Signal-to-noise optimization: gaining insight into information processing in neural networks

Benjamín Pascual[1,2] and Salva Ardid [2,3*]

[1] PhD in Design, Manufacture and Management of Industrial Projects, Universitat Politècnica de València
[2] Research Institute for Integrated Management of Coastal Areas, Universitat Politècnica de València
[3] Department of Applied Physics, Universitat Politècnica de València
* E-mail: sardid@upv.es

LAB OF NATURAL AND DESIGNED INTELLIGENCE

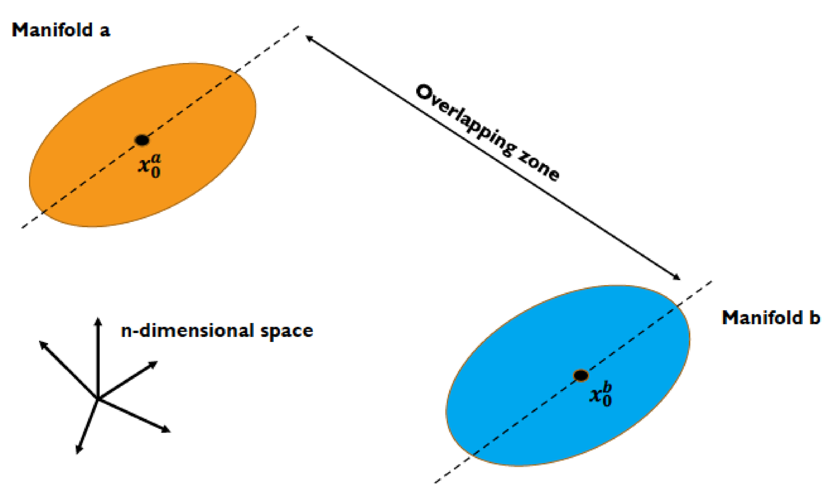
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Introduction

- Mainstream debate in neuroscience and machine learning arguing if neural networks benefit from **Low** [Op de Beeck et al., 2001, Gao and Ganguli, 2015, Gallego et al., 2017, Ansuini et al., 2019, Recanatesi et al., 2019] **vs. High** [Elmoznino and Bonner, 2022] **dimensional representations**.
- We suggest that learning in deep neural networks optimizes **signal-to-noise** processing.
- We also speculate that **nonlinearities** (e.g., in activation functions) facilitate this process.
- To test these hypotheses, we defined a measure of the signal-to-noise ratio (SNR) which can be applied to neural representations associated with predictions of unseen data.

Methods

- In neural networks, patterns of activity define a **manifold**.
- We can analyze these manifolds in **feature space**, e.g., for each class in a categorization task.
- Qualitatively, manifolds' separability can be expressed in terms of the distance between centroids minus their overlap, i.e., the projection of the manifolds in that axis.
- Let's consider two different categories:



- Our definition of the **SNR** then becomes:

$$SNR = \frac{\|\Delta \mathbf{x}_0\| - N_{lineal}}{\|\Delta \mathbf{x}_0\|} = 1 - \frac{N_{lineal}}{\|\Delta \mathbf{x}_0\|} \quad (1)$$

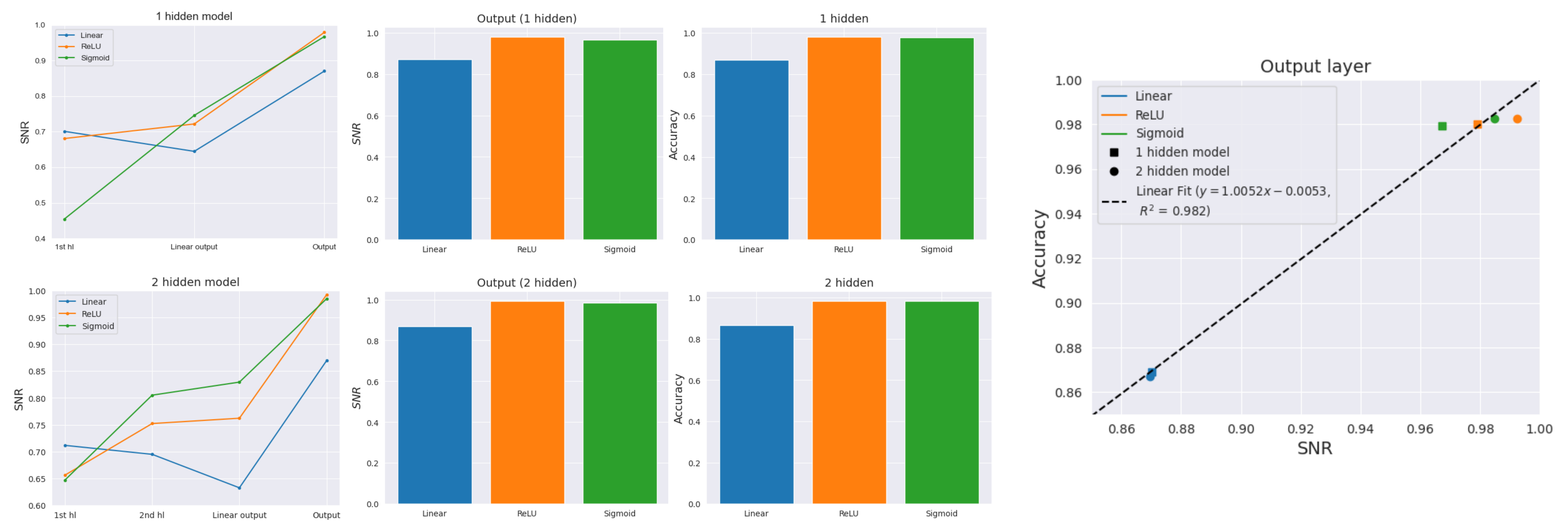
- $\|\Delta \mathbf{x}_0\|$: distance between centroids
- $N_{lineal} = \frac{1}{N_T^0}\sum proj_0^- + \frac{1}{N_T^1}\sum proj_1^-$ quantifies the **overlapping zone**.
- Equation (1) can be used to quantify the **SNR** of a subset adding the term $\sum proj$ (because we are considering all cluster, $\sum proj = 0$)
- We calculate the probability of one input image to belong to one category as $p(a_i) = \frac{a_i}{\sum_j a_j}$
- We used the MNIST image dataset.
- Feedforward neural networks were trained to classify digits as even or odd. These neural networks have one or two hidden layers with 784 neurons each one.

Conclusions

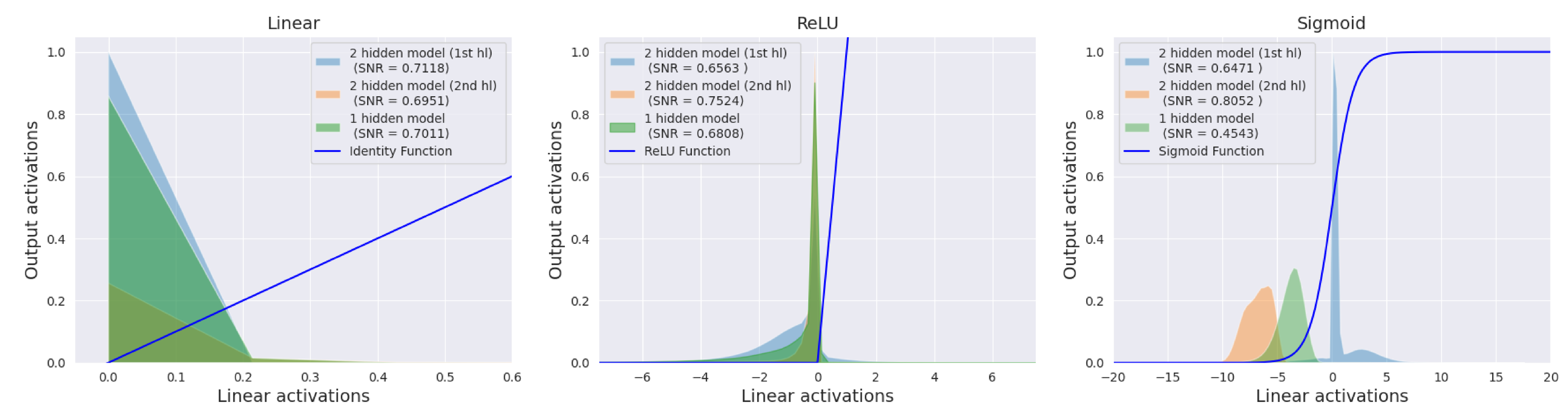
- **High correlation** between **Accuracy** and **SNR** supports our hypothesis that learning optimizes the SNR in neural networks.
- **Early stopping** based on **SNR** better avoids overfitting, when using sigmoid and linear function, than the **loss function**.

Results

- After training **SNR** is highly predictive of the **Accuracy** (top: 1 hidden layer model, bottom: 2 hidden layers model, right: linear fit Acc. vs. SNR)

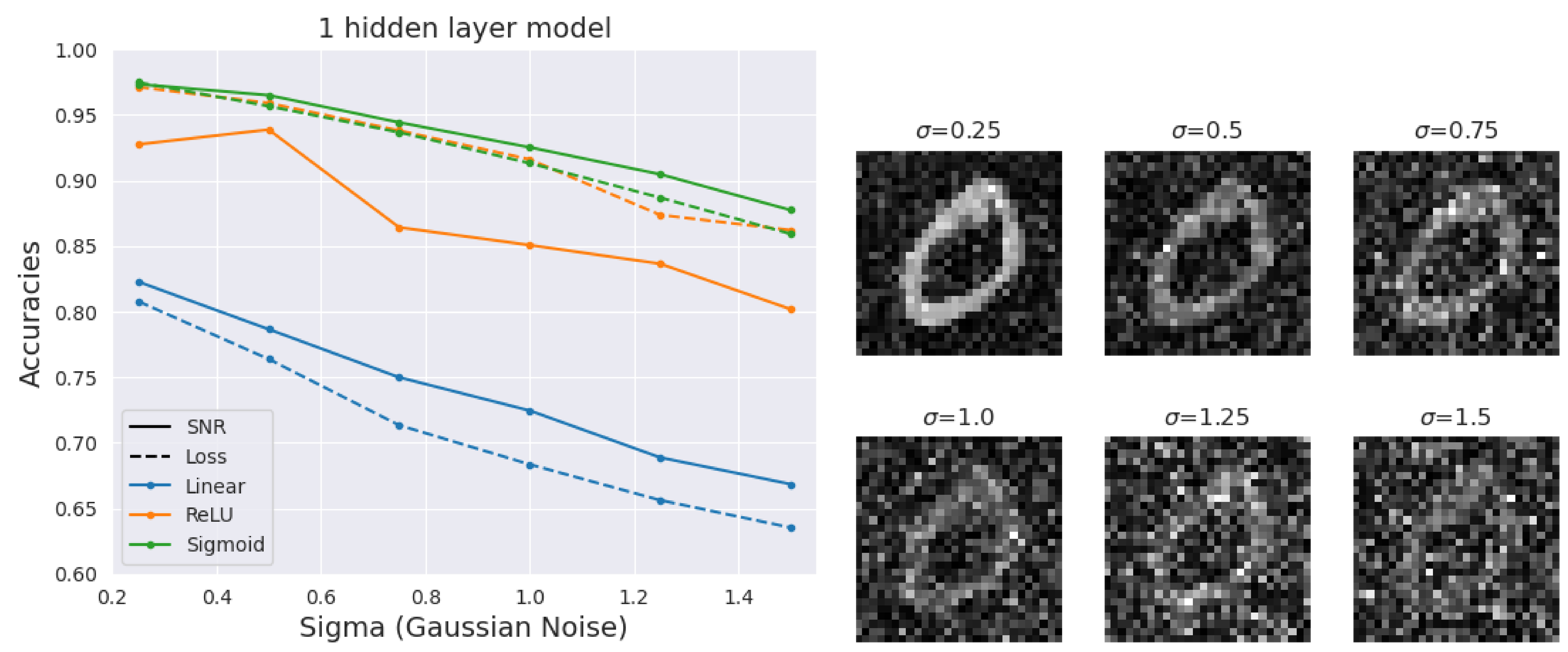


- Dimensionality in output layer (probabilities) is always equal to 1 (two-dimensional space with one constraint), so is not predictive of the accuracy.
- **Distributions of activations** after applying non-linearities show **two modes**: silent (non-preferred input) and non-silent (preferred input) activity



- We analyze if using **SNR** compared to the **loss function** better avoids overfitting when using **early stopping**.

State-of-the-art early stopping approach is using the minimum of the loss function in a reduced dataset (validation). However, this method is sensitive to noise in the validation set [Genkin and Engel, 2020]. Here we adapt the SNR metric (1) assuming that the training and validation sets belong to the same distribution.



- We show that better performance can be achieved by this method when noise is present in the data.
- Remarkably the two non-linearities behave very differently when using early stopping based on SNR: better performance is achieved with the sigmoid function, whereas the ReLU function shows stronger irregularity and worse performance.

References

H. Op de Beeck, J. Wagemans, and R. Vogels. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat Neurosci*, 4(12):1244-1252, Dec 2001

P. Gao and S. Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr Opin Neurobiol*, 32:148-155, Jun 2015.

J.A. Gallego, M. G. Perich, L. E. Miller, and S. A. Solla. Neural Manifolds for the Control of Movement. *Neuron*, 94(5):978-984, Jun 2017.

A. Ansuini, A. Laio, J. H. Macke, and D. Zoccolan. Intrinsic dimension of data representations in deep neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

S. Recanatesi, M. Farrell, M. Advani, T. Moore, G. Lajoie, and E. Shea-Brown. Dimensionality compression and expansion in deep neural networks. *arXiv preprint arXiv:1906.00443*, 2019

E. Elmoznino and M. F. Bonner. High-performing neural network models of visual cortex benefit from high latent dimensionality. *bioRxiv*, pages 2022-07, 2022.

B. Sorscher, S. Ganguli, and H. Sompolinsky. Neural representational geometry underlies few-shot concept learning. *Proc Natl Acad Sci U S A*, 119(43):E2200800119, Oct 2022.

Genkin, M., Engel, T.A. Moving beyond generalization to accurate interpretation of flexible models. *Nat Mach Intell* **2**, 674–683 (2020). https://doi.org/10.1038/s42256-020-00242-6

This work was supported by Grant PID2020-120037GA-I00 funded by MCIN/AEI/10.13039/501100011 and the Generalitat Valenciana Gen-T Program CIDEGENT/2019/043