

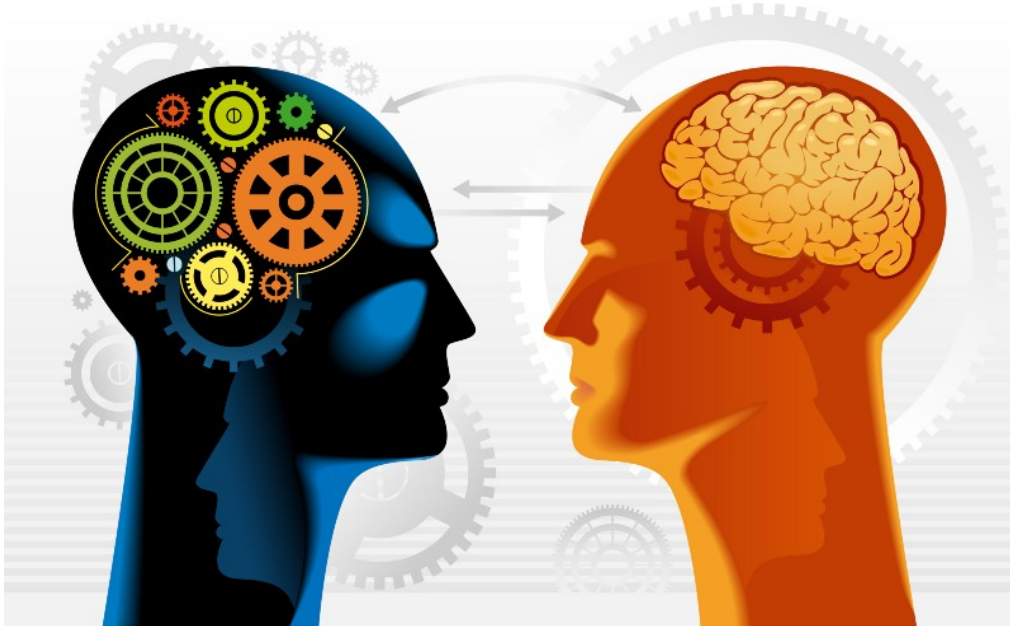


Pedro Domingos

[Follow](#)

Professor of computer science at U. Washington and author of "The Master Algorithm". pedrod...  
Sep 14 · 5 min read

## Ten Myths About Machine Learning



Machine learning used to take place behind the scenes: Amazon mined your clicks and purchases for recommendations, Google mined your searches for ad placement, and Facebook mined your social network to choose which posts to show you. But now machine learning is on the front pages of newspapers, and the subject of heated debate. Learning algorithms drive cars, translate speech, and win at *Jeopardy!* What can and can't they do? Are they the beginning of the end of privacy, work, even the human race? This growing awareness is welcome, because machine learning is a major force shaping our future, and we need to come to grips with it. Unfortunately, several misconceptions have grown up around it, and dispelling them is the first step. Let's take a quick tour of the main ones:

***Machine learning is just summarizing data.*** In reality, the main purpose of machine learning is to predict the future. Knowing the movies you watched in the past is only a means to figuring out which ones you'd like to watch next. Your credit record is a guide to whether you'll pay your bills on time. Like robot scientists, learning algorithms formulate hypotheses, refine them, and

only believe them when their predictions come true. Learning algorithms are not yet as smart as scientists, but they're millions of times faster.

***Learning algorithms just discover correlations between pairs of events.*** This is the impression you get from most mentions of machine learning in the media. In one famous example, an increase in Google searches for “flu” is an early sign that it's spreading. That's all well and good, but most learning algorithms discover much richer forms of knowledge, such as the rule *If a mole has irregular shape and color and is growing, then it may be skin cancer.*

***Machine learning can only discover correlations, not causal relationships.*** In fact, one of the most popular types of machine learning consists of trying out different actions and observing their consequences—the essence of causal discovery. For example, an e-commerce site can try many different ways of presenting a product and choose the one that leads to the most purchases. You've probably participated in thousands of these experiments without knowing it. And causal relationships can be discovered even in some situations where experiments are out of the question, and all the computer can do is look at past data.

***Machine learning can't predict previously unseen events, a.k.a. “black swans.”*** If something has never happened before, its predicted probability must be zero—what else could it be? On the contrary, machine learning is the art of predicting rare events with high accuracy. If A is one of the causes of B and B is one of the causes of C, A can lead to C, even if we've never seen it happen before. Every day, spam filters correctly flag freshly concocted spam emails. Black swans like the housing crash of 2008 were in fact widely predicted—just not by the flawed risk models most banks were using at the time.

***The more data you have, the more likely you are to hallucinate patterns.*** Supposedly, the more phone records the NSA looks at, the more likely it is to flag an innocent as a potential terrorist because he accidentally matched a terrorist detection rule. Mining more attributes of the same entities can indeed increase the risk of hallucination, but machine learning experts are very good at keeping it to a minimum. On the other hand, mining more entities with the same set of attributes decreases the risk, because the rules learned from them will have stronger support. And some learning algorithms can find patterns involving multiple entities, which makes them even more robust: a person videotaping the New York City Hall may not be suspicious, and another buying large quantities of ammonium nitrate may not be either;

but if the two are in close phone contact, perhaps the FBI should take a look, just to make sure it's not a bomb plot.

***Machine learning ignores preexisting knowledge.*** Experts in many fields that machine learning has permeated look askance at the “blank slate” approach of the learning algorithms they know. Real knowledge is the result of a long process of reasoning and experimentation, which you can't mimic by running a generic algorithm on a database. But not all learning algorithms start with a blank slate; some use data to refine a preexisting body of knowledge, which can be quite elaborate, provided it's encoded in a form the computer can understand.

***The models computers learn are incomprehensible to humans.*** This is naturally a cause for concern. If a learning algorithm is a black box, how can we trust its recommendations? Some types of models are indeed very hard to understand, like the deep neural networks responsible for some of machine learning's most notable successes (like recognizing cats in YouTube videos). But others are quite intelligible, like the rule for diagnosing skin cancer we saw earlier.

All of these myths are pessimistic, in the sense that they assume machine learning to be more limited than it really is. But there are also some optimistic myths around:

***Simpler models are more accurate.*** This belief is sometimes equated with Occam's razor, but the razor only says that simpler explanations are preferable, not why. They're preferable because they're easier to understand, remember, and reason with. Sometimes the simplest hypothesis consistent with the data is less accurate for prediction than a more complicated one. Some of the most powerful learning algorithms output models that seem gratuitously elaborate—sometimes even continuing to add to them after they've perfectly fit the data—but that's how they beat the less powerful ones.

***The patterns computers discover can be taken at face value.*** If a learning algorithm outputs the rule for skin cancer diagnosis we saw earlier and the rule is very accurate (in the sense that almost all the moles that match it are indeed tumors), that doesn't necessarily mean you should believe it. A slight change in the data could cause the algorithm to induce a very different—but equally accurate—rule. Only rules that are reliably induced despite random variations in the data can be trusted to mean what they say, as opposed to just being useful tools for prediction.

***Machine learning will soon give rise to superhuman intelligence.*** From the daily news of AI's advances, it's easy to get the impression that computers are on the verge of seeing, speaking and reasoning as well as we do, after which they'll quickly leave us in the dust. We've certainly come a long way in the first fifty years of artificial intelligence, and machine learning is the main reason for its recent successes, but we have a much longer way to go. Computers can do many narrow tasks very well, but they still have no common sense, and no one really knows how to teach it to them.

So there you have it. Machine learning is both more powerful than we often assume it to be and less. What we make of it is up to us—provided we start with an accurate understanding of it.