Why Machines Learn

The Elegant Math Behind Modern AI

Anil Ananthaswamy

"Some books about the development of neural networks describe the underlying mathematics while others describe the social history. This book presents the mathematics in the context of the social history. It is a masterpiece. The author is very good at explaining the mathematics in a way that makes it available to people with only a rudimentary knowledge of the field, but he is also a very good writer who brings the social history to life."

—**GEOFFREY HINTON,** deep learning pioneer, Turing Award winner, former VP at Google, and professor emeritus at the University of Toronto

"After just a few minutes of reading *Why Machines Learn*, you'll feel your own synaptic weights getting updated. By the end you will have achieved your own version of deep learning—with deep pleasure and insight along the way."

—**STEVEN STROGATZ,** *New York Times* bestselling author of *Infinite Powers* and professor of mathematics at Cornell University

"If you were looking for a way to make sense of the AI revolution that is well under way, look no further. With this comprehensive yet engaging book, Anil Ananthaswamy puts it all into context, from the origin of the idea and its governing equations to its potential to transform medicine, quantum physics —and virtually every aspect of our life. An essential read for understanding both the possibilities and limitations of artificial intelligence."

—**SABINE HOSSENFELDER,** physicist and *New York Times* bestselling author of *Existential Physics: A Scientist's Guide to Life's Biggest Questions*

"*Why Machines Learn* is a masterful work that explains—in clear, accessible, and entertaining fashion—the mathematics underlying modern machine learning, along with the colorful history of the field and its pioneering

researchers. As AI has increasingly profound impacts in our world, this book will be an invaluable companion for anyone who wants a deep understanding of what's under the hood of these often inscrutable machines."

—**MELANIE MITCHELL,** author of *Artificial Intelligence* and professor at the Santa Fe Institute

"Generative AI, with its foundations in machine learning, is as fundamental an advance as the creation of the microprocessor, the internet, and the mobile phone. But almost no one, outside of a handful of specialists, understands how it works. Anil Ananthaswamy has removed the mystery by giving us a gentle, intuitive, and human-oriented introduction to the math that underpins this revolutionary development."

—**PETER E. HART,** AI pioneer, entrepreneur, and co-author of *Pattern Classification*

"Anil Ananthaswamy's *Why Machines Learn* embarks on an exhilarating journey through the origins of contemporary machine learning. With a captivating narrative, the book delves into the lives of influential figures driving the AI revolution while simultaneously exploring the intricate mathematical formalism that underpins it. As Anil traces the roots and unravels the mysteries of modern AI, he gently introduces the underlying mathematics, rendering the complex subject matter accessible and exciting for readers of all backgrounds."

—**BJÖRN OMMER,** professor at the Ludwig Maximilian University of Munich and leader of the original team behind Stable Diffusion

Through Two Doors at Once

The Man Who Wasn't There

The Edge of Physics

*Data Communications Using Object-Oriented
Design and C++*

Why Machines Learn

The Elegant Math Behind Modern AI

Anil Ananthaswamy

DUTTON

CONTENTS

to teachers everywhere, sung and unsung

—

Whatever we do, we have to make our life vectors.
Lines with force and direction.

—LIAM NEESON AS FBI AGENT MARK FELT IN THE 2017 MOVIE

OF THE SAME NAME

Prologue

Buried on [this page](#) of the July 8, 1958, issue of *The New York Times* was a rather extraordinary story. The headline read, "New Navy Device Learns by Doing: Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser." The opening paragraph raised the stakes: "The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."

With hindsight, the hyperbole is obvious and embarrassing. But *The New York Times* wasn't entirely at fault. Some of the over-the-top talk also came from Frank Rosenblatt, a Cornell University psychologist and project engineer. Rosenblatt, with funding from the U.S. Office of Naval Research, had invented the perceptron, a version of which was presented at a press conference the day before the *New York Times* story about it appeared in print. According to Rosenblatt, the perceptron would be the "first device to think as the human brain" and such machines might even be sent to other planets as "mechanical space explorers."

None of this happened. The perceptron never lived up to the hype. Nonetheless, Rosenblatt's work was seminal. Almost every lecturer on artificial intelligence (AI) today will harken back to the perceptron. And that's justified. This moment in history—the arrival of large language models (LLMs) such as ChatGPT and its ilk and our response to it—which some have likened to what it must have felt like in the 1910s and '20s, when physicists were confronted with the craziness of quantum mechanics, has its roots in research initiated by Rosenblatt. There's a line in the *New York Times* story

that only hints at the revolution the perceptron set in motion: "Dr. Rosenblatt said he could explain *why the machine learned* only in highly technical terms" (italics mine). The story, however, had none of the "highly technical" details.

This book does. It tackles the technical details. It explains the elegant mathematics and algorithms that have, for decades, energized and excited researchers in "machine learning," a type of AI that involves building machines that can learn to discern patterns in data without being explicitly programmed to do so. Trained machines can then detect similar patterns in new, previously unseen data, making possible applications that range from recognizing pictures of cats and dogs to creating, potentially, autonomous cars and other technology. Machines can learn because of the extraordinary confluence of math and computer science, with more than a dash of physics and neuroscience added to the mix.

Machine learning (ML) is a vast field populated by algorithms that leverage relatively simple math that goes back centuries, math one learns in high school or early in college. There's, of course, elementary algebra. Another extremely important cornerstone of machine learning is calculus, co-invented by no less a polymath than Isaac Newton. The field also relies heavily on the work of Thomas Bayes, the eighteenth-century English statistician and minister who gave us the eponymous Bayes's theorem, a key contribution to the field of probability and statistics. The work of German mathematician Carl Friedrich Gauss on the Gaussian distribution (and the bell-shaped curve) also permeates machine learning. Then there's linear algebra, which forms the backbone of machine learning. The earliest exposition of this branch of mathematics appears in a two-thousand-year-old Chinese text, *Nine Chapters on the Mathematical Art.* The modern version of linear algebra has its roots in the work of many mathematicians, but mainly Gauss, Gottfried Wilhelm Leibniz, Wilhelm Jordan, Gabriel Cramer, Hermann Günther Grassmann, James Joseph Sylvester, and Arthur Cayley.

By the mid-1850s, some of the basic math that would prove necessary to building learning machines was in place, even as other mathematicians continued developing more relevant mathematics and birthed and advanced the field of computer science. Yet, few could have dreamed that such early mathematical work would be the basis for the astounding developments in AI over the past half century, particularly over the last decade, some of which may legitimately allow us to envision a semblance of the kind of future Rosenblatt was overoptimistically foreshadowing in the 1950s.

This book tells the story of this journey, from Rosenblatt's perceptron to modern-day deep neural networks, elaborate networks of computational units called artificial neurons, through the lens of key mathematical ideas underpinning the field of machine learning. It eases gently into the math and then, ever so slowly, ratchets up the difficulty, as we go from the relatively simple ideas of the 1950s to the somewhat more involved math and algorithms that power today's machine learning systems.

Hence, we will unabashedly embrace equations and concepts from at least four major fields of mathematics—linear algebra, calculus, probability and statistics, and optimization theory—to acquire the minimum theoretical and conceptual knowledge necessary to appreciate the awesome power we are bestowing on machines. It is only when we understand the inevitability of learning machines that we will be prepared to tackle a future in which AI is ubiquitous, for good and for bad.

Getting under the mathematical skin of machine learning is crucial to our understanding of not just the power of the technology, but also its limitations. Machine learning systems are already making life-altering decisions for us: approving credit card applications and mortgage loans, determining whether a tumor is cancerous, predicting the prognosis for someone in cognitive decline (will they go on to get Alzheimer's?), and deciding whether to grant someone bail. Machine learning has permeated science, too: It is influencing chemistry, biology, physics, and everything in between. It's being used in the study of genomes, extrasolar planets, the intricacies of quantum systems, and much more. And as of this writing, the

world of AI is abuzz with the advent of large language models such as ChatGPT. The ball has only just gotten rolling.

We cannot leave decisions about how AI will be built and deployed solely to its practitioners. If we are to effectively regulate this extremely useful, but disruptive and potentially threatening, technology, another layer of society—educators, politicians, policymakers, science communicators, or even interested consumers of AI—must come to grips with the basics of the mathematics of machine learning.

In her book *Is Math Real?,* mathematician Eugenia Cheng writes about the gradual process of learning mathematics: "It can...seem like we're taking very small steps and not getting anywhere, before suddenly we look behind us and discover we've climbed a giant mountain. All these things can be disconcerting, but accepting a little intellectual discomfort (or sometimes a lot of it) is an important part of making progress in math."

Fortunately, the "intellectual discomfort" in store for us is eminently endurable and more than assuaged by the intellectual payoff, because underlying modern ML is some relatively simple and elegant math—a notion that's best illustrated with an anecdote about Ilya Sutskever. Today, Sutskever is best known as the co-founder of OpenAI, the company behind ChatGPT. More than a decade ago, as a young undergraduate student looking for an academic advisor at the University of Toronto, Sutskever knocked on Geoffrey Hinton's door. Hinton was already a well-known name in the field of "deep learning," a form of machine learning, and Sutskever wanted to work with him. Hinton gave Sutskever some papers to read, which he devoured. He remembers being perplexed by the simplicity of the math, compared to the math and physics of his regular undergrad coursework. He could read these papers on deep learning and understand powerful concepts. "How can it be that it's so simple...so simple that you can explain it to high school students without too much effort?" he told me. "I think that's actually miraculous. This is also, to me, an indication that we are probably on the right track. [It can't] be a coincidence that such simple concepts go so far."

Of course, Sutskever already had sophisticated mathematical chops, so what seemed simple to him may not be so for most of us, including me. But let's see.

This book aims to communicate the conceptual simplicity underlying ML and deep learning. This is not to say that everything we are witnessing in AI now—in particular, the behavior of deep neural networks and large language models—is amenable to being analyzed using simple math. In fact, the denouement of this book leads us to a place that some might find disconcerting, though others will find it exhilarating: These networks and AIs seem to flout some of the fundamental ideas that have, for decades, underpinned machine learning. It's as if empirical evidence has broken the theoretical camel's back in the same way experimental observations of the material world in the early twentieth century broke classical physics; we need something new to make sense of the brave new world awaiting us.

As I did the research for this book, I observed a pattern to my learning that reminded me of the way modern artificial neural networks learn: With each pass the algorithm makes through data, it learns more about the patterns that exist in that data. One pass may not be enough; nor ten; nor a hundred. Sometimes, neural networks learn over tens of thousands of iterations through the data. This is indeed the way I grokked the subject in order to write about it. Each pass through some corner of this vast base of knowledge caused some neurons in my brain to make connections, literally and metaphorically. Things that didn't make sense the first or second time around eventually did upon later passes.

I have used this technique to help readers make similar connections: I found myself repeating ideas and concepts over the course of writing this book, sometimes using the same phrasing or, at times, a different take on the same concept. These repetitions and rephrasings are intentional: They are one way that most of us who are not mathematicians or practitioners of ML can come to grips with a paradoxically simple yet complex subject. Once an idea is exposed, our brains might see patterns and make connections when

encountering that idea elsewhere, making more sense of it than would have been possible at first blush.

I hope your neurons enjoy this process as much as mine did.

Desperately Seeking Patterns

When he was a child, the Austrian scientist Konrad Lorenz, enamored by tales from a book called *The Wonderful Adventures of Nils*—the story of a boy's adventures with wild geese written by the Swedish novelist and winner of the Nobel Prize for Literature, Selma Lagerlöf—"yearned to become a wild goose." Unable to indulge his fantasy, the young Lorenz settled for taking care of a day-old duckling his neighbor gave him. To the boy's delight, the duckling began following him around: It had imprinted on him. "Imprinting" refers to the ability of many animals, including baby ducks and geese (goslings), to form bonds with the first moving thing they see upon hatching. Lorenz would go on to become an ethologist and would pioneer studies in the field of animal behavior, particularly imprinting. (He got ducklings to imprint on him; they followed him around as he walked, ran, swam, and even paddled away in a canoe.) He won the Nobel Prize for Physiology or Medicine in 1973, jointly with fellow ethologists Karl von Frisch and Nikolaas Tinbergen. The three were celebrated "for their discoveries concerning organization and elicitation of individual and social behavior *patterns*."

Patterns. While the ethologists were discerning them in the behavior of animals, the animals were detecting patterns of their own. Newly hatched ducklings must have the ability to make out or tell apart the properties of things they see moving around them. It turns out that ducklings can imprint not just on the first living creature they see moving, but on inanimate things as well. Mallard ducklings, for example, can imprint on a pair of moving objects that are similar in shape or color. Specifically, they imprint on the

relational concept embodied by the objects. So, if upon birth the ducklings see two moving red objects, they will later follow two objects of the same color (even if those latter objects are blue, not red), but not two objects of different colors. In this case, the ducklings imprint on the *idea* of similarity. They also show the ability to discern *dis*similarity. If the first moving objects the ducklings see are, for example, a cube and a rectangular prism, they will recognize that the objects have different shapes and will later follow two objects that are different in shape (a pyramid and a cone, for example), but they will ignore two objects that have the same shape.

Ponder this for a moment. Newborn ducklings, with the briefest of exposure to sensory stimuli, detect patterns in what they see, form abstract notions of similarity/dissimilarity, and then will recognize those abstractions in stimuli they see later and act upon them. Artificial intelligence researchers would offer an arm and a leg to know just how the ducklings pull this off.

While today's AI is far from being able to perform such tasks with the ease and efficiency of ducklings, it does have something in common with the ducklings, and that's the ability to pick out and learn about patterns in data. When Frank Rosenblatt invented the perceptron in the late 1950s, one reason it made such a splash was because it was the first formidable "brain-inspired" algorithm that could learn about patterns in data simply by examining the data. Most important, given certain assumptions about the data, researchers proved that Rosenblatt's perceptron will always find the pattern hidden in the data in a finite amount of time; or, put differently, the perceptron will converge upon a solution without fail. Such certainties in computing are like gold dust. No wonder the perceptron learning algorithm created such a fuss.

But what do these terms mean? What are "patterns" in data? What does "learning about these patterns" imply? Let's start by examining this table:

x1	x2	y
4	2	8
1	2	5
0	5	10
2	1	4

Each row in the table is a triplet of values for variables *x1, x2,* and *y.* There's a simple pattern hidden in this data: In each row, the value of *y* is related to the corresponding values of *x1* and *x2.* See if you can spot it before reading further.

In this case, with a pencil, paper, and a little effort one can figure out that *y* equals *x1* plus two times *x2.*

$$y = x1 + 2x2$$

A small point about notation: We are going to dispense with the multiplication sign ("×") between two variables or between a constant and a variable. For example, we'll write

$$2 \times x2 \text{ as } 2x2 \text{ and } x1 \times x2 \text{ as } x1x2$$

Ideally, we should write $2x2$ as $2x_2$ and $x1x2$ as $x_1 x_2$, with the variables subscripted. But we'll dispense with the subscripts, too, unless it becomes absolutely necessary to use them. (Purists will cringe, but this method helps keep our text less cluttered and easy on the eye; when we do encounter subscripts, read x_i as "x sub-i.") So, keep this in mind: If there's a symbol such as "*x*" followed by a digit such as "2," giving us *x2,* take the entire symbol to mean one thing. If a symbol (say, *x* or *x2*) is preceded by a number (say, 9), or by another symbol (say, *w1*), then the number and the symbol, or the two symbols, are being multiplied. So: