Deep learning is revolutionising our world, but why does it work so well?

Deep learning is making breakthroughs across many scientific and commercial domains, but we don't have a theory which explains why it works so well. Much of the complexity of deep learning arises because it can solve the same problem in many different ways. We can view these solutions as a shape, where complex attributes of the model manifest as geometric features called singularities. We believe that these singularities encode rich information which governs the behaviour of models, and that the path to understanding deep learning will be found via understanding these singularities.

Deep learning: a recent history

The accomplishments of deep learning in the last five years alone are dizzying. In 2016 Google's AlphaGo defeated the world's best Go players, a board-game long-thought to demand levels of creativity beyond artificial players. In 2020 Google's AlphaFold determined the structure of protein molecules to near-experimental levels of accuracy, which biologists described as a "game changer" for medical and biological research. Also in 2020 OpenAI unveiled GPT-3, which can write sophisticated articles which are often indistinguishable from those written by humans – see the picture below. And just last week Microsoft announced GitHub Copilot, a deep learning model which Microsoft claims can write entire computer functions based on short descriptions. Deep learning is proving to be a technology which could revolutionise a number of fields, and we are just starting to see its potential.

Why do we need a theory?

Currently deep learning practitioners rely on experiments to develop and test models. But having a large amount of data is essential to how deep learning works: many phenomena are simply not observed when working with small amounts of data. This means prototyping with scaled-down models can be misleading. On the other hand experiments involving a large amount of data can be eye-wateringly expensive: running the computations necessary to produce the final version of GPT-3 is estimated to have cost over USD \$4 000 000. These makes studying and improving upon cutting-edge deep learning models very difficult.

A theory could both explain deep learning's extraordinary results and inform practitioners as they develop new models, just as our understanding of physics informs the engineering discipline.

A crash course in deep learning

Let's take a step back and discuss, in broad terms, what a deep learning model does and how it is taught to solve a problem. A model takes in some input and produces an output.

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**
The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

Figure 1: An example of the output of GPT-3. The text in grey is a human-written prompt – the input to GPT-3 – and the text in black is the resulting output. The authors of the paper found that, of 25 similarly generated news pieces, people found this example the most difficult to identify as being written by GPT-3. Reproduced from Figure 3.14 in Brown et al, (2020) "Language Models are Few-Shot Learners", arXiv:2005.14165.

For example, GPT-3 takes in a text prompt, and outputs a completion of that prompt. The behaviour of the model is controlled by adjustable parameters, which change how different parts of the input affect the output. The process of training a model amounts to repeatedly making small tweaks to these parameters which nudge the model closer to the desired output.

Towards a theory

A key reason deep learning models are hard to understand is that there are many different parametrisations of the model which give the correct output. In other words, a model can solve a given problem in many different ways. We can view the set of possible solutions as a shape by considering each parametrisation of the model as a point in space. It turns out that this shape can be extremely complex. Where a circle, sphere or donut shape changes smoothly and predictably as you move over it, this shape often has sharp twists, creases, self-intersections and other complexities. These complexities are called *singularities* and are the key reason why classical statistical techniques do not work for deep learning.

We believe that understanding these singularities is key to developing a theory of deep learning. When a model is close to a solution the effect of making small adjustments to its parameters can be hard to understand. Recent work by Sumio Watanabe has shown that by systematically removing these singularities we can simplify the behaviour of the model near solutions. Singularities also encode information about the model in a way which is not immediately obvious. By analysing the geometry of the singularities we can predict certain properties of the model, for example how well a model is likely to generalise from the examples it was trained with.

It is likely that these singularities have a lot more to tell us. There are rich theories about singularities arising in other areas of mathematics and physics which have not been applied to deep learning. Our project aims to adapt results about singularities from other areas to the context of deep learning.

So how do we tackle these singularities? One technique – the one used by Watanabe – is the *blow-up*. This is a way of pulling apart singularities and removing them from the shape, and is best illustrated with a picture – see below. The red curve is a simple example of the set of solutions available to some deep learning model. At the point labelled A we have a singularity – a 'self-intersection'. The purple spiral is the result of blowing up. One can see that the shadow cast on the plane by the purple curve is exactly the red curve. The singular point A has been separated into two points on the purple curve, and the purple curve no longer intersects itself. In general you can think about the process of blowing up as follows. We take our shape, and think about it as being the shadow cast by some other shape without any singularities. Our goal is to find a shape which casts this shadow. A remarkable fact is that we can *always* remove all singularities in this way, no matter how complex. This was proved in 1964 by Heisuke Hironaka, in a paper over 200 pages long!

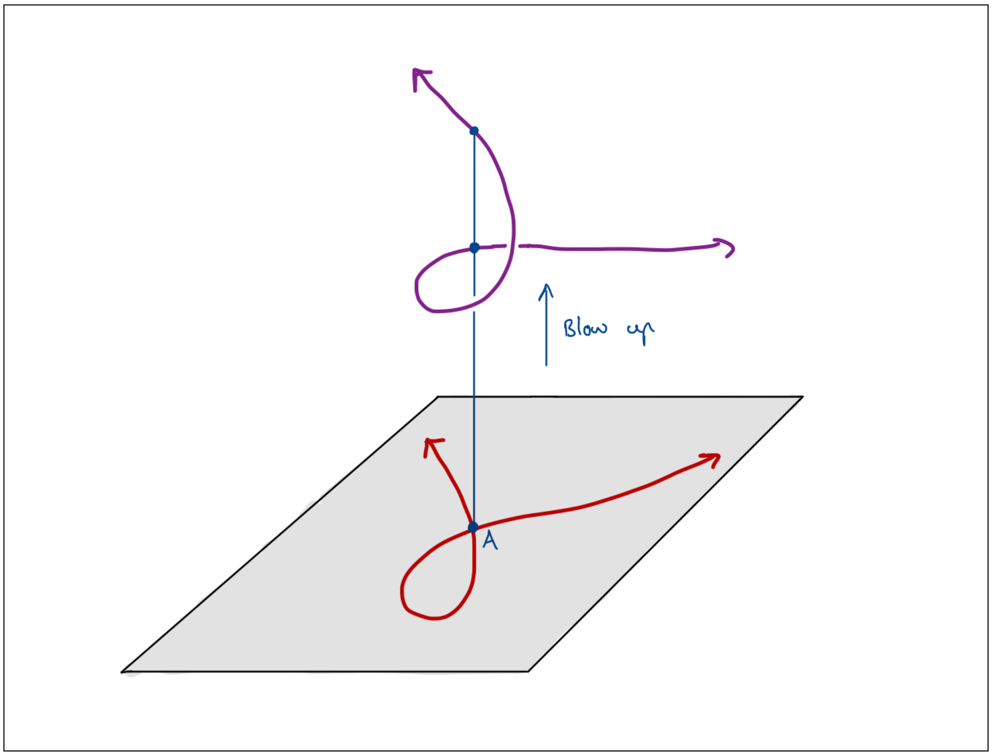

Figure 2: An illustration of a blow-up. The curve in red lies entirely within the plane (grey) and has a singularity at the point A. The result of blowing up in order to remove the singularity at A is illustrated lying above in purple.

Singularities arise across across a variety of areas of mathematics, from quantum physics, to differential equations, to the study of knots. When they arise, understanding the singularities is often key to understanding the problem at hand. As such, there are many sophisticated techniques – like the blow-up – that we can use to analyse the singularities arising in deep learning. We don't know exactly what these singularities have to tell us about deep learning, but we firmly believe that the path to understanding deep learning lies via understanding these singularities. By focusing the study of deep learning around these singularities we are best positioned to discover what makes deep learning special.

Words: 1024

References

- Brown, T. B. et al. (2020). Language models are few-shot learners. arXiv:2005.14165.
- Callaway, E. (2020). 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature*, 588(7837), 203–204. doi:10.1038/d41586-020-03348-4
- Friedman, N. (2021). *Introducing GitHub Copilot: your AI pair programmer*. The Github Blog. https://github.blog/2021-06-29-introducing-github-copilot-ai-pair-programmer/
- Murfet, D., Wei, S., Gong, M., Li, H., Gell-Redman, J., & Quella, T. (2020). Deep Learning is Singular, and That's Good. arXiv:2010.11560 [cs].
- Silver, D. et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. doi:10.1038/nature16961
- Watanabe, S. (2009). *Algebraic geometry and statistical learning theory*. Cambridge University Press.