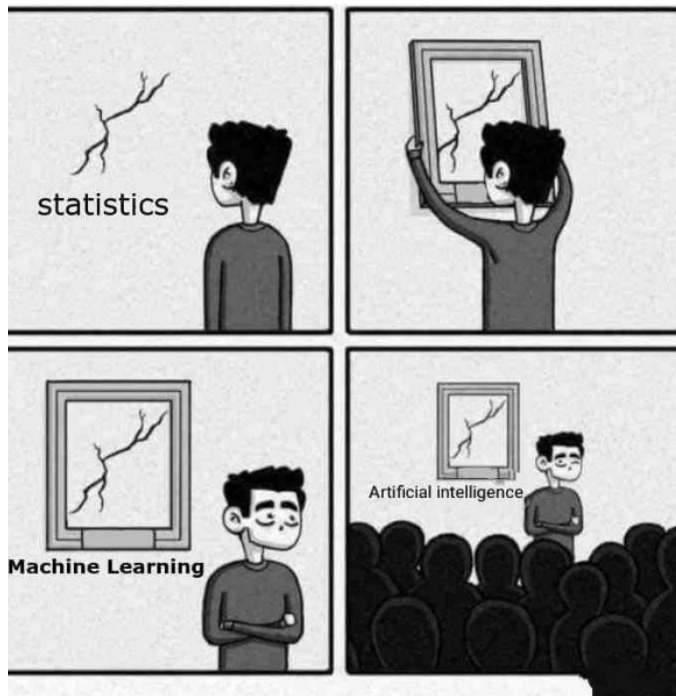**Joe Davison** [Follow]

I do ML @ Pluralsight & Harvard.

Jun 27 · 8 min read

No, Machine Learning is not just glorified Statistics

This meme has been all over social media lately, producing appreciative chuckles across the internet as the hype around deep learning begins to subside. The sentiment that machine learning is really nothing to get excited about, or that it's just a redressing of age-old statistical techniques, is growing increasingly ubiquitous; the trouble is it isn't true.

I get it—it's not fashionable to be part of the overly enthusiastic, hype-drunk crowd of deep learning evangelists. ML experts who in 2013 preached deep learning from the rooftops now use the term only with a hint of chagrin, preferring instead to downplay the power of modern neural networks lest they be associated with the scores of people that still seem to think that `import keras` is the leap for every hurdle, and that they, in knowing it, have some tremendous advantage over their competition.

While it's true that deep learning has outlived its usefulness as a buzzword, as Yann LeCun put it, this overcorrection of attitudes has yielded an unhealthy skepticism about the progress, future, and usefulness of artificial intelligence. This is most clearly seen by the influx of discussion about a looming *AI winter*, in which AI research is prophesied to stall for many years as it has in decades past.

Yann LeCun
@ylecun

There is almost as much BS being written about a purported impending AI winter as there is around a purported impending AGI explosion.

10:15 PM - Jun 13, 2018

♡ 425   ◯ 127 people are talking about this

This purpose of this post isn't to argue against an AI winter, however. It is also not to argue that one academic group deserves the credit for deep learning over another; rather, it is to make the case that credit *is* due; that the developments seen go beyond big computers and nicer datasets; that machine learning, with the recent success in deep neural networks and related work, represents the world's foremost frontier of technological progress.

Machine Learning != Statistics

> *"When you're fundraising, it's AI. When you're hiring, it's ML. When you're implementing, it's logistic regression."*

> *—everyone on Twitter ever*

The main point to address, and the one that provides the title for this post, is that machine learning is not just glorified statistics—the same-old stuff, just with bigger computers and a fancier name. This notion comes from statistical concepts and terms which are prevalent in machine learning such as *regression*, *weights, biases, models,* etc. Additionally, many models approximate what can generally be considered statistical functions: the softmax output of a classification model consists of logits, making the process of training an image classifier a *logistic regression*.

Though this line of thinking is technically correct, reducing machine learning as a whole to nothing more than a subsidiary of statistics is quite a stretch. In fact, the comparison doesn't make much sense. Statistics is the field of mathematics which deals with the understanding and interpretation of data. Machine learning is nothing more than a class of computational algorithms (hence its emergence from computer science). In many cases, these algorithms are completely useless in aiding with the understanding of data and assist only in certain types of uninterpretable predictive modeling. In some cases, such as in reinforcement learning, the algorithm may not use a pre-existing dataset at all. Plus, in the case of image processing, referring to images as *instances* of a dataset with pixels as *features* was a bit of a stretch to begin with.

The point, of course, is not that computer scientists should get all the credit or that statisticians should not; like any field of research, the contributions that led to today's success came from a variety of academic disciplines, statistics and mathematics being first among them. However, in order to correctly evaluate the powerful impact and potential of machine learning methods, it is important to first dismantle the misguided notion that modern developments in artificial intelligence are nothing more than age-old statistical techniques with bigger computers and better datasets.

Machine Learning Does Not Require An Advanced Knowledge of Statistics

Hear me out. When I was learning the ropes of machine learning, I was lucky enough to take a fantastic class dedicated to deep learning techniques that was offered as part of my undergraduate computer science program. One of our assigned projects was to implement and train a Wasserstein GAN in TensorFlow.

At this point, I had taken only an introductory statistics class that was a required general elective, and then promptly forgotten most of it. Needless to say, my statistical skills were not very strong. Yet, I was able to read and understand a paper on a state-of-the-art generative machine learning model, implement it from scratch, and generate quite

convincing fake images of non-existent individuals by training it on the MS Celebs dataset.

Throughout the class, my fellow students and I successfully trained models for cancerous tissue image segmentation, neural machine translation, character-based text generation, and image style transfer, all of which employed cutting-edge machine learning techniques invented only in the past few years.

Yet, if you had asked me, or most of the students in that class, how to calculate the variance of a population, or to define marginal probability, you likely would have gotten blank stares.

That seems a bit inconsistent with the claim that AI is just a rebranding of age-old statistical techniques.

True, an ML expert probably has a stronger stats foundation than a CS undergrad in a deep learning class. Information theory, in general, requires a strong understanding of data and probability, and I would certainly advise anyone interested in becoming a Data Scientist or Machine Learning Engineer to develop a deep intuition of statistical concepts. But the point remains: *If machine learning is a subsidiary of statistics, how could someone with virtually no background in stats develop a deep understanding of cutting-edge ML concepts?*

It should also be acknowledged that many machine learning algorithms require a stronger background in statistics and probability than do most neural network techniques, but even these approaches are often referred to as _statistical machine learning_ or _statistical learning_, as if to distinguish themselves from the regular, less statistical kind. Furthermore, most of the hype-fueling innovation in machine learning in recent years has been in the domain of neural networks, so the point is irrelevant.

Of course, machine learning doesn't live in a world by itself. Again, in the real world, anyone hoping to do cool machine learning stuff is probably working on data problems of a variety of types, and therefore needs to have a strong understanding of statistics as well. But that doesn't mean they're the same thing.

Machine Learning = Representation + Evaluation + Optimization

To be fair to myself and my classmates, we all had a strong foundation in algorithms, computational complexity, optimization approaches, calculus, linear algebra, and even some probability. All of these, I would argue, are more relevant to the problems we were tackling than knowledge of advanced statistics.

Machine learning is a class of computational algorithms which iteratively "learn" an approximation to some function. Pedro Domingos, a professor of computer science at the University of Washington, laid out three components that make up a machine learning algorithm: representation, evaluation, and optimization.

Representation involves the transformation of inputs from one space to another more useful space which can be more easily interpreted. Think of this in the context of a Convolutional Neural Network. Raw pixels are not useful for distinguishing a dog from a cat, so we transform them to a more useful representation (e.g., logits from a softmax output) which can be interpreted and evaluated.

Evaluation is essentially the loss function. How effectively did your algorithm transform your data to a more useful space? How closely did your softmax output resemble your one-hot encoded labels (classification)? Did you correctly predict the next word in the unrolled text sequence (text RNN)? How far did your latent distribution diverge from a unit Gaussian (VAE)? These questions tell you how well your representation function is working; more importantly, they define *what it will learn to do.*

Optimization is the last piece of the puzzle. Once you have the evaluation component, you can *optimize* the *representation function* in order to improve your *evaluation metric*. In neural networks, this usually means using some variant of stochastic gradient descent to update the weights and biases of your network according to some defined loss function. And voila! You have the world's best image classifier (at least, if you're Geoffrey Hinton in 2012, you do).

When training an image classifier, it's quite irrelevant that the learned representation function has logistic outputs, except for in defining an appropriate loss function. Borrowing statistical terms like *logistic regression* do give us useful vocabulary to discuss our model space, but they do not redefine them from problems of optimization to problems of data understanding.

Aside: The term artificial intelligence is stupid. An AI problem is just a problem that computers aren't good at solving yet. In the 19th century, a mechanical calculator was considered intelligent (link). Now that the term has been associated so strongly with deep learning, we've started saying artificial general intelligence (AGI) to refer to anything more intelligent than an advanced pattern matching mechanism. Yet, we still don't even have a consistent definition or understanding of general intelligence. The only thing the term AI does is inspire fear of a so-called "singularity" or a terminator-like killer robot. I wish we could stop using such an empty, sensationalized term to refer to real technological techniques.

Techniques For Deep Learning

Further defying the purported statistical nature of deep learning is, well, almost all of the internal workings of deep neural networks. Fully connected nodes consist of weights and biases, sure, but what about convolutional layers? Rectifier activations? Batch normalization? Residual layers? Dropout? Memory and attention mechanisms?

These innovations have been central to the development of high-performing deep nets, and yet they don't remotely line up with traditional statistical techniques (probably because they are not statistical techniques at all). If you don't believe me, try telling a statistician that your model was overfitting, and ask them if they think it's a good idea to randomly drop half of your model's 100 million parameters.

And let's not even talk about model interpretability.

Regression Over 100 Million Variables — No Problem?

Let me also point out the difference between deep nets and traditional statistical models by their scale. Deep neural networks are huge. The VGG-16 ConvNet architecture, for example, has approximately 138 million parameters. How do you think your average academic advisor would respond to a student wanting to perform a multiple regression of over *100 million variables*? The idea is ludicrous. That's because training VGG-16 is not multiple regression—it's machine learning.

New Frontiers

You've probably spent the last several years around endless papers, posts, and articles preaching the cool things that machine learning can now do, so I won't spend too much time on it. I will remind you, however, that not only is deep learning more than previous techniques, it has enabled to us address an ***entirely new class of problems.***

Prior to 2012, problems involving unstructured and semi-structured data were challenging, at best. Trainable CNNs and LSTMs alone were a huge leap forward on that front. This has yielded considerable progress in fields such as computer vision, natural language processing, speech transcription, and has enabled huge improvement in technologies like face recognition, autonomous vehicles, and conversational AI.

It's true that most machine learning algorithms ultimately involve fitting a model to data—from that vantage point, it is a statistical procedure. It's also true that the space shuttle was ultimately just a flying machine with wings, and yet we don't see memes mocking the excitement around NASA's 20th century space exploration as an overhyped rebranding of the airplane.

As with space exploration, the advent of deep learning did not solve all of the world's problems. There are still significant gaps to overcome in many fields, especially within "artificial intelligence". That said, it has made a significant contribution to our ability to attack problems with complex unstructured data. Machine learning continues to represent the world's frontier of technological progress and innovation. It's much more than a crack in the wall with a shiny new frame.

. . .

If you're like me and enjoy connecting with fellow ML aficionados, feel free to add me on Twitter and/or LinkedIn!