NEURAL NETWORK FORMALISM

(Beta version 1.0)

Daniel Crespin
Facultad de Ciencias
Universidad Central de Venezuela

Abstract

Neural networks are defined using only elementary concepts from set theory, without the usual connectionistic graphs. The neural diagrams are *derived* from these definitions. This approach provides mathematical techniques and insight to develop theory and applications of neural networks.

1.- **Parametric maps.** Let W be a set to be called *weight space*, *parameter space*, *control space* or *space of states*; correspondingly, the elements $w \in W$ will be *weights, parameters, controls* or *states*. A *parametric map* with domain X, range Y and with *parameters in* W is a function $f : W \times X \to Y$. By convention the *input* (*independent variable, point in domain*) of this parametric map is $x \in X$ (instead of $(w, x) \in W \times X$); the *output* (*dependent variable, image point*) is $y \in Y$. Such f can also be regarded as a family of maps $\{f_w\}_{w \in W}$ indexed by W. Parametric maps $f : W \times X \to Y$ will be represented in the following way: An input arrow with input label x on its tail, the parameter w as a label over the input arrow, a circle labeled by f and an output arrow with label y on its tip:

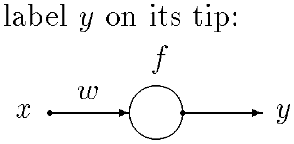

Figure 1. *Parametric map.*

When no confusion arises some or all labels can be omited. The terms *unit*, *neuron*, *node*, *cell*, *processor*, etc. are often used. Any map $f : X \to Y$ can be considered as a parametric map with a singleton parameter space $W = \{w_0\}$.

2.- **Multiple inputs.** Let now f have *multiple input*. This means that $y = f(w, x_1, \ldots, x_m)$. In other words, assume $f : W \times X \to Y$ with $X \subseteq X_1 \times \ldots \times X_m$, the element $x_i \in X_i$ being the *i-th input*. Then, by convention, f will be represented by the graph below. The weight label w is common to all the input arrows.

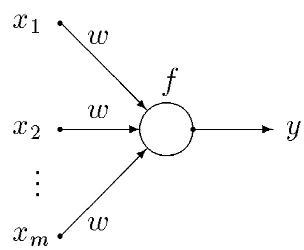

Figure 2. *Parametric map with multiple input.*

Remark: The number of inputs is essentially conventional. For example, several real variables (multiple inputs) can be considered as a single vector variable (single input). Similarly for the number of outputs.

3.- **Paired maps.** There is a special functional form of parametric maps that often appears in practice. Suppose that the parameter set and the input set are contained in products $W \subseteq W_1 \times \ldots \times W_m$, $X \subseteq X_1 \times \ldots \times X_m$; in this situation w_i is the *i-th weight*. The single output parametric map $f : W \times X \to Y$ has *paired weights and inputs* or is a *paired map* if there are functions $\xi_i : W_i \times X_i \to X_i$, $i = 1, \ldots, m$, and $\phi : X_1 \times \ldots \times X_m \to Y$ such that

$$f(w_1, \ldots, w_m, x_1, \ldots, x_m) = \phi(\xi_1(w_1, x_1), \ldots, \xi_m(w_m, x_m))$$

Neither ϕ nor the ξ_i are uniquely determined by f. For example if $W_i = X_i = Y = \mathbf{R}$ and $f(w_1, \ldots, w_m, x_1, \ldots, x_m) = |w_1|^{\alpha_1}|x_1|^{\beta_1} + \ldots + |w_m|^{\alpha_m}|x_m|^{\beta_m}$ (α_i, β_i positive integers) then one can take $\xi_i(w_i, x_i) = |w_i|^{\alpha_i}|x_i|^{\beta_i}$ and $\phi(\xi_1, \ldots, \xi_m) = \xi_1 + \ldots + \xi_m$; but it is also possible to choose the functions $\xi_i(w_i, x_i) = |w_i|^{q\alpha_i}|x_i|^{q\beta_i}$ and $\phi(\xi_1, \ldots, \xi_m) = \xi_1^{1/q} + \ldots + \xi_m^{1/q}$, with any $q > 0$, obtaining the same f. The graph of a paired map is the one shown below.

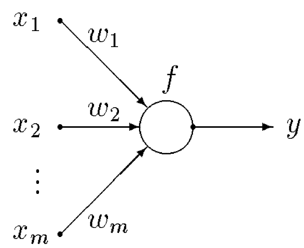

Figure 3. *Neural graph of a paired map.*

There are m inputs and a single output. For each input labeled x_i there is an arrow labeled w_i with tip attached to the circle labeled f. Equivalently, a single $1 \times m$ row-matrix label $[w_i]_{1\times m}$ can be used instead of several individual labels. Heuristically speaking, this matrix 'acts on the left' of the input 'column matrix' with entries $x_1, \ldots, x_m$. Neither ϕ nor the ξ_i's appear explicitly in the diagram.

Single input parametric maps are trivially paired. Since any parametric map can be considered to be single input (remark at end of section 2), any map can be considered a paired map.

4.- **Parametric products.** Assume now that n single input-output maps $f_j : W_j \times X \to Y_j$, $j = 1, \ldots, n$ are given, all with a common input $x \in X$. Their *parametric product*, or simply *product* if danger of confusion is absent, is the map $\widehat{\prod}_{j=1}^{n} f_j = f_1 \widehat{\times} \cdots \widehat{\times} f_n : W_1 \times \cdots \times W_n \times X \to Y_1 \times \cdots \times Y_n$ defined as $f_1 \widehat{\times} \cdots \widehat{\times} f_n(w_1, \ldots, w_n, x) = (f_1(w_1, x), \ldots, f_n(w_n, x))$ or equivalently, $f_{(w_1,\ldots,w_n)} = f_{w_1} \times \cdots \times f_{w_n}$. This has input $x \in X$, output $y = (y_1, \ldots, y_n) \in Y_1 \times \cdots \times Y_n$, parameters $w = (w_1, \ldots, w_n) \in W_1 \times \cdots \times W_n$ and the maps f_j are the *parametric factors*. A parametric product will be also called *layer*, see section 7. They are represented as in Figure 4 below.

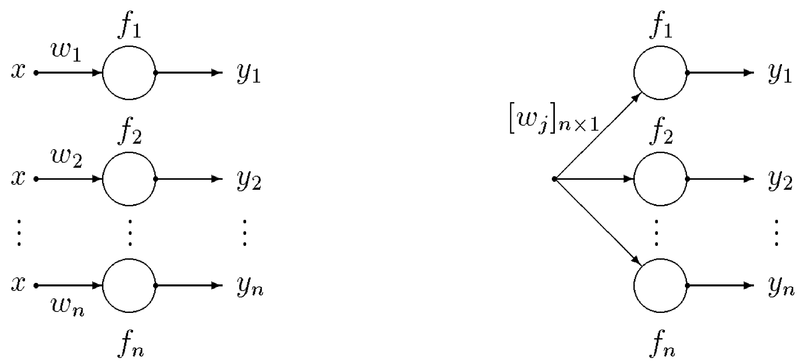

Figure 4. *Single input parametric factors and their parametric product.*

The label on the right is a column matrix $\mathbf{w} = [w_j]_{n\times 1}$ indicating weight w_j for the *j-th* arrow. Alternatively, each input arrow can carry its own label w_j. In case the parametric factors f_j have common input an m-tuple, that is, if $X \subseteq X_1 \times \ldots \times X_m$, then the parametric product can be represented as in Figure 5.

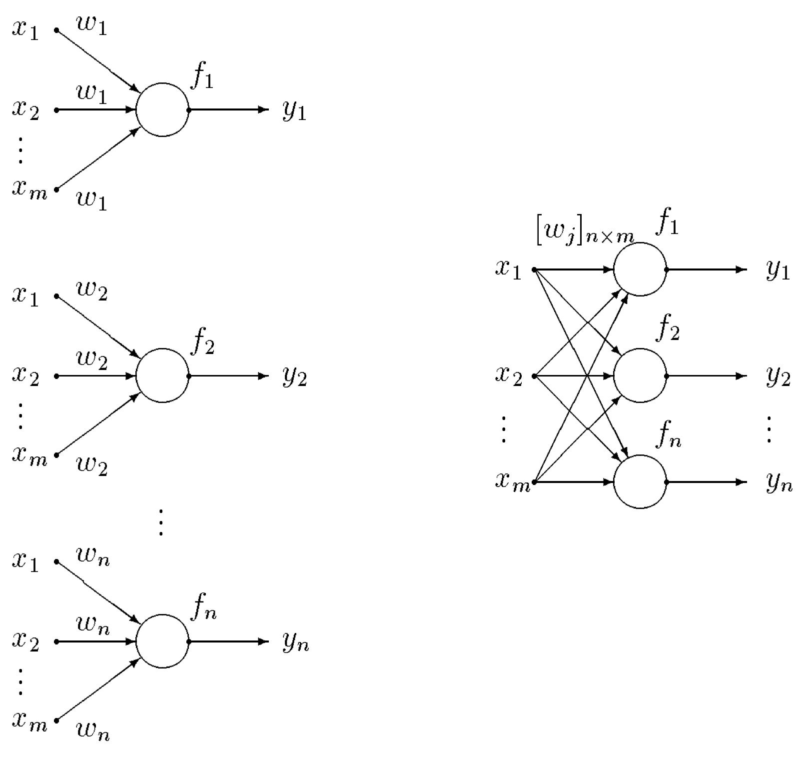

Figure 5. *Multiple input parametric factors and their product.*

The matrix label is a matrix with n rows and m columns and the notation $[w_i]_{n\times m}$ indicates that the rows are constant, $w_{ij} = w_i$. Alternatively, each arrow entering the circle f_j can carry label w_j.

If furthermore each parametric factor is a paired map then there is a similar graph with a different matrix label as shown in Figure 6.

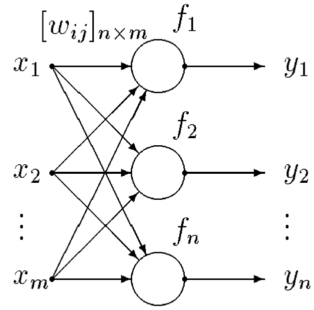

Figure 6. *Product of multiple input paired maps.*

The label is a $m \times n$ matrix $[w_{ij}]_{n\times m}$ indicating that the arrow from input x_i to circle f_j carries label w_{ij}.

5.- **Parametric compositions of single-input maps.** Let $f^k : W^k \times X^k \to X^{k+1}, k = 1, \dots, p$ be parametric maps. Their *parametric composition* is the map $f = \widehat{\bigcirc}_{k=1}^{p} f^k = f^p \widehat{\circ} f^{p-1} \widehat{\circ} \dots \widehat{\circ} f^1 : W^1 \times \dots \times W^p \times X^1 \to X^{p+1}$ defined recursively by the expression $(\widehat{\bigcirc}_{k=1}^{p} f^k)(w^1, \dots, w^{p-1}, w^p, x^1) = f^p(w^p, (\widehat{\bigcirc}_{k=1}^{p-1} f^k)(w^1, \dots, w^{p-1}, x^1))$ Equivalently $f_{(w^1,\dots,w^{p-1},w^p)} = f^p_{w^p} \widehat{\circ} f^{p-1}_{w^{p-1}} \widehat{\circ} \cdots \widehat{\circ} f^1_{w^1}$. This composition will be represented as shown in Figure 7.

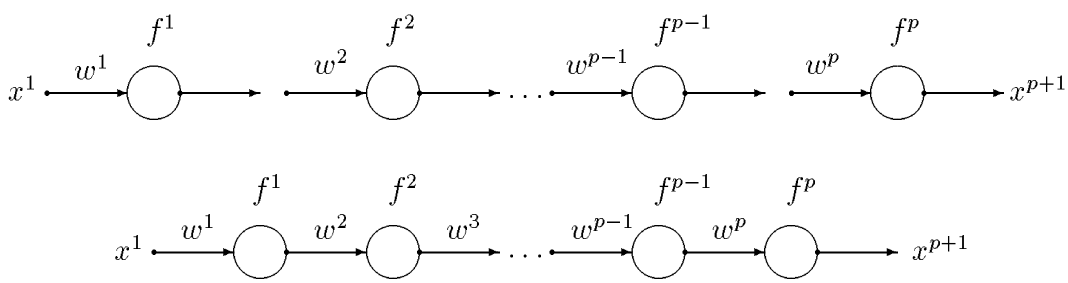

Figure 7. *Parametric composition of parametric maps.*

6.- **Parametric composition of multiple input products.** The hypothesis can now be made that each layer f^k is a multiple input product of parametric maps. Hence add the assumptions that each X^k is a product of n_k sets, that is, $X^k = X_1^k \times \cdots \times X_{n_k}^k, k = 1, \ldots, p+1$, that each W^k is a product of n_{k+1} sets $W^k = W_1^k \times \cdots \times W_{n_{k+1}}^k$, $k = 1, \ldots, p$ and that the maps $f^k : W^k \times X^k \to X^{k+1}$ are multiple input parametric products $f^k = \widehat{\prod}_{i_{k+1}=1}^{n_{k+1}} f_{i_{k+1}}^k : W_1^k \times \cdots \times W_{n_{k+1}}^k \times X_1^k \times \cdots \times X_{n_k}^k \to X_1^{k+1} \times \cdots \times X_{n_{k+1}}^{k+1}$. Such composition $f = f^p \hat{\circ} f^{p-1} \hat{\circ} \ldots \hat{\circ} f^1$ is represented in Figure 8.

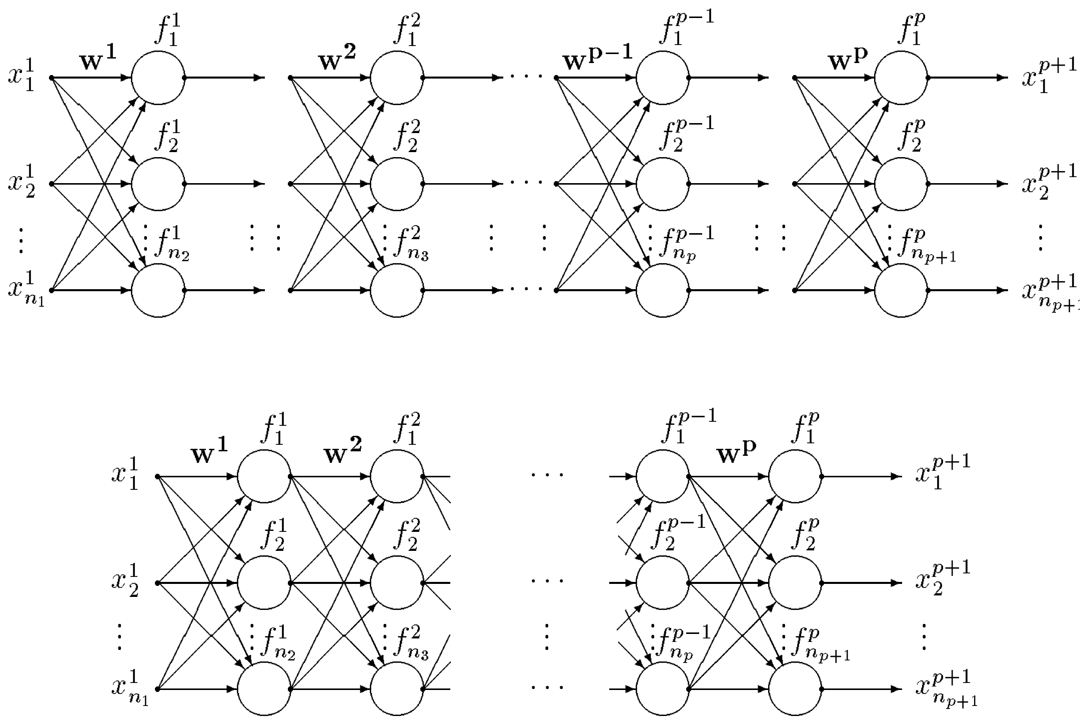

Figure 8. *Several multiple input products and their parametric composition.*

The weight labels $\mathbf{w}^k$ are $n_k \times n_{k+1}$ matrices, $\mathbf{w}^k = [w_{i_{k+1}i_k}^k]_{n_{k+1} \times n_k}$. If it is not assumed that the maps are paired then for each matrix the elements on a row are equal to each other: $w_{i_{k+1}i_k}^k = w_{i_{k+1}i'_k}^k = w_{i_{k+1}}^k$. If the maps are paired this is no longer the case. As an alternative to matrices each input arrow can carry its own individual label.

7.- **Neural Networks.** The basic objects of neural network theory have already been defined. The following is essentially a repetition of previously discussed concepts. A *processing unit*, or simply *unit*, is a parametric map $f : W \times X \to Y$. A *layer* is a parametric product of units $f = f_1 \widehat{\times} \cdots \widehat{\times} f_n : W_1 \times \cdots \times W_n \times X \to Y_1 \times \cdots \times Y_n$ with $f_j : W_j \times X \to Y_j$, $j = 1, \ldots, n$, $f_1 \widehat{\times} \cdots \widehat{\times} f_n(w_1, \ldots, w_n, x) = (f_1(w_1, x), \ldots, f_n(w_n, x))$. And a *neural network*, or simply *network*, is a parametric composition of layers $f = f^p \widehat{\circ} \cdots \widehat{\circ} f^1$. Each f^k being a layer it follows that $f^k = f_1^k \widehat{\times} \cdots \widehat{\times} f_{n_{k+1}}^k : W_1^k \times \cdots \times W_{n_{k+1}}^k \times X_1^k \times \cdots \times X_{n_k}^k \to X_1^{k+1} \times \cdots \times X_{n_{k+1}}^{k+1}$, $k = 1, \ldots, p$. The neural graph for networks is as shown in the lower part of Figure 8. Units, layers and networks are parametric maps.

If the units are paired maps the network can be described as follows. For each $k = 1, \ldots, p$ let $1 \leq i_k \leq n_k$, $1 \leq i_{k+1} \leq n_{k+1}$. The parameter spaces are contained in products $W_{i_{k+1}}^k \subseteq W_{i_{k+1}1}^k \times \cdots \times W_{i_{k+1}n_k}^k$ and there are functions $\phi_{i_{k+1}}^k : X_1^k \times \cdots \times X_{n_k}^k \to X_{i_{k+1}}^{k+1}$, $\xi_{i_{k+1}i_k}^k : W_{i_{k+1}i_k}^k \times X_{i_k}^k \to X_{i_k}^k$ such that $f_{i_{k+1}}^k(w_{i_{k+1}1}^k, \ldots, w_{i_{k+1}n_k}^k, x_1^k, \ldots, x_{n_k}^k) = \phi_{i_{k+1}}^k(\xi_{i_{k+1}1}^k(w_{i_{k+1}1}^k, x_1^k), \ldots, \xi_{i_{k+1}n_k}^k(w_{i_{k+1}n_k}^k, x_{n_k}^k))$. In the corresponding graph each arrow pointing to $f_{i_{k+1}}^k$ carries a label $w_{i_{k+1}i_k}^k$. All this information regarding the architecture of the network can be concisely presented using neural graphs.

It is a remarkable fact that these basic neural-theoretic definitions are just elementary concepts from set theory independent of connectionist diagrams, graphs or topologies. Units, layers and networks are, parameters notwithstanding, the familiar maps, products and compositions. The associated graphs give, once more, a concise and useful description. But the graphs are not part of the definitions, they are a consequence. The function-theoretical viewpoint emphasizes the more mathematical aspects and provides geometrical insight as well as technical language and tools to state and prove results while the neural graphs underline or remind the neurophysiological origins and concerns of the theory.

Applications can be found in [2]-[6].

REFERENCES

[1] Crespin, D. Neural Network Formalism (this present paper).
[2] Crespin, D. Generalized Backpropagation. To appear.
[3] Crespin, D. Geometry of Perceptrons. To appear.
[4] Crespin, D. Neural Polyhedra.
[5] Crespin, D. Pattern recognition with untrained perceptrons.
[6] Crespin, D. Feature Extraction. To appear.

Daniel Crespin
dcrespin@euler.ciens.ucv.ve
Caracas, December 14, 1995.