Statistical Machine Learning vs Deep Learning

Introduction

There are three broad categories within machine learning: statistical models, deep learning or neural networks and reinforcement learning. This is a huge oversimplification, but it captures the general idea. In this article, I will discuss the differences between statistical and deep methods in machine learning, and leave reinforcement learning entirely since there is a clear demarcation in RL versus the other methods. Most of the confusion lies between statistical models and deep learning from my experience, and before writing this article, I myself hadn't given it too much thought either.

Defining Statistical and Deep Machine Learning

For the purposes of this article, we will draw a distinction between the two methods as gradient based methods and non gradient based methods. This might be slightly inaccurate since something like linear regression can be both statistical and deep (it's used in neural nets), we will think of it as a deep learning method according to our definition. This is by no means the absolute correct categorization, it's just the best way I could find to differentiate the two methods and since a comparison between the two methods requires a rigorous definition for both, this is what we'll work with.

Statistical vs Deep learning

The first and most obvious difference between the two is that statistical methods are usually a lot more interpretable than deep learning methods. For example, think of a decision tree and an MLP. By looking at the decision condition at each node in the tree, you can understand the reasoning behind the underlying classification or regression model whereas looking at the individual perceptrons in an MLP doesn't give you any intuition about what the underlying model really is.

The second, and one of the reasons for this article, is the predictive power of each type of model. Now ofcourse, it heavily depends on the architecture of the model in each category, for example a CNN will almost always outperform an MLP in vision tasks, but here we are contrasting the architectures in each category, for example a CNN outperforms statistical models in foreground detection in images. This is the type of problem we are concerning ourselves with. Which category of machine learning has a better method and/or architecture for that specific task. After learning about deep networks, it seems like there will always be a deep method that out-performs statistical methods in any given task, but this is, surprisingly (atleast surprising to me) far from the truth. Statistical methods seem to perform better on tasks that don't really have an underlying function that approximates their behaviour. One of the main reasons for my hypothesis is that neural networks are universal function approximators and

therefore, by definition, if there is some underlying function that approximates the behaviour of the system we are trying to predict, neural networks should have no problem doing so. To support my hypothesis, I tried to use both methods in various domains and found this example that supports my claim: [GITHUB LINK]. This is a stock market prediction project where I used off the shelf statistical methods as well as some fine tuned and thought out deep learning methods for stock price prediction, and found that the statistical methods consistently outperformed the deep learning methods. This was surprising to me, but it supported my hypothesis since the stock market seems to move randomly. It doesn't really make sense. This could be because of how influential financial institutions, like hedge funds, are in manipulating the markets and how most decisions are made because everyone else is making them rather than thinking about why they should be made. That could also be a personal bias I have stemming from the fact that I think most institutional investors can't think for themselves and just follow the herd, but that is a whole discussion in and of itself and not the point of this article. In any case, this shows the importance of statistical methods and why we can't rely on deep learning frameworks for every task.