Spectral Architecture Search for Neural Networks

G. Peri [1] L. Chicchi [1] D. Fanelli [1] L. Giambagli [2]

[1]Department of Physics and Astronomy, University of Florence, Sesto Fiorentino, Italy
[2]Department of Physics, Freie Universität Berlin, Berlin, Germany

Feedforward Neural Networks

In the fields of data analysis and cognitive task automation, neural networks represent the current leading technology.

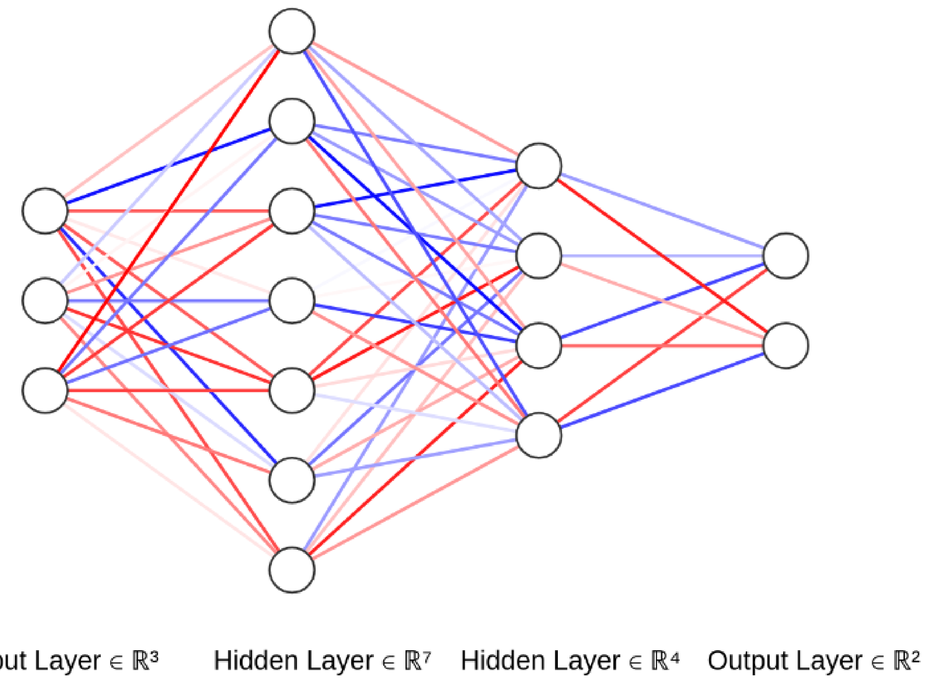
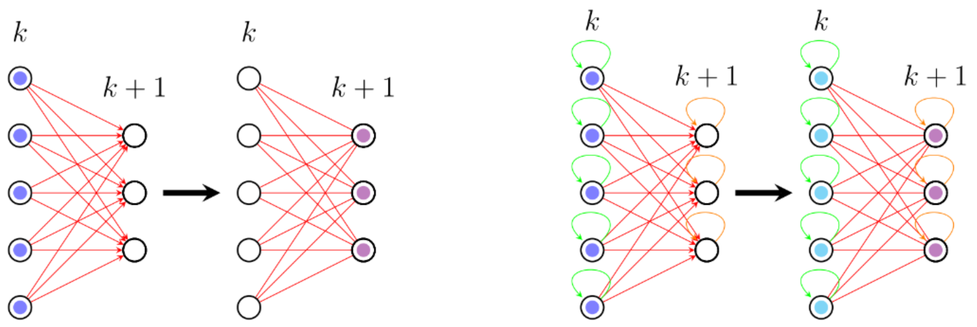


Figure 1. Depiction of a standard feedforward neural network with four layers.

The connections between two layers can be encoded in a rectangular matrix of trainable weights $\mathcal{W}_{i,j}$, where i and j are the indices of the two layers. $\mathcal{W}_{i,j}$ describes the linear transfer of neural activities: non-linearities are then applied on each landing neuron to ensure a rich hypothesis space.

Spectral Description of a Two-Layer Information Transfer

Two layers of the network can be interpreted as a directed bipartite graph, as shown in Figure 2.



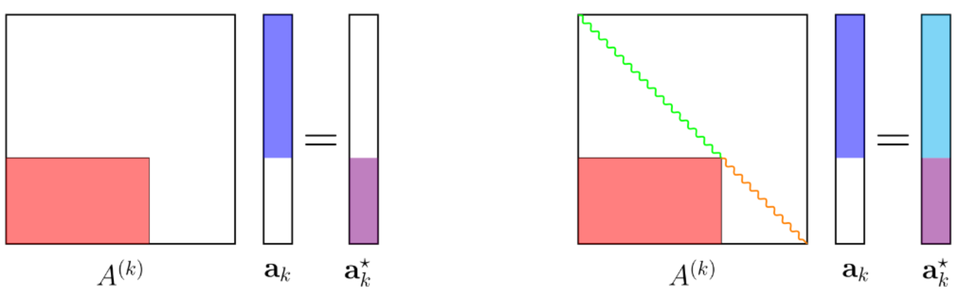


Figure 2. **On the left:** The linear transfer of information between layers k and k+1 can be traced back to a bipartite graph. **On the right:** representation of the same graph, but with the insertion of *self-connections*; notice that the self-connections do not influence the transfer of information between layers.

Matrix $A = A^{(k)}$ with the inclusion of self-loops (see left panel of Figure 2) can be decomposed as $A = \Phi\Lambda\Phi^{-1}$. The structure of the matrices Φ and Λ is illustrated in Figure 3. We also remark that $\Phi^{-1} = 2\mathbb{I} - \Phi$.

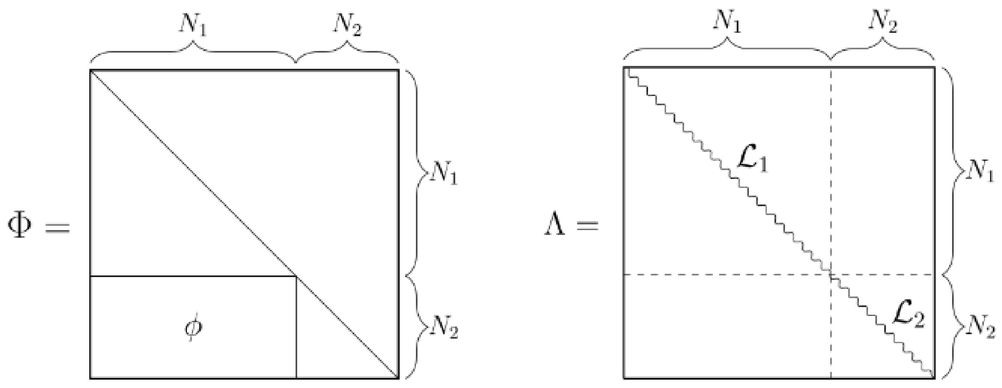


Figure 3. Spectral description of a bipartite graph with two layers.

Spectral parametrization of two layers

The weight matrix $\mathcal{W}$, that describes the linear transfer of activities between two adjacent layers, admits the following parametrization in terms of the eigenvalues and eigenvectors of the adjacency matrix:

$$\mathcal{W} = \phi\mathcal{L}_1 - \mathcal{L}_2\phi \tag{1}$$

The elements of ϕ, as well as the diagonals of $\mathcal{L}_i$, become the parameters of the model; on which we can perform a gradient descent optimization.

Global Spectral Description

We now wish to go beyond the two-layers setting and consider a generic network made of $B+1$ layers. The spectral parametrization can be generalized as depicted in Figure 4.

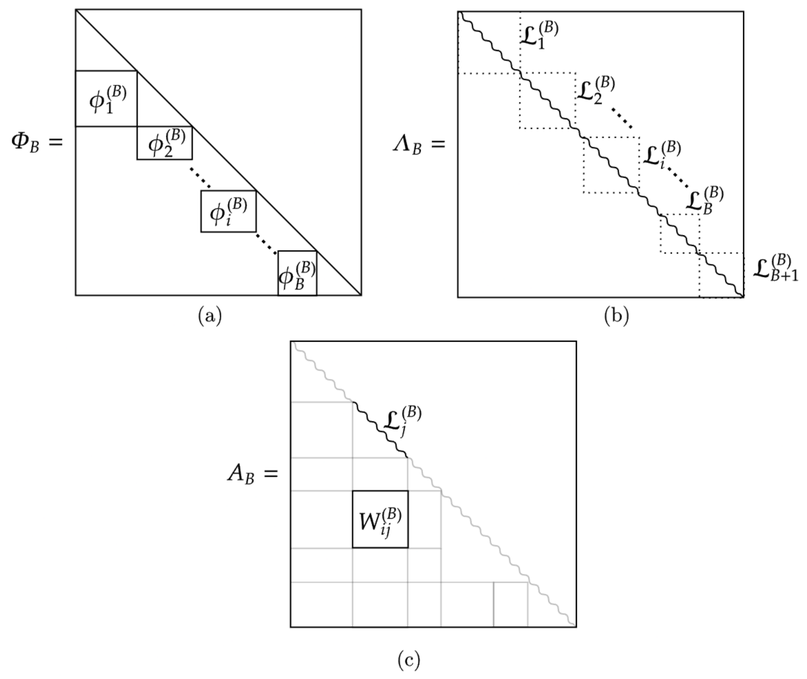


Figure 4. Global spectral parametrization of a neural network. A_B is of course given by $A_B = \Phi_B\Lambda_B\Phi_B^{-1}$

The general inversion formula

It can be shown that Φ_B^{-1}, for any B, is given by the following expression:

$$\Phi_B^{-1} = \sum_{i=0}^{B} (-1)^i \Phi^i \binom{B+1}{i+1}. \tag{2}$$

This general analytical form for Φ_B^{-1} allows us to find the direct expressions for the transfer blocks $\mathcal{W}_{i,j}$, as shown below.

Global spectral parametrization

It is possible to show that under the global spectral parametrization, the blocks $W_{i,j}$ of A read:

$$\mathcal{W}_{i,j} = (-1)^{i-1-j}\left[\phi_{i-1}\mathcal{L}_{i-1} - \mathcal{L}_i\phi_{i-1}\right] \prod_{k=1}^{i-1-j} \phi_{i-1-k}. \tag{3}$$

How to perform Neural Architecture Search

Parametrize an entire neural network made of $B+1$ layers with spectral attributes. Then impose that all eigenvalues must be zero except for the ones corresponding to the output layer. We hence obtain a network with a topology that is equivalent to a linear perceptron (as shown in Figure 5).

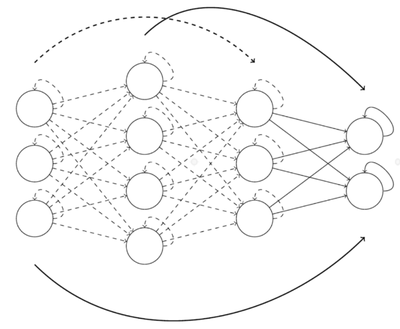


Figure 5. Collapse of the topology to a perceptron when initializing with $\mathcal{L}_i = \mathbf{0}$ con $i \neq B+1$.

This opens the door to a new way of performing Neural Architecture Search (NAS), where we train on the spectral parameters of the network instead of the weights in direct space. In doing so the network can in principle activate, depending on the complexity of the processed task, all other bundle of connections, including long-ranged ones.

A simple Spectral NAS result

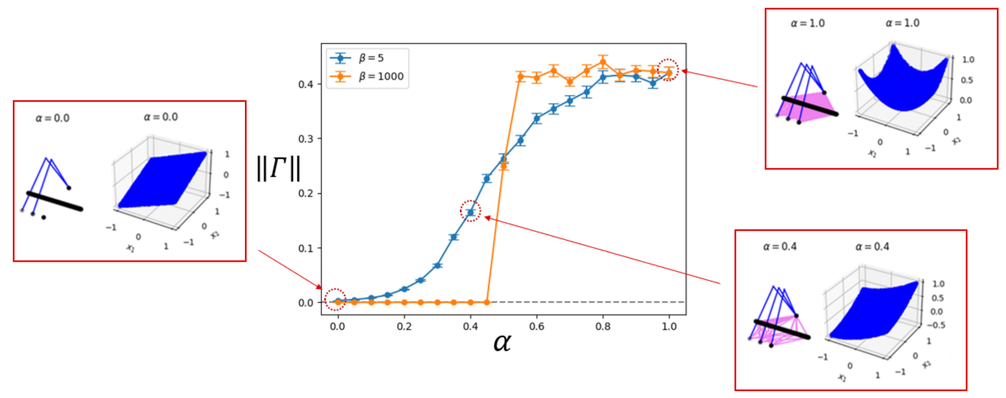


Figure 6. Richness of the topology ||Γ|| vs. complexity (non-linearity) of assigned regression problem (α). β is a parameter dictating the abruptness of the transition between simple (linear) regression problems and harder (non-linear) ones. We can see that the network topology changes appropriately to adapt to the complexity of the task.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Scan the QR code for the full paper! –>