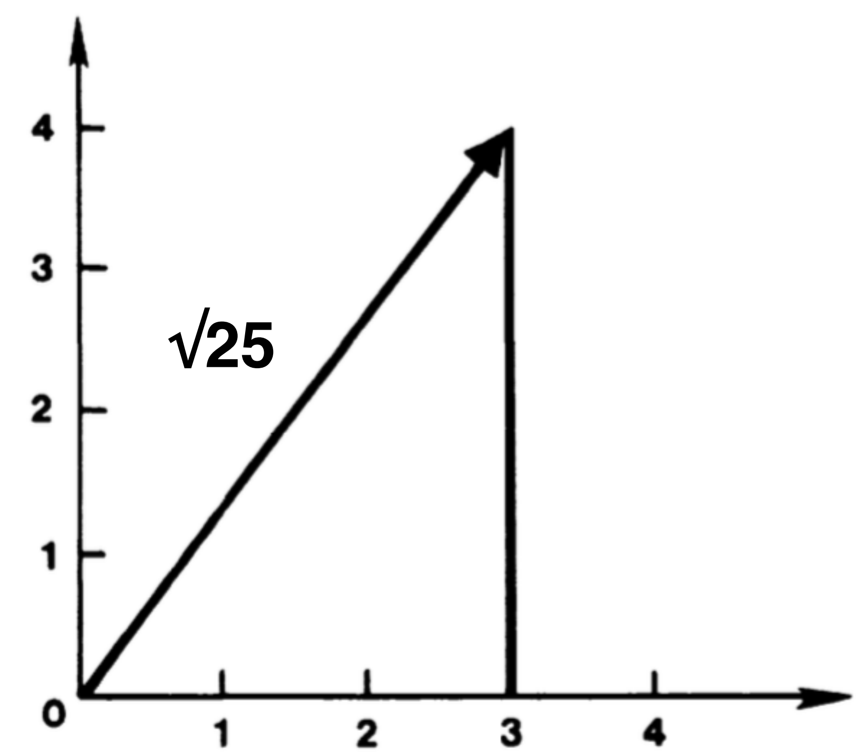
$$\mathbf{v} \cdot \mathbf{w} = (3 \cdot 1) + (-1 \cdot 2) + (2 \cdot 1) = 3.$$

- 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**:

$$\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \mathbf{v} \cdot \mathbf{v} = 3^2 + 4^2 = 25.$$



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

$$\|\mathbf{v}\| = (\mathbf{v} \cdot \mathbf{v})^{1/2} \quad = \sqrt{(\mathbf{v} \cdot \mathbf{v})}$$

- This definition includes our intuitions about length:

$$\|c\mathbf{v}\| = |c| \, \|\mathbf{v}\| \qquad\qquad \|\mathbf{v}_1 + \mathbf{v}_2\| \leqslant \|\mathbf{v}_1\| + \|\mathbf{v}_2\|$$

(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):

$$\cos\theta = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\,\|\mathbf{w}\|} \quad = \quad \cos\theta = \frac{\sum_{i=1}^{n} v_i\, w_i}{(\sum_{i=1}^{n} v_i^2)^{1/2}(\sum_{i=1}^{n} w_i^2)^{1/2}}$$

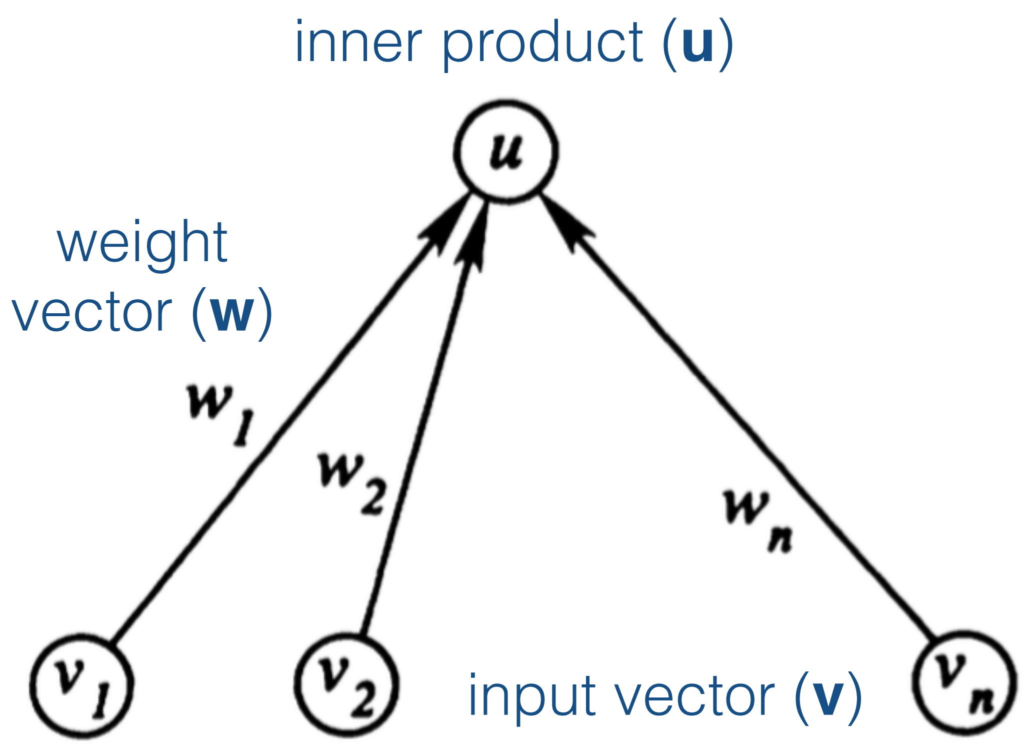
- Example: The angle between $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is found through:

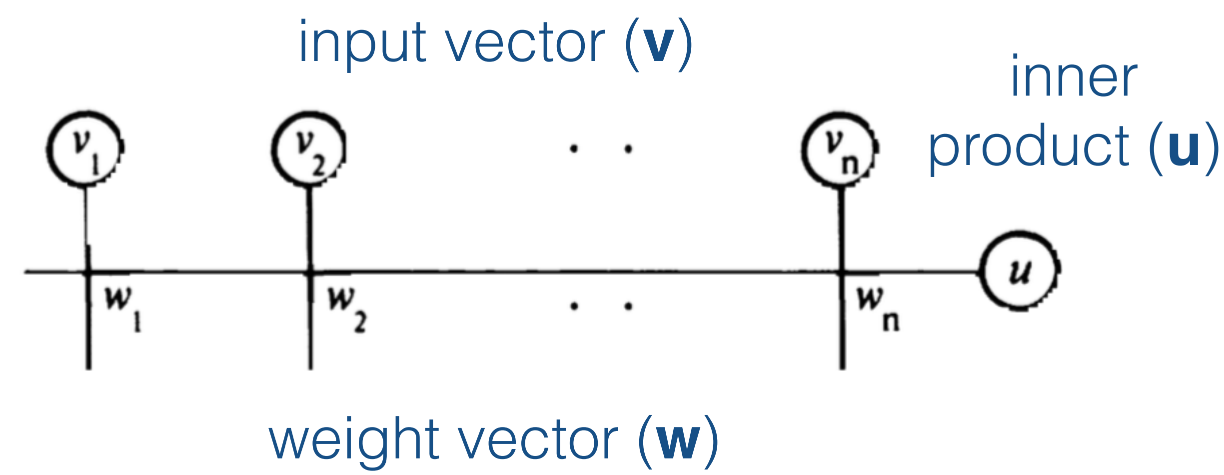
$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 1 \qquad \|\mathbf{v}_1\| = 1 \qquad \|\mathbf{v}_2\| = \sqrt{2} \qquad \cos\theta = \frac{1}{1\cdot\sqrt{2}} = 0.707$$

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

- If **cos P = 0 (=90°),** two vectors are **orthogonal** (at right angles) to one another

Connectionism: A single unit





- The activation of unit **u** computes the inner product of **w** and **v**: **u = w · v**

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

- 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**

Matrices

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

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

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

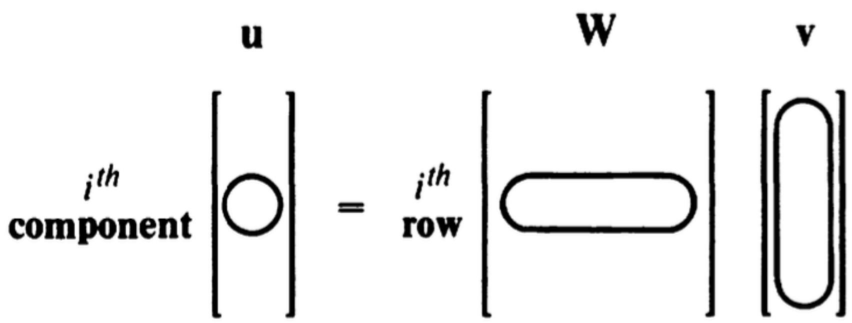
$$3\,\mathbf{M} = 3 \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 12 & 15 \\ 3 & 0 & 3 \end{bmatrix}$$

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

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

Multiplying a Vector by a Matrix

- A **m x 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**:



$$\mathbf{u} = \mathbf{W}\mathbf{v} = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 3\cdot 1 + 4\cdot 0 + 5\cdot 2 \\ 1\cdot 1 + 0\cdot 0 + 1\cdot 2 \end{bmatrix} = \begin{bmatrix} 13 \\ 3 \end{bmatrix}$$

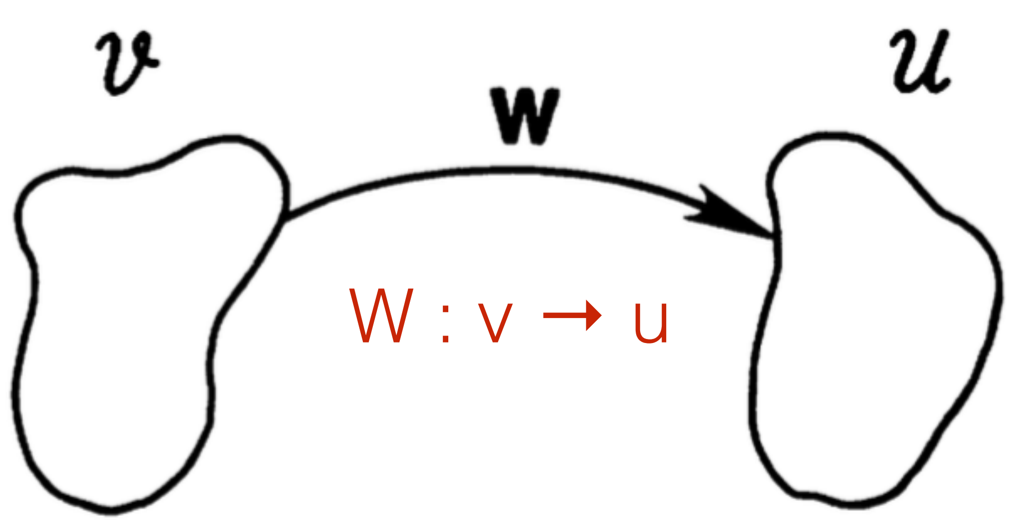
- From another perspective, **u** is linear combination of the **column vectors w*j*** of **W** with the components of **v** as coefficients:

$$\begin{array}{cccc} \mathbf{W} & \mathbf{v} & & \mathbf{u} \\ \begin{bmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ | & & | \end{bmatrix} & \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} & = & \begin{bmatrix} & | & \\ v_1\mathbf{w}_1 + & \cdots & + v_n\mathbf{w}_n \\ & | & \end{bmatrix} \end{array}$$

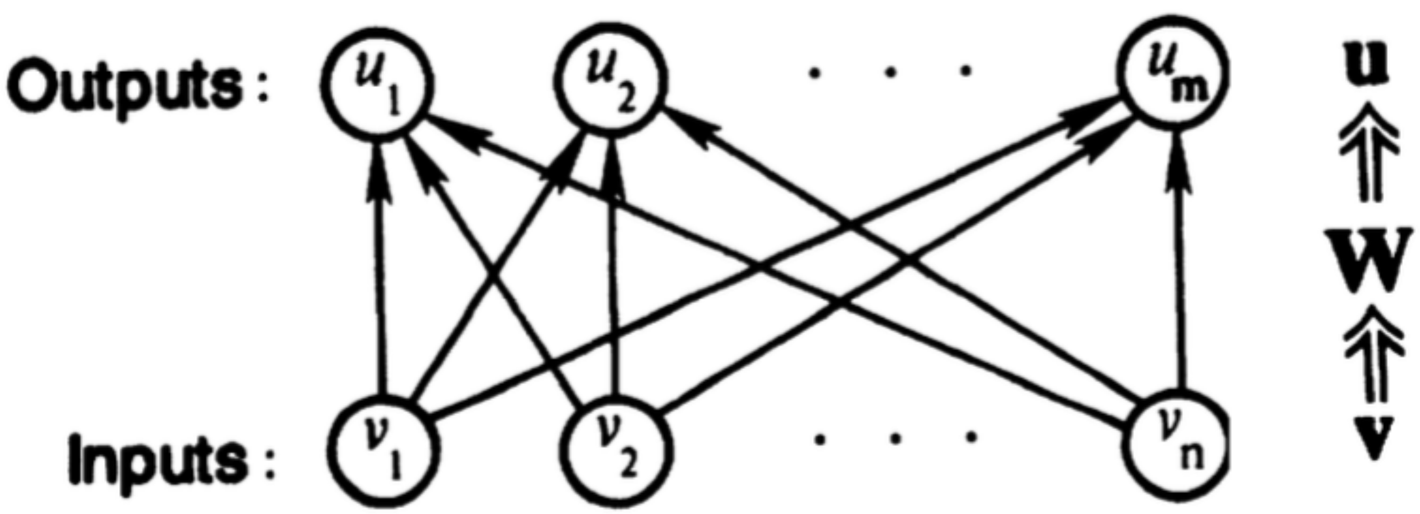
$$\mathbf{u} = v_1\mathbf{w}_1 + v_2\mathbf{w}_2 + v_3\mathbf{w}_3 = \left[1\begin{bmatrix} 3 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 4 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 5 \\ 1 \end{bmatrix} \right] = \begin{bmatrix} 13 \\ 3 \end{bmatrix}$$

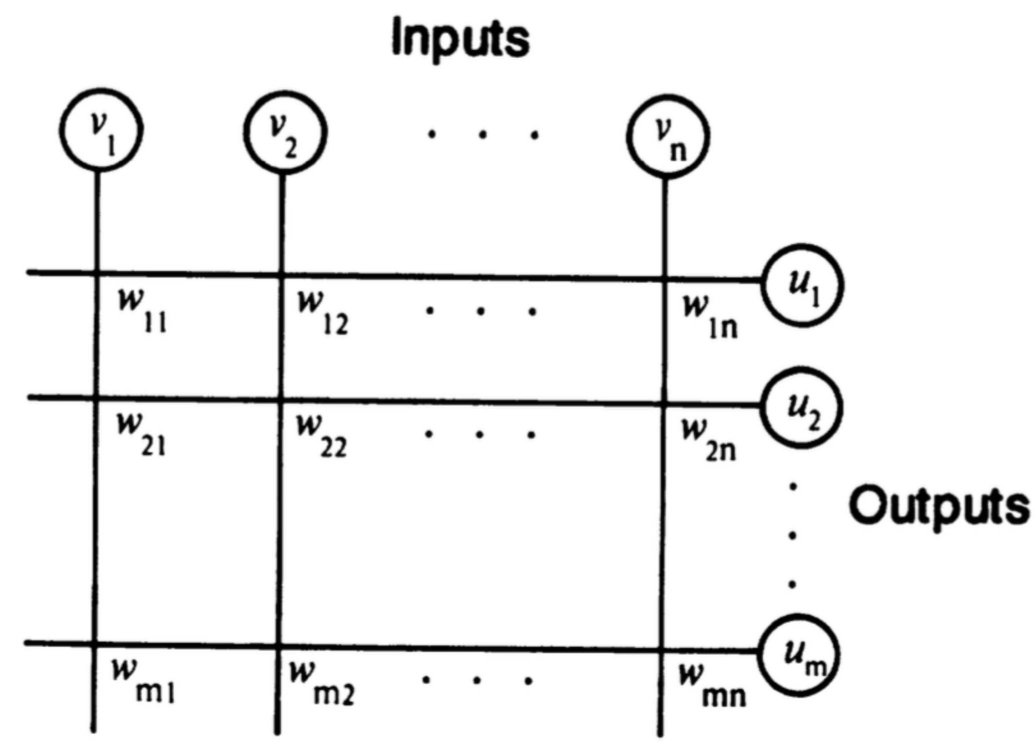
Vector-Matrix Multiplication as a Function

- 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

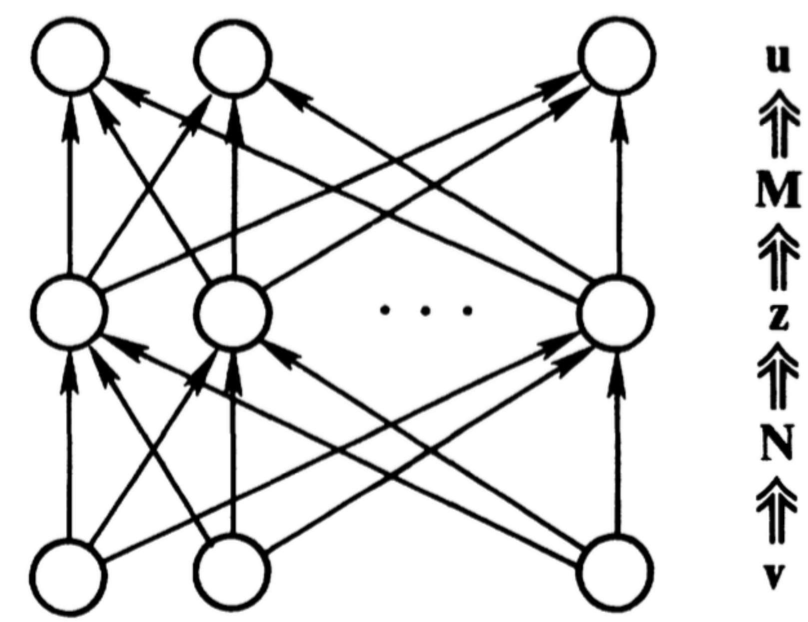




- Each unit **u*i*** has its own weight vector **w*i***, and the activation of unit **u*i*** is the inner product of **w*i*** and input vector **v**: **u*i*** = **w*i*** · **v**
- If we define a matrix **W** that has weight vectors **w*i*** as its row vectors, the activation of all units **u*i*** is 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**:

$$\underset{\mathbf{M}}{\begin{bmatrix} \\ \mathbf{M} \\ \\ \end{bmatrix}} \underset{\mathbf{N}}{\begin{bmatrix} | & | & & | \\ \mathbf{n}_1 & \mathbf{n}_2 & \cdots & \mathbf{n}_s \\ | & | & & | \end{bmatrix}} = \underset{\mathbf{P}}{\begin{bmatrix} | & | & & | \\ \mathbf{Mn}_1 & \mathbf{Mn}_2 & \cdots & \mathbf{Mn}_s \\ | & | & & | \end{bmatrix}}$$

inner product

$$\begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \boxed{(3+8-5)} & (6+0+5) \\ (1+0-1) & (2+0+1) \\ (0+2-2) & (0+0+2) \end{bmatrix} = \begin{bmatrix} 6 & 11 \\ 0 & 3 \\ 0 & 2 \end{bmatrix}$$

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 $\mathbf{y} = \boldsymbol{f}(\mathbf{x})$ describes a **linear system**, if for any inputs **x1** and **x2**, the following equations hold:

    - $\boldsymbol{f}(\mathbf{cx}) = \mathbf{c}\boldsymbol{f}(\mathbf{x})$

    - $\boldsymbol{f}(\mathbf{x1} + \mathbf{x2}) = \boldsymbol{f}(\mathbf{x1}) + \boldsymbol{f}(\mathbf{x2})$

- **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 …

$$\begin{bmatrix} \cos 90^\circ & \sin 90^\circ \\ -\sin 90^\circ & \cos 90^\circ \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a_2 \\ a_1 \end{bmatrix}$$