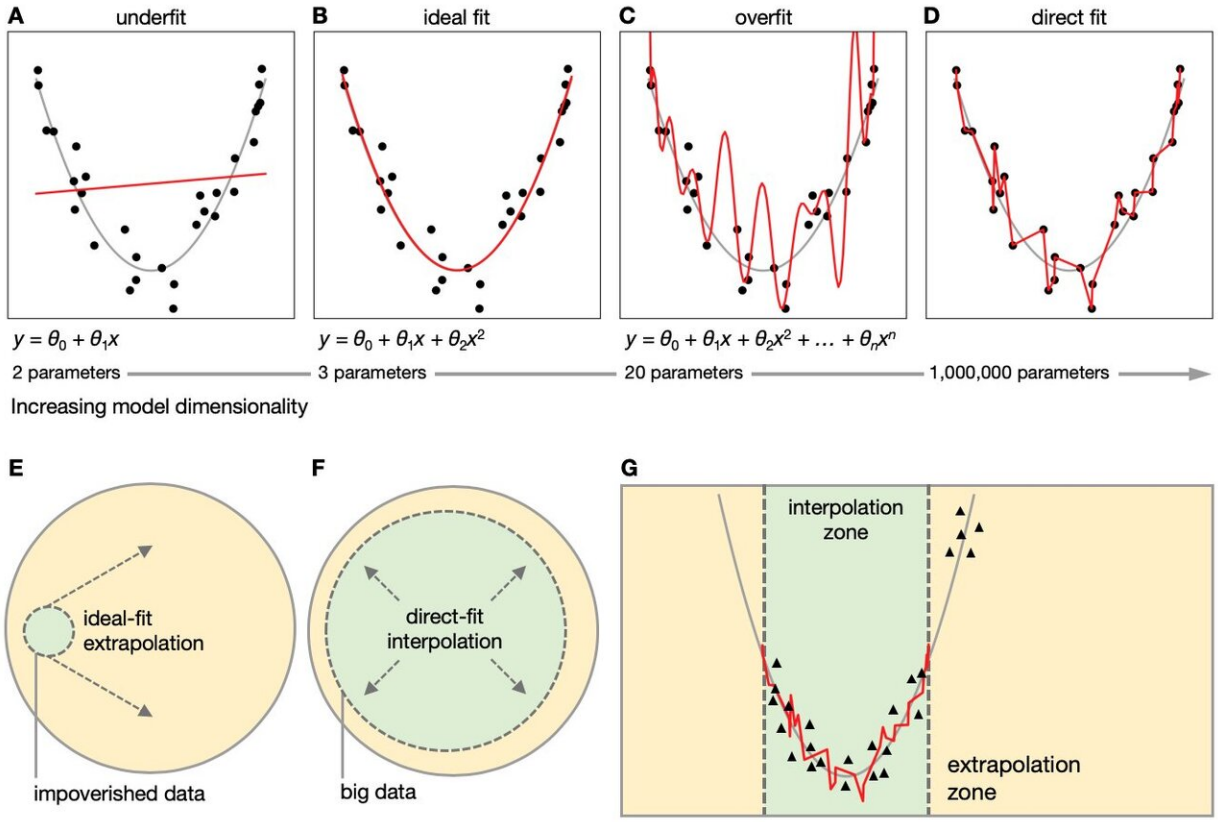
Medical Xpress

A perspective on the study of artificial and biological neural networks

February 24 2020, by Ingrid Fadelli



Direct-Fit Learning with Dense Sampling Supports Interpolation-Based Generalization. (A) An overly simplistic model will fail to fit the data. (B) The ideal-fit model will yield a good fit with few parameters in the context of data relying on a relatively simple generative process; in fact, this is the model used to generate the synthetic data (with noise) shown here. (C) An overly complex (i.e., over-parameterized) model may fixate on noise and yield an explosive overfit. (A)–(C) capture the "textbook" description of underfitting and overfitting. (D) Complex models, such as ANNs, however, can nonetheless yield a fit that both

captures the training data and generalizes well to novel data within the scope of the training sample (see G and Bansal et al., 2018 for a related discussion). (E) Traditional experimentalists typically use highly controlled data to construct rule-based, ideal-fit models with the hope that such models will generalize beyond the scope of the training set into the extrapolation zone (real-life data). (F) Direct-fit models—like ANNs and, we argue, BNNs—rely on dense sampling to generalize using simple interpolation. Dense, exhaustive sampling of real-life events (which the field colloquially refers to as "big data") effectively expands the interpolation zone to mimic idealized extrapolation. (G) A direct-fit model will generalize well to novel examples (black triangles) in the interpolation zone but will not generalize well in the extrapolation zone. Credit: Hasson, Nastase & Goldstein.

Evolution, the process by which living organisms adapt to their surrounding environment over time, has been widely studied over the years. As first hypothesized by Darwin in the mid 1800s, research evidence suggests that most biological species, including humans, continuously adapt to new environmental circumstances and that this ultimately enables their survival.

In recent years, researchers have been developing advanced computational techniques based on artificial neural networks, which are architectures inspired by biological neural networks in the human brain. Models based on artificial neural networks are trained to optimize millions of synaptic weights over millions of observations in order to make accurate predictions or classify data.

Researchers at Princeton University have recently carried out a study investigating the similarities and differences between artificial and biological neural networks from an evolutionary standpoint. Their paper, published in *Neuron*, compares the evolution of biological neural networks with that of artificial ones using psychology theory.

"This project originates from the puzzle of why modern deep neural networks excel—and learn to be as good, if not better, than humans—in many complex cognitive tasks," Uri Hasson and Sam Nastase, the primary authors of the paper, told Medical Xpress. "This puzzle drew our attention to the similarities as well as differences between artificial and biological neural networks."

While quite a few psychology and neuroscience researchers have tried to better understand the structure and functioning of deep neural networks, many have found them hard or impossible to interpret, due to their sheer complexity. Hasson, Nastase and their colleague Ariel Goldstein wanted to show that trying to describe complex neural networks inspired by those in the brain using simple and intuitive representations may be highly challenging and unrealistic.

"We argue that both artificial neural networks and biological neural networks aim to guide action in the real world," Hasson and Nastase explained. "Scientists try to understand the underlying structure of the world, but neural networks learn a direct mapping from useful elements of the world to useful behaviors (or outputs). They don't construct ideal or intuitive models of the world—rather, they use an overabundance of parameters (e.g., connection weights or synapses) to reflect whatever task-relevant structure is out in the world."

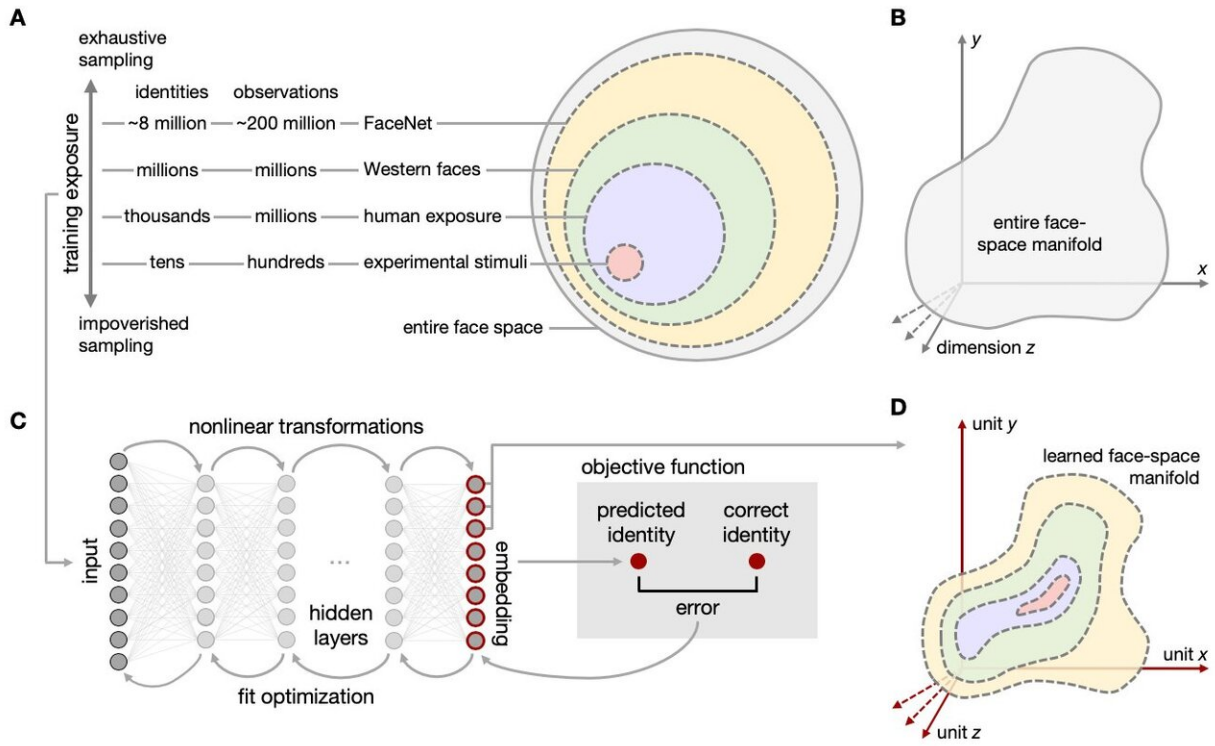


Figure showing how the interpolation zone in ANNs is bound by the training set. .Credit: Hasson, Nastase & Goldstein.

In their paper, the researchers essentially propose that models based on artificial neural networks do not learn rules or representations of the world around them that are easy for humans to interpret. On the contrary, they typically use local computations to analyze different aspects of data in a high-dimensional parameter space.

To explain their reasoning further, Hasson, Nastase and Goldstein used two simple examples of existing neural networks. In their paper, they describe how these two networks learn to interpolate different data they encountered and their features without learning ideal 'rules' for generalization. They then draw an analogy between the 'mindless' optimization-based learning strategies employed by artificial neural

networks and the 'blind' adaptation of species, including humans, over several years of biological evolution.

"We argue that similarities between the brain and artificial neural networks may undermine some standard practices in experimental neuroscience," Hasson and Nastase said. "More specifically, we suggest that the tradition of using tightly controlled experimental manipulations to probe the brain for simple, intuitive responses may be misleading. On the other hand, we suspect that artificial neural networks will benefit tremendously from the development of more ecological objective functions (i.e., goals)."

The theoretical analyses carried out by this team of researchers could have numerous interesting implications for future psychology and neuroscience research, particularly for studies aimed at better understanding artificial neural networks. Notably, their findings highlight the need for a change in the methods generally used to investigate the human brain, as well as new computational architectures inspired by it.

Essentially, Hasson, Nastase and Goldstein feel that the use of contrived experimental manipulations with the hope of uncovering simple rules or representations is unlikely to yield models that can be effectively applied to the real world. In contrast, mindless fitting of big data to big models is likely to provide the brain with the needed biases to act and predict phenomena in the real world. Such blind optimization, like evolutionary theory, may be far better suited for the brain, as it is a relatively simple and yet powerful way of guiding learning of different real-world phenomena.

"We now plan to develop further naturalistic experimental paradigms to explore the human brain in more ecological contexts," Hasson and Nastase said. "We are also interested in exploring how human brains and

artificial neural networks interact, and whether we can identify shared signals between the two."

More information: Uri Hasson et al. Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks, *Neuron* (2020). DOI: 10.1016/j.neuron.2019.12.002