The heresy of unheard-of simplicity Comment on "The unreasonable effectiveness of small neural ensembles in high-dimensional brain" by A.N. Gorban, V.A. Makarov, and I.Y. Tyukin

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
El Paso, Texas 79968, USA, email vladik@utep.edu

"Grandmother" neurons: a puzzle. From biology, we know that originally, life consisted of simple primitive single-cell organisms – like modern amoebas. Eventually, cells learned how to collaborate, and thus, multi-cell organisms appeared, organisms that were able to perform more complex functions. Similarly, a single ant or a single bee are rather primitive creatures, but together, ants and bees can perform very complex tasks. We can go on and on with such examples.

In general, when there is a simple function, often, we can find a single cell that performs this function. For example, the perception of light in our eyes is performed by simple cells, each of which detects the light of certain color coming from a certain direction. When we need to explain a more complex behavior – e.g., breathing, digestion, or reproduction – we usually find a large number of different cells acting in collaboration.

Because of all these examples, we expect the same from the brain: to explain simple brain activity, it may be sufficient to consider single cells (neurons) responsible for this activity, but to explain more complex intellectual phenomena – e.g., recognizing one's grandmother – one expects a perfectly orchestrated collaboration between many different cells. But this is *not* what the analysis of brain activity shows, what it shows is that, crudely speaking, for each complex notion like recognizing one's grandmother (or recognizing Jennifer Aniston), there is an individual neuron responsible for exactly this complex task.

In other words, our brain – which is responsible for our most complex behavior and is, therefore, expected to have very complex structure – is much simpler than we expected.

The brain is the most puzzling example, but the same phenomenon has been observed in social sciences, in many other disciplines. Even poets noticed – the Nobelist Boris Pasternak wrote, in his 1931 poem "The Waves" [9]:

Assured of kinship with all things
And with the future closely knit

We can't but fall – a heresy! –
To unbelievable simplicity.

But to be spared we can't expect
If we do not conceal it closely.
Men need it more than anything,
But complex things are easier for them ...

Maybe this is in our genes? Maybe God, in his infinite wisdom, created us in such a way, so that we can be smart? Religion aside, maybe this is in our genes? One neuron is designed for sines, one for cosines, one for integration, etc. – maybe this is how we are born?

This might have been a reasonable explanation until deep learning appeared, a modern technique that enables computer to learn from data as well as (and often better than) we do; see, e.g., [4]. For example, by showing a deep learning algorithm several samples of Go play, we can train it to beat the world Go champion.

The main idea behind deep learning is rather simple: in a nutshell, it is a network consisting of artificial neurons, each of which is a model of a biological model (a very simplified model). The main difference from previously used neural networks – which usually had a few layers (e.g., 3) – deep neural networks have many layers. And here is a surprise: it turns out that what neurons on intermediate layers recognize are exactly corresponding human concepts. So, it is not just in our genes.

There are some explanations of some features of deep learning, but this remains a puzzle. Many empirical observations related to deep learning can be theoretically explained, from the efficiency of piece-wise linear activation functions [2] to the use of geometric mean for averaging [3], the use of softmax instead of exact optimization [8], and the use of Kullbach-Leibler (KL) divergence instead of the usual least squares [4, 7]. There are also general explanations for surprisingly simplicity of many physical models [6]. But the extreme simplicity of grandmother neurons remains a puzzle.

Explanation by Gorban et al. The main purpose of the paper [5] is to explain this phenomenon. This explanation is based on the known fact that for the multi-D distribution, the vast majority of points are located in a smaller-dimensional subspace. To non-mathematicians, this may sound like a weird and unusual result, but this is exactly what statistics (see, e.g., [10]) and statistical physics (see, e.g., [1, 11]) are based on. For example, while flipping coins is a random process, if we flip a coin many times, then, with high confidence, the number of tails will be close to the half of the number of all experiments. In mathematical terms, this means that all the trajectories are close to the hyperplane corresponding to 1/2 of tails and 1/2 of heads. Similarly, in statistical physics, while individual gas molecules randomly fluctuate, these fluctuations average out, and we get, in effect, a deterministic behavior.

In mathematics, this is a known result. It is known in applications like statistics or statistical physics, where individual events are independent. In many real-life situations, however – e.g., in complex situations analyzed by our brains – individual events are *not* independent. To cover these cases, Gorban et al. proved a natural extension of the above known result to the not-necessarily-independent case, namely, to the case when events with small probability in the independent case remain small-probability ones if we take dependence into account.

It is also known that with high probability, two random vectors in a multi-D space are almost orthogonal. The authors also extend this result to the possibly-dependent case. They prove an even more general result: namely, they prove that the same phenomenon holds if we take many random vectors instead of just two. For example, in the 100-dimensional space, with high probability, 2 million random vectors all have angles larger than 45^o between every two of them. This new result shows that it is possible to separate each of these vectors from all the others by a linear separating function – which is exactly how neurons make their accept-reject decisions. Thus, the results from [5] indeed explain why such simple single neurons are ubiquitous in human reasoning.

Beyond explanations. Explanations are important, but a post-facto explanation is not that impressive, it is what Lev Landau (another Nobelist from Russia) used to derogately call "Neue Begrundung" ("New Foundations"). Good news is that Gorban et al. went beyond explanations: they used their ideas to come up with efficient algorithms, e.g., for correcting errors in AI systems.

Overall impression. This is a very interesting paper, with unexpected new ideas, ideas that have a potential to bring revolutionary changes to our understanding of complexity – and to algorithms based on this understanding.

This paper reminded me of a story well-known by mathematicians: when Riemann – father of Riemannian geometry that describes curved space (and later curved space-time) – finished his first lecture, the great Gauss (who attended this lecture) left the lecture hall "in deep thought". This is how I feel after reading this paper, this is how many readers will feel – and these deep thoughts will hopefully help us go even further forward in these studies!

References

[1] R. Feynman, R. Leighton, and M. Sands, *The Feynman lectures on physics*, Addison Wesley, Boston, Massachusetts, 2005.

[2] O. Fuentes, J. Parra, E. Anthony, and V. Kreinovich, "Why rectified linear neurons are efficient: a possible theoretical explanations", In: O. Kosheleva, S. Shary, G. Xiang, and R. Zapatrin (eds.), *Beyond Traditional Probabilistic Data Processing Techniques: Interval, Fuzzy, etc. Methods and Their Applications*, Springer, Cham, Switzerland, 2018, to appear.

[3] A. Gholamy, J. Parra, V. Kreinovich, Olac Fuentes, and E. Anthony, "How to best apply deep neural networks in geosciences: towards optimal 'averaging' in dropout training", In: J. Watada, S. C. Tan, P. Vasant, E. Padman-

abhan, and L. C. Jain (eds.), *Smart unconventional modelling, simulation and optimization for geosciences and petroleum engineering*, Springer Verlag, 2019, pp. 15–26.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Leaning*, MIT Press, Cambridge, Massachusetts, 2016.

[5] A. N. Gorban, V. A. Makarov, and I. Y. Tyukin, "The unreasonable effectiveness of small neural ensembles in high-dimensional brain", *Physics of Life Reviews*, 2019, this issue.

[6] O. Kosheleva and V. Kreinovich, "Explaining Boris Pasternak's observation that complex ideas are sometimes easier to understand", *Journal of Innovative Technology and Education*, 2016, Vol. 3, No. 1, pp. 9–12.

[7] O. Kosheleva and V. Kreinovich, "Why deep learning methods use KL divergence instead of least squares: a possible pedagogical explanation", *Mathematical Structures and Modeling*, 2018, Vol. 46, pp. 102–106.

[8] V. Kreinovich, "From traditional neural networks to deep learning: towards mathematical foundations of empirical successes", In: S. N. Shahbazova, J. Kacprzyk, V. E. Balas, and V. Kreinovich (eds.), *Proceedings of the World Conference on Soft Computing*, Baku, Azerbaijan, May 29–31, 2018.

[9] G. Reavey, *The poetry of Boris Pasternak*, G. P. Putnam, New York, 1960.

[10] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[11] K. S. Thorne and R. D. Blandford, *Modern classical physics: optics, fluids, plasmas, elasticity, relativity, and statistical physics*, Princeton University Press, Princeton, New Jersey, 2017.