

# *Machine learning, Deep Learning, and (R | Python)*

Mark Andrews  
Psychology Department, Nottingham Trent University

`mark.andrews@ntu.ac.uk`

# What is machine learning?

- ▶ Put very informally, machine learning is a means to engineer knowledge or behaviour into a computer or machine by providing examples rather than explicit programming.
- ▶ For example, in an email spam filter program, rather than explicitly programming rules that determine whether an email is *spam* or *ham*, we can use a machine learning algorithm and provide large number of examples of *spam* and *ham* and let the program learn the rules.
- ▶ Machine learning can be much more efficient at learning the rules than relying on explicit programming.
- ▶ It can arguably acquire knowledge that is impossible to explicitly program.

# *Examples of machine learning*

- ▶ Recommender systems for YouTube, Amazon, Netflix
- ▶ Detecting credit card fraud
- ▶ Face identification, face recognition, object recognition
- ▶ Sentiment analysis
- ▶ Text classification and topic identification
- ▶ Natural language translation
- ▶ Game playing programs (e.g. AlphaGo)

# Machine learning “algorithms”

- ▶ Often machine learning “algorithms” are reinventions or extensions of traditional statistical methods, e.g. nonlinear regression.
- ▶ However, machine learning’s relationship with statistics is comparable to, say, computer science’s relationship with electrical engineering, or neuroscience with psychology, etc.
- ▶ Machine learning obviously differs from applied statistics in ways such as:
  - ▶ A focus on real world engineering (often tech) problems rather than (scientific) data analysis.
  - ▶ Focus on large real-time systems, and so with a major focus on efficiency (e.g compare machine learning with computational statistics using Stan).
  - ▶ A major focus on prediction and “behaviour” rather than developing scientific models.
  - ▶ Side-stepping preoccupations with classical statistics like hypothesis tests, p-values
  - ▶ A new terminology (e.g. *features, inputs, targets, training*, rather than say *predictor variables, outcome variables, estimation, inference*, etc.).

# *Subdivisions of machine learning*

- ▶ Machine learning is traditionally broken into three sub-divisions:
  - ▶ Supervised learning: Classification and regression, where explicit answers or “targets” corresponding to data examples are given.
  - ▶ Unsupervised learning: Statistical pattern identification and discovery. Data is provided and patterns are found. E.g. topic models, natural language grammar induction.
  - ▶ Reinforcement learning: Behaviours of an “agent” are rewarded or punished, leading to improved behaviours.

# *Deep learning*

- ▶ Deep learning is another name, arguably a rebranding, of artificial neural networks, a technology originating in the 1950s.
- ▶ The first major artificial neural network technology were the perceptrons. In the 1980s, there were extended to multilayer perceptrons, which are a type of nonlinear regression.
- ▶ In the late 2000s, many artificial neural net technologies were increasingly scaled up to ever more “layers”, hence “deep”.
- ▶ Artificial neural nets were a side-show in machine learning until around 2010. Since then, they’ve been widely celebrated (and hyped?).

# *Why deep learning, why now?*

- ▶ The following interrelated factors have all contributed to deep learning's growth:
  - ▶ Hardware. Deep learning models are very computationally demanding. GPUs are ideally suited to deep learning calculating and so cheap powerful GPUs have accelerated their performance.
  - ▶ Training data. Massive collections of texts, images, user data are now available.
  - ▶ Investment. Big tech, who often own the data, have invested heavily in deep learning.
  - ▶ Software. Tensorflow (Google), PyTorch (Facebook), Autograd have made using deep learning at scale easier.

# *Why Python, not R?*

- ▶ Why has Python become the lingua franca of machine learning?
- ▶ Arguably
  - ▶ It is widely taught in computer science.
  - ▶ It is a major language for the web back-end.
  - ▶ It has extensive scientific libraries.
- ▶ And then, as with R for computational statistics, it has benefited from a feedback loop of users and developers.
- ▶ But is R just as powerful and extensive for machine learning research, or the use of machine learning tools in scientific data analysis? Probably.