

Definition of Neural Networks

Samuel Lippl

April 15, 2018

Contents

1 Structure of Neural Networks	1
1.1 General Structure	1
1.2 Additive neural networks	2
1.3 Predictive Coding	3
Open Problems: Overview	

Neural networks are very general notions that encompass many different varieties. While there have been attempts to restrict make restrictions that are derived from their purpose (see, for instance, [2]), I argue that these restrictions make the definition more complicated to prove but not more powerful. Therefore, I provide a definition in the sense of [4]. However, I have attempted to make the notation more compact for my purposes and I have worked out the different aspects of neural networks more clearly. Finally, my definition is broader than Rojas' definition in [4].

1 Structure of Neural Networks

1.1 General Structure

1.1 DEFINITION (ELEMENTARY STRUCTURE OF NEURAL NETWORKS). A *neural network* is defined by its *elementary structure* which is a directed graph $(\mathcal{V}, \mathcal{C})$, $\mathcal{C} \subseteq \mathcal{V} \times \mathcal{V}$, that is *completely connected*. We call the nodes in \mathcal{V} *Processing Units* (PU) und the edges in \mathcal{C} *Unit Connections* (UC). We will leave open the possibility to associate a class attribute with PUs or UCs that may restrict different network aspects (see, for instance, example ??).

1.2 NOTATION. We associate \mathcal{C} with the relations $\rightarrow_{\mathcal{C}}$ and $--_{\mathcal{C}}$ on \mathcal{V} , i. e.

$$\begin{aligned} v \rightarrow w &\equiv (v, w) \in \mathcal{C} \equiv v \leftarrow w \\ v -- w &\equiv v \rightarrow w \vee w \rightarrow v \end{aligned} \tag{1}$$

Where the context is clear, we write \rightarrow and $--$ instead of $\rightarrow_{\mathcal{C}}$ and $--_{\mathcal{C}}$.

We now complete Definition 1.1:

1.3 DEFINITION (COMPLETE CONNECTEDNESS). We call a graph completely connected if and only if

$$\forall_{v, w \in \mathcal{V}} \exists_{k \in \mathbb{N}^+} \exists_{u_1, \dots, u_k \in \mathcal{V}} u_1 = v \wedge \forall_{i \in \{1, \dots, k-1\}} u_i -- u_{i+1} \wedge u_k = w \tag{2}$$

1.4 The restriction that a neural network must be completely connected is sensible. It prevents networks from incorporating PUs or sets of PUs that are independent from the rest of the network. A single isolated PU may simply be ignored and a set of isolated PUs may be considered as a separate network¹.

(Where does this restriction come to use?)

1.5 NOTATION (PRECEDING AND SUCCEEDING UNITS). If $u \in \mathcal{V}$ is a PUs we define the set of *preceding* PUs

$$\mathcal{C}_u := \{v \in \mathcal{V} | v \rightarrow u\} \quad (3)$$

and the set of *succeeding* PUs

$$\mathcal{C}^u := \{w \in \mathcal{V} | u \leftarrow w\} \quad (4)$$

1.6 DEFINITION (ACYCLICAL NETWORKS). The network $(\mathcal{V}, \mathcal{C})$ is called *acyclical* or *feed-forward* if and only if the relation \rightarrow is acyclical.

1.7 The elementary structure of a neural network only specifies how the PUs are connected but not how they use these connections. With respect to this task, we have to differentiate between the *network assumptions* that cannot be change during its applications and its *parameters* that can be changed during learning (see section ??). While specific values of the latter should not be considered as part of the network, the values they may take are as relevant as the way they parametrize the network.

1.8 DEFINITION (FUNCTIONAL STRUCTURE). We define the *parameter space* Θ and the *integrative functions* $\mathcal{G} = (g_u)_{u \in \mathcal{V}}$ where

$$g_u : \mathbb{R}^{\mathcal{C}_u} \times \Theta \rightarrow \mathbb{R} \quad (x, \theta) \mapsto g_u(x|\theta) \quad (5)$$

We call (\mathcal{G}, Θ) the *functional structure* of the neural network and $(\mathcal{V}, \mathcal{C}, \mathcal{G}, \Theta)$ its *structure*.

1.9 I will note that we have not associated these parameters with a specific PU or UC. The main reason for this is the possibility of shared parameters, see definition ?. I will also note that this chapter is actually not concerned with how the network generates its values which will be of concern in chapter ?. The reader may however have inferred from (5) that I expect the PUs to take real values. This is only because there has been no reason to generalize this definition. The relevant concepts can easily be generalized to more general spaces, however.

1.2 Additive neural networks

1.10 DEFINITION (ADDITIVE NEURAL NETWORK). An *additive neural network* is defined by certain restrictions on the functional structure. The parameter space is now $\Theta \subseteq \mathbb{R}^{\mathcal{C}}$ and we define a set of *unit functions* $\mathcal{F} = (f_u)_{u \in \mathcal{V}}$ where

$$f_u : \mathbb{R} \rightarrow \mathbb{R} \quad (6)$$

Then, the integrative functions are defined by

$$g_u(x|\theta) := \sum_{v \in \mathcal{C}_u} \theta_{(v,u)} f_v(x_v) \quad (7)$$

¹ Of course, a single PU could also be considered as a neural network.

By defining

$$\theta_u^v \equiv \theta_{(v,u)} \quad \theta^v \equiv (\theta_u^v)_{u \in \mathcal{C}^v} \quad \theta_u \equiv (\theta_u^v)_{v \in \mathcal{C}_u} \quad (8)$$

$$f_V : \mathbb{R}^V \rightarrow \mathbb{R}^V \quad f_V(x) := (f_v(x_v))_{v \in V} \quad V \subseteq \mathcal{V} \quad (9)$$

and using the ordinary definition of matrix multiplication we can notate

$$g_u(x|\theta) = \theta_u^T f_{\mathcal{C}_u}(x) \quad (10)$$

where θ^T is the transpose of θ .

1.11 In many neural networks all unit functions are provided by a unique function f . Most unit functions map their input to $(-1, 1)$ or $(0, 1)$. Popular choices are the sigmoid function $f(x) = \frac{1}{1+e^{-cx}}$ or the tangens hyperbolicus $f(x) = \tanh(x)$. If we define $f = \text{id}$, the resulting network provides a linear model.

Finally, many neural networks work in layers: they are structured into L layers of PUs where $v_1^{(l)}, \dots, v_{n_i}^{(l)} \in \mathcal{V}$ and $\mathcal{C} = \{(v_i^{(l)}, v_j^{(l+1)}) | l = 0, \dots, L-1, i = 1, \dots, n_l, j = 1, \dots, n_{l+1}\}$ where input is provided to layer 0 and the network produces some kind of output at layer L . It may helpful to use such a structure to visualize certain notions. However, as far as I see it, a general feed-forward architecture provides all the benefits of a layered structure which may be justified by Proposition ??.

1.3 Predictive Coding

1.12 (PREDICTIVE CODING I: THE GENERATIVE MODEL) Predictive Coding serves as a model of how our brain processes sensations to perceptions and is based on [3]. It poses that our brain processes data by inferring hidden *causes* which are, in this case, "simply the states of processes generating sensory data" [1, p. 819].

References

- [1] Karl Friston. "A theory of cortical responses". In: *Philosophical Transactions of the Royal Society B* 360 (2005), pp. 815–836. DOI: [10.1098/rstb.2005.1622](https://doi.org/10.1098/rstb.2005.1622).
- [2] Erkam Guresen and Gulgun Kayakutlu. "Definition of artificial neural networks with comparison to other networks". In: *Procedia Computer Science* 3 (2011), pp. 426–433. DOI: [10.1016/j.procs.2010.12.071](https://doi.org/10.1016/j.procs.2010.12.071).
- [3] D. Mumford. "On the computational architecture of the neocortex - II The role of cortico-cortical loops". In: *Biological Cybernetics* 66.3 (1992), pp. 241–251. ISSN: 03401200. DOI: [10.1007/BF00198477](https://doi.org/10.1007/BF00198477). arXiv: [arXiv:1408.1149](https://arxiv.org/abs/1408.1149).
- [4] Raúl Rojas. *Neural Networks*. Berlin: Springer-Verlag, 1996.