Lecture 7 - Deep Learning

Sondre Elstad

2022-06-25

Introduction

We have now spent some time getting an overview of the different functionalities of machine learning. We have seen different techniques and we have had a very technical perspective. However, if one wish to understand something, it is not enough to be able to DO. Rather, one needs an understanding of WHY one is doing what one is doing. Yes, of course, you could argue that the reason why we are doing supervised learning is that we wish to gather information, and thus being able to draw inferences, about the problems for which we are carrying out the machine learning analysis in the first place. There might be someone applying for a job at our company, and we want to estimate the wages this candidate could get at a rival company. In fact, we have seen that supervised learning is more or less a rephrasing of conventional statistics. Albeit with a slightly different perspective. We have also seen that unsupervised learning can help us make sense of a large and messy data set. But why are these techniques called learning? I have used the term learning algorithm here and there in these notes. Why do we stress the word 'learning' so much? If most of these techniques are simply a redressing of old techniques, why do we bother? In the second lecture of this machine learning part we talked about interpretability vs. flexibility. I argued that techniques such as support vector machines and neural networks tended to overfit the data. Why is there, then, so much focus on neural networks? In this lecture we are going to talk about deep learning and we are going to reflect a little bit on the learning 'paradigm'. The discussion in this lecture is based on *Deep Learning* (written by Ian Goodfellow, Yoshua Bengio and Aaron Courville).

Human vs. machine thinking

We all know that computers are able to solve some problems much faster than the human brain is. Try computing 2^{32} in your head! Using R on my laptop, I got to the correct answer - 4 294 967 296 - in about 0.0003 seconds. That's $\frac{3}{10000}$th of a second! The first 31.4 trillion decimals of π are known. We can store all these decimals on a large computer. However, no single human being has ever been or will ever be able to remember 31.4 trillion numbers without making a mistake. Why, then, is it so difficult to teach a computer how to pour itself a glass of orange juice? Computers are remarkable at solving problems that can be described by a list of formal, mathematical rules. But there are many tasks, which we as humans find trivial, but which a computer struggles to perform.

But what we must remember is we must have learnt all of these trivial tasks at some point. I have yet to encounter a newborn who is able to reach out their hand and present themselves. We do so everyday. We talk to new people regularly. But learning all of the written and (perhaps even more troubling, the unwritten) social codes and norms for proper conduct in society is a massive task! There is an incredible amount of information which must be processed in order to be able to live everyday life. We are able to learn from experience. We do mistakes. All of us can be impolite from time to time. We do not wish to be (well, some might and who am I to judge?), but most people try to be polite. The point is that we learn from experience. The difference between learning mathematics and learning how to eat dinner at a fancy restaurant, is that the dinner requires subjective and intuitive knowledge whereas the mathematical problem requires logic.

When learning something new (machine learning for example), we build new intuition on top of old intuition. Although it would save us a lot of time, we cannot start directly at the master's level. It sounds interesting to do a master's in Data Science, but I do believe there are many concepts we need to learn before we can make any sense of the syllabus at master's level. We need to learn the basic concepts before we can apprehend the more complex ones.

Deep Learning

Deep learning has been presented as a solution to the problem of solving tasks that humans find easy, but which computers find really hard. Just as humans learn from experience, it has been found that computers ought to be able to learn from experience as well. It uses experience and previously learned concepts to learn new concepts. In fact, there are many layers going into a deep learning model. The reason why we call it deep learning has to do with the depth of the network of concepts that the model uses in making sense of new knowledge.

The reason why people are now turning to deep learning is that previous methods have failed. Many artificial intelligence projects have tried to create hard-coded knowledge of the world. This entrails trying to describe the subjective and intuitive considerations in formal and logic-based language. However, devising such rules is very difficult. The failings of these hard-coded knowledge projects suggests that a machine learning cannot rely on a static knowledge base. Rather, such a machine learning model must be able to generate knowledge on its own. It can do this by connecting previous knowledge. The model must be able to learn on its own!

Different modes of learning

One of the biggest challenges with deep learning is finding a way in which to represent the input data in such a way that the machine learning algorithm can draw benefit from it. Representation learning tries to discover mappings between representation to output as well as the representation itself. A quintessential example of representation learning are autoencoders. Autoencoders consists of an encoder function. This converts input data into a different representation (e.g., just like we encoded our dummy variables from qualitative classes to 0 and 1). There is also a decoder function. This converts the new representation back to the old one.

However, representation learning can only take us thus far. Consider a classic causal inference problem. We build a linear regression model with a set of regressors. We want to keep those regressors which are statistically significant, but omit those who aren't. We want to disentangle factors of variation (i.e., the regressors) and discard those we don't need. However, we have seen that issues such as collinearity, correlated errors etc. can impair our ability to do so. In cases such as this, in which it is difficult to disentangle factors of variation and discard insignificant ones, representation learning does not take us very far.

Deep learning seek to circumvent this problem by 'doing more of the work'. Instead of directly considering complex representations, it generates these complex representations as combinations of much simpler ones. Much like a typical 6-year old starts learning mathematics by learning how to count instead of being exposed to Lebesgue measures. Feed-forward deep neural networks, or multilayer perceptrons (MLP) aids in this process. A multilyaer perceptron is a mathematical function. Mathematical functions applies specific rules on input in order to generate some output. The function in questions is, itself, generated by composing many simpler functions. This results in a steady inflow of new representations.

Big Data

Machine learning isn't something new. In fact, most of the techniques have been around for decades! However, they have found newfound popularity in recent years because they are becoming more and more tractable as data sets keeps growing and growing. This means that it is no longer so difficult to generalize from sample to population. Furthermore, we have much better technology today. This makes it possible to estimate more and more advanced models such as deep neural networks. We simply have much more computational resources today than ever before. In fact, we even have programs which are able to learn how to program! Neural Turing machines can learn to read from and write to memory cells. This is clearly in its infancy, but it is promising. Reinforcement learning is another crowning achivement of deep learning.

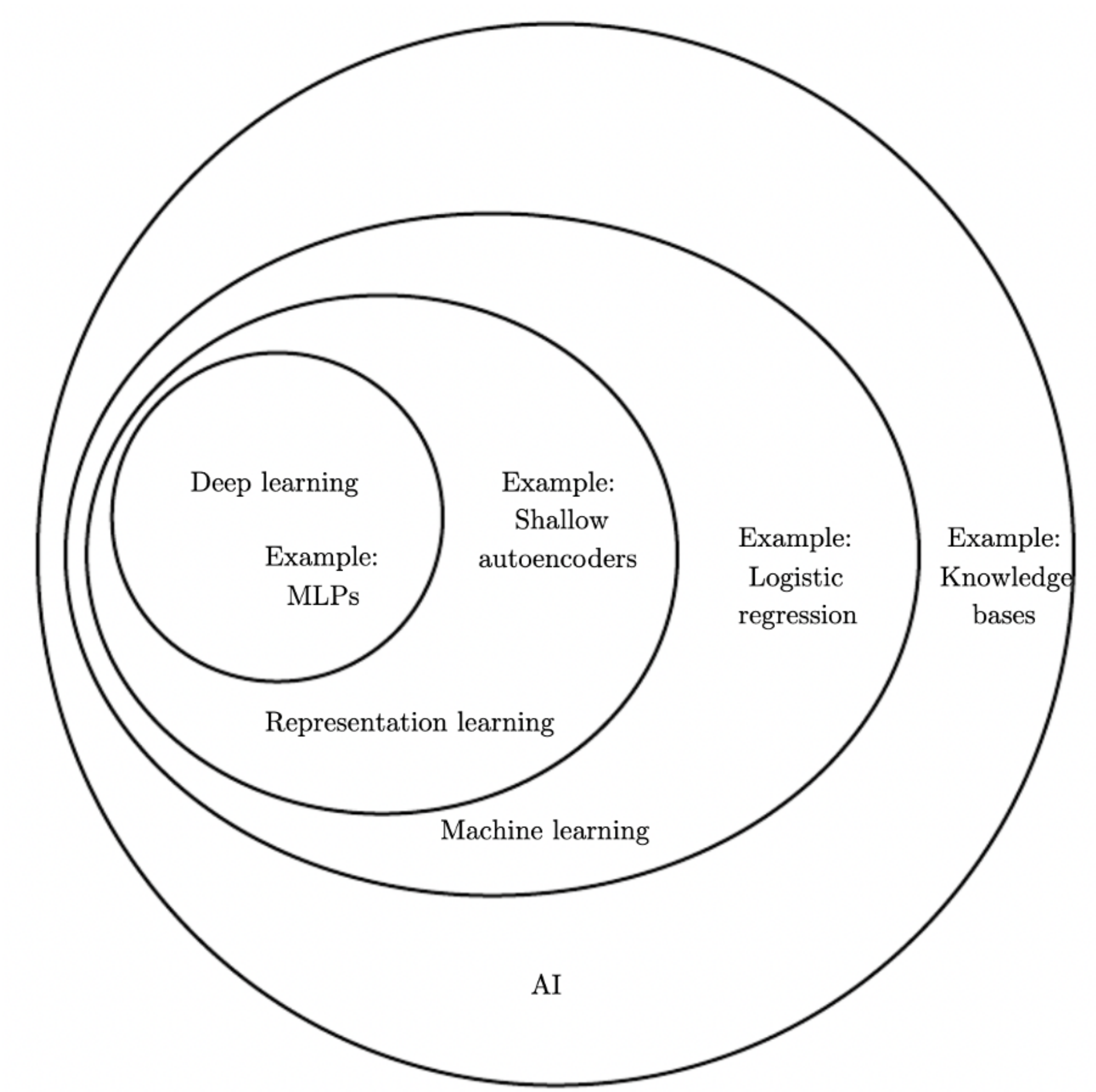

Figure 1: Venn diagram showing that deep learning is a kind of representation learning, which is a kind of machine learning which is a apart of AI. Source: Goodfellow, Bengio & Courville (2016).