General (Artificial) Neural Networks

Basic graph theoretic notions

A (directed) **graph** is a pair $G = (V, E)$ consisting of a (finite) set V of **vertices** or **nodes** and a (finite) set $E \subseteq V \times V$ of **edges**.

We call an edge $e = (u, v) \in E$ **directed** from vertex u to vertex v.

Let $G = (V, E)$ be a (directed) graph and $u \in V$ a vertex.
Then the vertices of the set

$$\mathrm{pred}(u) = \{v \in V \mid (v, u) \in E\}$$

are called the **predecessors** of the vertex u
and the vertices of the set

$$\mathrm{succ}(u) = \{v \in V \mid (u, v) \in E\}$$

are called the **successors** of the vertex u.

General Neural Networks

General definition of a neural network

An (artificial) **neural network** is a (directed) graph $G = (U, C)$, whose vertices $u \in U$ are called **neurons** or **units** and whose edges $c \in C$ are called **connections**.

The set U of vertices is partitioned into

- the set U_{in} of **input neurons**,
- the set U_{out} of **output neurons**, and
- the set U_{hidden} of **hidden neurons**.

It is

$$U = U_{\text{in}} \cup U_{\text{out}} \cup U_{\text{hidden}},$$

$$U_{\text{in}} \neq \emptyset, \qquad U_{\text{out}} \neq \emptyset, \qquad U_{\text{hidden}} \cap (U_{\text{in}} \cup U_{\text{out}}) = \emptyset.$$

General Neural Networks

Each connection $(v, u) \in C$ possesses a **weight** w_{uv} and
each neuron $u \in U$ possesses three (real-valued) state variables:

- the **network input** net_u,
- the **activation** act_u, and
- the **output** out_u.

Each input neuron $u \in U_{\mathrm{in}}$ also possesses a fourth (real-valued) state variable,

- the **external input** ext_u.

Furthermore, each neuron $u \in U$ possesses three functions:

- the **network input function** $f_{\mathrm{net}}^{(u)} : \mathbb{R}^{2|\,\mathrm{pred}(u)|+\kappa_1(u)} \to \mathbb{R}$,
- the **activation function** $f_{\mathrm{act}}^{(u)} : \mathbb{R}^{\kappa_2(u)} \to \mathbb{R}$, and
- the **output function** $f_{\mathrm{out}}^{(u)} : \mathbb{R} \to \mathbb{R}$,

which are used to compute the values of the state variables.

General Neural Networks

Types of (artificial) neural networks:

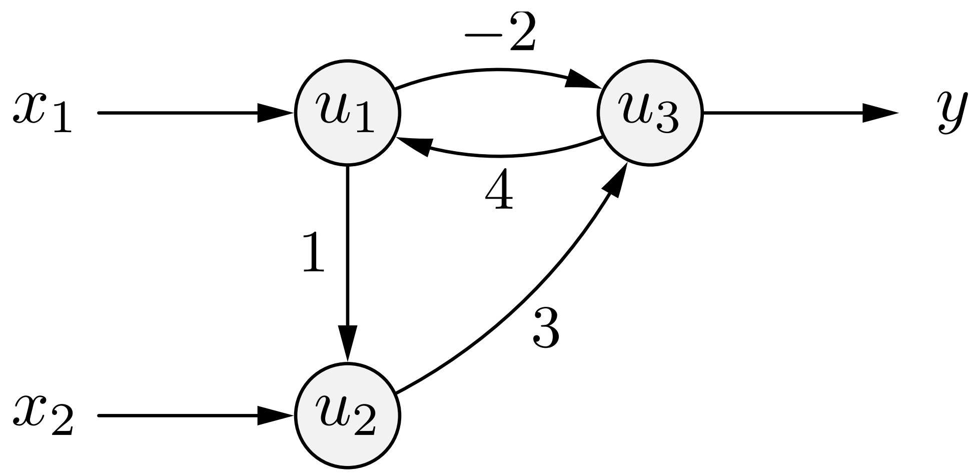
- If the graph of a neural network is **acyclic**, it is called a **feed-forward network**.
- If the graph of a neural network contains **cycles** (backward connections), it is called a **recurrent network**.

Representation of the connection weights as a matrix:

$$\begin{array}{cc} \begin{array}{cccc} u_1 & u_2 & \dots & u_r \end{array} & \\ \left(\begin{array}{cccc} w_{u_1 u_1} & w_{u_1 u_2} & \dots & w_{u_1 u_r} \\ w_{u_2 u_1} & w_{u_2 u_2} & & w_{u_2 u_r} \\ \vdots & & & \vdots \\ w_{u_r u_1} & w_{u_r u_2} & \dots & w_{u_r u_r} \end{array} \right) & \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_r \end{array} \end{array}$$

General Neural Networks: Example

A simple recurrent neural network

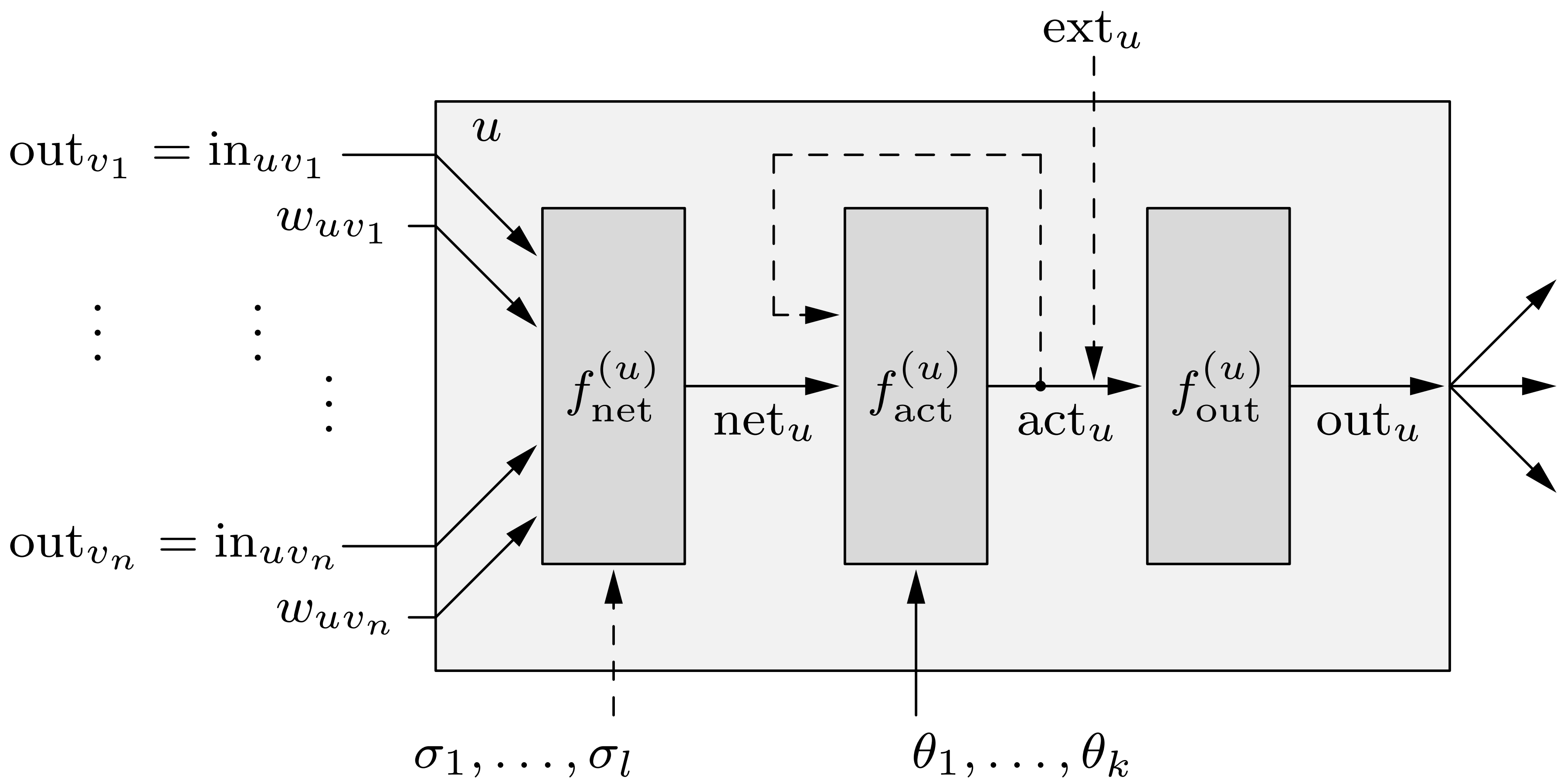


Weight matrix of this network

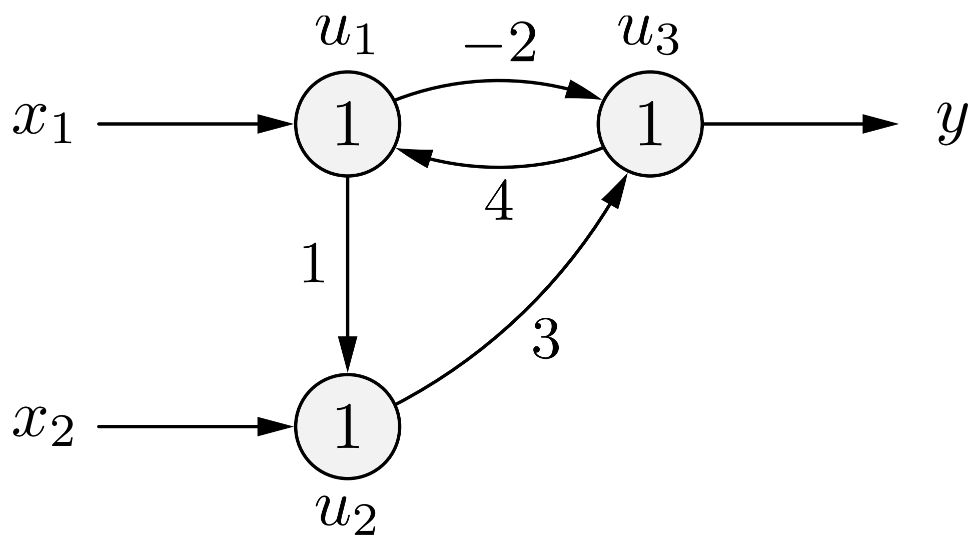
$$
\begin{array}{c}
\begin{array}{ccc} u_1 & u_2 & u_3 \end{array} \\
\left(\begin{array}{rrr} 0 & 0 & 4 \\ 1 & 0 & 0 \\ -2 & 3 & 0 \end{array}\right)
\end{array}
\begin{array}{c} \\ u_1 \\ u_2 \\ u_3 \end{array}
$$

Structure of a Generalized Neuron

A generalized neuron is a simple numeric processor



General Neural Networks: Example



$$f_{\text{net}}^{(u)}(\vec{w}_u, \vec{\text{in}}_u) = \textstyle\sum_{v\in\text{pred}(u)} w_{uv}\text{in}_{uv} = \sum_{v\in\text{pred}(u)} w_{uv}\,\text{out}_v$$

$$f_{\text{act}}^{(u)}(\text{net}_u, \theta) = \begin{cases} 1, & \text{if} \quad \text{net}_u \geq \theta, \\ 0, & \text{otherwise.} \end{cases}$$

$$f_{\text{out}}^{(u)}(\text{act}_u) = \text{act}_u$$

General Neural Networks: Example

Updating the activations of the neurons

	u_1	u_2	u_3	
input phase	**1**	**0**	**0**	
work phase	1	0	**0**	$\text{net}_{u_3} = -2 \quad < 1$
	0	0	0	$\text{net}_{u_1} = 0 \quad < 1$
	0	**0**	0	$\text{net}_{u_2} = 0 \quad < 1$
	0	0	**0**	$\text{net}_{u_3} = 0 \quad < 1$
	0	0	0	$\text{net}_{u_1} = 0 \quad < 1$

- Order in which the neurons are updated:
  $u_3, u_1, u_2, u_3, u_1, u_2, u_3, \ldots$
- Input phase: activations/outputs in the initial state.
- Work phase: activations/outputs of the next neuron to update (bold) are computed from the outputs of the other neurons and the weights/threshold.
- A stable state with a unique output is reached.

General Neural Networks: Example

Updating the activations of the neurons

	u_1	u_2	u_3	
input phase	**1**	**0**	**0**	
work phase	1	0	**0**	$\text{net}_{u_3} = -2 \quad < 1$
	1	**1**	0	$\text{net}_{u_2} = 1 \quad \geq 1$
	0	1	0	$\text{net}_{u_1} = 0 \quad < 1$
	0	1	**1**	$\text{net}_{u_3} = 3 \quad \geq 1$
	0	**0**	1	$\text{net}_{u_2} = 0 \quad < 1$
	1	0	1	$\text{net}_{u_1} = 4 \quad \geq 1$
	1	0	**0**	$\text{net}_{u_3} = -2 \quad < 1$

- Order in which the neurons are updated:
  $u_3, u_2, u_1, u_3, u_2, u_1, u_3, \ldots$
- No stable state is reached (oscillation of output).

General Neural Networks: Training

Definition of learning tasks for a neural network

A **fixed learning task** L_{fixed} for a neural network with

- n input neurons $U_{\text{in}} = \{u_1, \ldots, u_n\}$ and
- m output neurons $U_{\text{out}} = \{v_1, \ldots, v_m\}$,

is a set of **training patterns** $l = (\vec{\imath}^{\,(l)}, \vec{o}^{\,(l)})$, each consisting of

- an **input vector** $\vec{\imath}^{\,(l)} = (\,\text{ext}_{u_1}^{(l)}, \ldots, \text{ext}_{u_n}^{(l)}\,)$ and
- an **output vector** $\vec{o}^{\,(l)} = (o_{v_1}^{(l)}, \ldots, o_{v_m}^{(l)})$.

A fixed learning task is solved, if for all training patterns $l \in L_{\text{fixed}}$ the neural network computes from the external inputs contained in the input vector $\vec{\imath}^{\,(l)}$ of a training pattern l the outputs contained in the corresponding output vector $\vec{o}^{\,(l)}$.

General Neural Networks: Training

Solving a fixed learning task: Error definition

- Measure how well a neural network solves a given fixed learning task.
- Compute differences between desired and actual outputs.
- Do not sum differences directly in order to avoid errors canceling each other.
- Square has favorable properties for deriving the adaptation rules.

$$e = \sum_{l \in L_{\text{fixed}}} e^{(l)} = \sum_{v \in U_{\text{out}}} e_v = \sum_{l \in L_{\text{fixed}}} \sum_{v \in U_{\text{out}}} e_v^{(l)},$$

$$\text{where} \qquad e_v^{(l)} = \left(o_v^{(l)} - \text{out}_v^{(l)} \right)^2$$

General Neural Networks: Training

Definition of learning tasks for a neural network

A **free learning task** L_{free} for a neural network with

- n input neurons $U_{\text{in}} = \{u_1, \ldots, u_n\}$,

is a set of **training patterns** $l = (\vec{\imath}^{\,(l)})$, each consisting of

- an **input vector** $\vec{\imath}^{\,(l)} = (\,\text{ext}_{u_1}^{(l)}, \ldots, \text{ext}_{u_n}^{(l)}\,)$.

Properties:

- There is no desired output for the training patterns.
- Outputs can be chosen freely by the training method.
- Solution idea: **Similar inputs should lead to similar outputs.**
  (clustering of input vectors)

Normalization of the input vectors

- Compute expected value and (corrected) standard deviation for each input:

$$\mu_k = \frac{1}{|L|} \sum_{l \in L} \mathrm{ext}_{u_k}^{(l)} \qquad \text{and} \qquad \sigma_k = \sqrt{\frac{1}{|L| \mathbf{-1}} \sum_{l \in L} \left(\mathrm{ext}_{u_k}^{(l)} - \mu_k \right)^2},$$

- Normalize the input vectors to expected value 0 and standard deviation 1:

$$\mathrm{ext}_{u_k}^{(l)(\mathrm{new})} = \frac{\mathrm{ext}_{u_k}^{(l)(\mathrm{old})} - \mu_k}{\sigma_k}$$

- Such a normalization avoids unit and scaling problems.
  It is also known as **z-scaling** or **z-score standardization**.