Emergent behaviour in complex systems

Daniel Ahlberg, Stockholm University *December 17, 2024*

Proposal for Master degree project in Mathematical Statistics

Emergent behaviour refers to phenomena observed in systems consisting of many small parts which cannot be attributed to the parts themselves, but which appear as a result of interaction between the parts. A example of such behaviour is the appearence of so-called phase transitions in physical systems, e.g. when a material transitions from a solid to liquid state, or loses its magnetic properties, both as a consequence of increasing temperature.

In recent years, a collection of methods referred to as machine learning and artificial intelligence have come to awareness of the general public, as a consequence of the striking success of generative artificial intelligence and large language models. These models are at the core based on neural networks, which can be thought of as a complex class of regression functions used to fit training data, often in a widely overparametrised setting. Nevertheless, careful training of such networks have led to a series of competent tools for generating text and images, and can be perceived as exhibiting near-human levels of intelligence.

The recent success with generative artificial intelligence sparks a number of questions regarding the nature of intelligence. For instance, whether this is the first step towards artificial general intelligence? Is the perceived intelligence in large language models simply emerging as a consequence of interacting neurons in a network of this size?

Studies over the past few decades have led to a mathematical understanding of emergent behaviour such as phase transitions. This includes the emergence of a gigant component in random graphs, the formation of an infinite cluster in percolation models, and the transition from an ordered to a disordered state in models for magnetism such as the Ising model. Together with non-rigorous arguments and simulation studies from the physics community, our understanding of emergent behaviour in complex systems is greater still. Nevertheless, the vast complexity of deep neural networks have so far prevented us from understanding what goes on under the hood.

The purpose of this proposal is to explore what probabilistic techniques and our understanding of emergent behaviour in complex systems can tell us about the striking success of deep neural networks. There are various aspects of deep neural networks that could be of interest to investigate from a theoretical perspective, possibly together with simulation studies.

1. Is it possible to explain the apparent intelligence of large language models as a phenomena emerging due to its large size?

2. How can we explain the fact that the training of neural networks gives rise to complex networks and simultaneously avoid overfitting?

3. What can be said about the structure of firing neurons in a trained neural network? Is there any resemblance with critical behaviour of random networks?

4. The scaling $\frac{1}{n}$ is recurring in the training of neural networks, bandit problems and reinforcement learning. What is the common factor in all these cases?

A video lecture from a leading scientists at Microsoft Research can serve as inspiration: `https://youtu.be/XLNmgviQHPA?si=3S0CqC2or4wUZ7cM`