The Deep Learning Revolution
Terrence J Sejnowski
MIT Press

Review: Andrew Baum

To a greater or lesser extent, we business, finance and real estate researchers are all empiricists, making a presumption that knowledge is based on experience and that data is essential to prove any hypothesis we might form about the way the world works. Of course, we know that this is flawed. Journals are full of papers that claim generalisations, formed from samples of data or experiences that are very specific to locations or time periods. We will use the present tense and claim that A causes B with 87% explanatory power, when we should say that A appears to have caused B over the past 15 years in the USA. Most of us will feel much better about making such generalisations if the finding supports, or is supported by, a normative theory to the effect that A *should* cause B. Usually, empirical tests are used to prove or disprove a prior (hence *a priori*) theory. Theory precedes empirical tests.

Artificial intelligence and machine learning, and in particular the deep learning approach to AI, present a challenging problem. Terrence Sejnowski's book sets up the empiricist v theorist debate in the context of machine learning. It would be (and was) natural for pioneers in AI to establish some theory-based rules about how a machine would become intelligent. But if the way to become intelligent – just as a baby becomes an intelligent child – is through pure experience, why use any theory? As the author says: information can be extracted from raw data; information can be used to create knowledge; knowledge leads to understanding; and understanding leads to wisdom. In other words, data precedes theory.

Theory and rules might be necessary where we are unable to collect enough data for our findings to be to be statistically significant, or where we lack appropriate statistical or econometric techniques. For example, panel regression became popular when it was realised that inadequate time series plus inadequate cross-sectional data could together form an adequate two-dimensional dataset. These theory-based rules are also important when the statistical or econometric model needs to be pre-specified, because (for example) e-views requires that the researcher not only specifies the form of the model she wants to test but also the dataset that will be used in the analysis. But what if there is so much data that it would take too long to examine the suitability of any one set? Or if data is multi-dimensional? Or if computing power were such that any number of models could be applied to any number of datasets to come up with the best possible fit?

Sejnowski uses the example of using images of faces to identify gender. If hair, jewellery, make up and so were neutralised, what purely facial features would identify a female? Before I read this book, I am not sure I would have any idea – I have no theory. Yet participants in a test achieved 88% accuracy, without being able to form any theory as to how.   And a computer achieved roughly the same result – in1995. The computer is able to establish how (the philtrum is larger in males, while upper cheeks are larger in females and so on). So the theory or rule emerges from pure experience, just as children learn and just, as the author argues, as machine learning is now working.

While the most advanced computer is still much less complex or powerful than the human brain, it is now sufficiently advanced to be able to handle very large datasets and to test many different models to explain the relationships within that data. No theory is necessary. Scrape as much data as you like from public sources or using APIs (application programming interfaces), load the data, write a simple program in Python and press go.

Using this approach in financial economics or (say) applied real estate research is likely to be highly fruitful.

Why? As the world is changed by technology, new relationships are forming and new data is becoming available. For example, how does the intensity of on-line retail sales in any postcode relate to the value of retail space nearby? How do Airbnb occupancy rates explain house prices or residential rents? The data is new; no theory has yet formed to connect the datasets.

If this is true, the next generation of research methods after 30 years of stagnation formed around multiple linear regression is already upon us. Machine learning based on data mining will lead us to more significant relationships and new theories about the way the world works. This book is a true revelation of the world that awaits us.