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

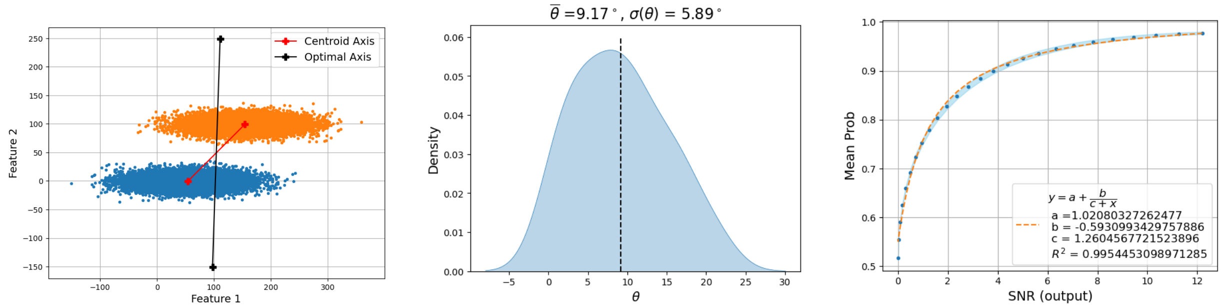
[2] Research Institute for Integrated Management of Coastal Areas, Universitat Politècnica de València

[3] Department of Applied Physics, Universitat Politècnica de València

[*] Corresponding Author E-mail: sardid@upv.es

Abstract

The current view in neuro-AI is that neural networks compress representations into low-dimensional manifolds [Gao and Ganguli, 2015]. A study challenges this view, arguing that neural networks benefit from high dimensionality [Elmoznino and Bonner, 2022]. We argue that learning in neural networks may optimize the signal-to-noise ratio (SNR), so neural networks can benefit from feature compression and expansion to: (i) increase signal processing and (ii) diminish noise, while (iii) mapping inputs into outputs. Yet, if SNR is optimized through learning, a causal relationship shall exist with model performance (e.g., accuracy or the probability to select the proper response). To test this, we introduce a SNR metric computed as the ratio between the signal (defined as the square distance between clusters), over the noise (defined as the sum of variances of the clusters). Unlike [Sorscher et al., 2022], our metric is optimized: we determine the *optimal axis* that maximizes SNR, which may differ from the *centroid axis*. Projecting data to this axis reduces the signal but also reduces, and more strongly, the noise, resulting in higher SNRs. We show this in panel (a): the SNR along the axis in black is larger than that along the centroid axis. In this example, there is no overlap with respect to the optimal axis, whereas the projections to the centroid axis do overlap significantly (not shown). In panel (b), we compute the distribution of the angles formed between these two axes when applied to the linear output of $n = 40$ neural networks classifying MNIST digits by parity. Using a perturbative analysis, we show that SNRs are predictive of the probability to select the proper response (panel c), and the accuracy (not shown). In the analysis, we gradually shortened the distance between the manifold centroids of each category keeping the variance untouched. This analysis shows that the performance relates to the SNR, while the dimensionality of the data remains constant.



(a) (b) (c)

Keywords — deep neural networks, manifold, high vs. low dimensionality, signal-to-noise ratio, optimization, overlap

References

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.

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