As [Judea Pearl] sees it, the state of the art in artificial intelligence today is merely a souped-up version of what machines could already do a generation ago: find hidden regularities in a large set of data. “All the impressive achievements of deep learning amount to just curve fitting,” he said recently. […]

The way you talk about curve fitting, it sounds like you’re not very impressed with machine learning [remarks the interviewer]. “No, I’m very impressed, because we did not expect that so many problems could be solved by pure curve fitting. It turns out they can.”

Judea Pearl as interviewed by Kevin Hartnett, QUANTA MAGAZINE (May 15, 2018)

Judea Pearl won the 2011 Turing Award, the “Nobel Prize for computer science,” for his work on probabilistic and causal reasoning. He describes machine learning as “just curve fitting,” the mechanical process of finding regularities in data. The term comes from draftsmen’s use of spline curves, flexible strips made from thin pieces of wood or metal or plastic, to draw smooth lines through a set of pins.

In this chapter, we posit a further, specific analogy. We posit an analogy between Pearl’s description of machine learning and Holmes’s view of law. According to Holmes, the main task in law is finding patterns in human experience; law should not be seen simply as an exercise in mathematical logic. Likewise, machine learning should be thought of as curve fitting, i.e. as finding regularities in large datasets, and not as algorithms that execute a series of logical steps.
We described in Chapter 2 why it is not helpful to view machine learning as an algorithm. It is not an adequate explanation of what makes machine learning the powerful tool it has become today to say that it is about executing a series of logical instructions, composed in a piece of programming code. To understand what makes machine learning distinctive one has to start with the role of datasets as input, a role we described in Chapter 3 above, and which may be analogized to Holmes’s view of the jurist’s experience. In this chapter we now examine pattern finding more closely, first in law then in machine learning, to see how far the analogy might go.

4.1 Pattern Finding in Law

Holmes said in *The Path of the Law* that identifying the law means “following the existing body of dogma into its highest generalizations.” Two years after *The Path*, Holmes described law as a proposition that emerges when certain “ideals of society have been strong enough to reach that final form of expression.” To describe the law as Holmes did is to call for “the scientific study of the morphology and transformation of human ideas in the law.” If the pattern is strong enough, then the proposition emerges, the shape becomes clear.

Holmes returned a number of times to this idea that law is to be identified in patterns in human nature and practice. In a Supreme Court judgment in 1904, he addressed the right of “title by prescription.” Under that right, a sustained and uncontested occupation of land can override a legal title to that land. Prescription is thus an example where the law explicitly recognizes that a pattern of reality on the ground is the law. Holmes described prescription like this:

Property is protected because such protection answers a demand of human nature, and therefore takes the place of a fight. But that demand is not founded more certainly by creation or discovery than it is by the lapse of time, which gradually shapes the mind to expect and demand the continuance of what it actually and long has enjoyed, even if without right, and dissociates it from a like demand of even a right which long has been denied.

This way of describing title by prescription evoked the search for pattern in experience. How society actually behaves and how people think about that behavior are facts in which a pattern may be discerned. If the pattern
is well enough engrained, if it “shapes the mind” to a sufficient degree, and one knows how to discern it, then legal conclusions follow.

In what is perhaps his most famous dissenting opinion, that in *Lochner v. New York*, Holmes applied this idea about the ideals of society and the shape of the law a good deal further. The Supreme Court concluded that a New York state statute limiting the hours employees worked in a bakery violated the freedom of contract as embodied in the 14th Amendment. Holmes, as against his colleagues’ formal reading of the 14th Amendment, argued that one should interpret the constitutional right in the light of the patterns of belief discernible in society:

> Every opinion tends to become a law. I think that the word ‘liberty’ in the 14th Amendment, is perverted when it is held to prevent the natural outcome of a dominant opinion, unless it can be said that a rational and fair man necessarily would admit that the statute proposed would infringe fundamental principles as they have been understood by the traditions of our people and our law.⁶

In the land title case, the rule of title by prescription acknowledged that the pattern in human experience *is* the law. A formal rule, exceptionally, there corresponded to what Holmes thought law is. In *Lochner*, by contrast, there was no formal rule that says you are to interpret the 14th Amendment by reference to “dominant opinion.” The reading that Holmes arrived at in *Lochner* thus illustrates just how far-reaching Holmes’s conception of the law as a process of pattern finding was. Even the plain text of the law, which a logician might think speaks for itself, Holmes said calls for historical analysis. The meaning of a text is not to be found only in its words, but in the body of tradition and opinion around it: “A word [in the Constitution] is not a crystal, transparent and unchanged, but the skin of a living thought.”⁷ Holmes believed that we identify the law by systematically examining the shape of what exists already and what might later come—“the morphology and transformation of human ideas.”

A good jurist reaches decisions by discerning patterns of tradition and practice. The bad jurist treats cases as exercises in logical deduction. According to Holmes, “a page of history is worth a volume of logic.”⁸
4.2 **So Many Problems Can Be Solved by Pure Curve Fitting**

Judea Pearl expressed surprise that so many problems could be solved by curve fitting. And to someone from outside machine learning, it may seem preposterous that Holmes’s pattern finding might be analogous to drawing a line through a collection of points, as illustrated in the figure above. To give some idea of the scope of what machine learning researchers express as curve fitting, we now consider some applications. We have chosen applications from law and data to keep with our analogy to legal pattern finding—but curve fitting applications from any number of application areas, such as those from the data science careers website that we listed in Chapter 1, would support the same point.

Our first application relates to Holmes’s famous epigram “The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law.” Suppose it were possible to draw a chart summarizing the body of relevant case law. Each case would be assigned an $x$ coordinate encoding the characteristics of the case (the type of plea, the set of evidence, the history of the judge, and so on) and a $y$ coordinate encoding the outcome of the case (the decision reached, the sentence, and so on), and a point would be plotted for each case at its assigned $x$ and $y$ coordinates. We could then draw a smooth curve that expresses how the $y$ coordinate varies as a function of the $x$ coordinate—i.e. find the pattern in the dataset—and we could use this curve to predict the likely outcome of a new case given its $x$ coordinate.

This may sound preposterous, a law school version of plotting poets on a chalk board like the English teacher in *Dead Poets Society* did to ridicule
a certain kind of pedantry. However, it is an accurate description of how machines are able to accomplish such tasks as translating text or captioning images. A chalkboard has only two dimensions; a machine learning system works in many more dimensions, represented through mathematical functions. The coordinates are expressed in sophisticated geometrical spaces (instead of $x$, use $x_1, x_2, \ldots, x_n$ for some large number of dimensions $n$) that go beyond human visualization abilities; but the method is nothing more than high dimensional curve fitting.

The above application is a thought experiment. Here are some actual examples borrowed from a recent book on Law As Data:

(i) Predicting whether a bill receives floor action in the legislature, given the party affiliation of the sponsor and other features, as well as keywords in the bill itself.

(ii) Predicting the outcome of a parole hearing, given keywords that the inmate uses.

(iii) Predicting the case-ending event (dismissal, summary judgement, trial, etc.), given features of the lawsuit such as claim type or plaintiff race or plaintiff attorney’s dismissal rate.

(iv) Predicting the topic of a case (crime, civil rights, etc.) given the text of an opinion. (To a human with a modicum of legal training this is laughably simple, but for machine learning it is a great achievement to turn a piece of text into a numerical vector $(x_1, x_2, \ldots, x_n)$ that can be used as the $x$ coordinate for curve fitting. The mathematics is called “doc2vec”.)

(v) Predicting the decision of an asylum court judge given features of the case. (If a prediction can be made based on features revealed in the early stages of a case, and if the prediction does not improve when later features are included, then perhaps the judge was sleeping through the later stages.)

We have used the word “predict” for all of these examples. Most of these tasks are predictive in the sense of forecasting, but in the case (iv) the word “predict” might strike a layperson as odd. In machine learning, the word “predict” is used even when the outcome being predicted is already known; what matters is that the outcome is not known to the machine making the prediction. Philosophers use the words “postdiction” or “retrodiction” for such cases. In Chapter 5 we address in detail why
computer scientists use the language of prediction to describe the outputs of a machine learning system—and why Holmes used it to describe the outputs of law.

### 4.3 Noisy Data, Contested Patterns

Holmes’s wrote that “a page of history is worth a volume of logic.” When lawmakers ask for “the logic involved” in automated decision making, they should really be asking for “a story about the training dataset.” It is the data—that which is a given and thus came before—that matters in machine learning, just as the history is what matters in Holmes’s idea of law—not some formal process of logic.

But history can be contested. Even when parties agree on the facts, there may be multiple narratives that can be fitted. Likewise, for a given dataset, there may be various curves that may be fitted, as the figure above illustrates. We might wish to remove the subjectivity, leaving us with a volume of irrefutable logic proving that the decision follows necessarily from the premises, but that is the nature neither of law nor of machine learning. The phrase “story about the training dataset” is meant to remind us of this.

For some datasets, there may be a clear curve that fits all the data points very closely. In Holmes’s language, this corresponds to finding patterns in experience that have attained the “final form of expression.” The process of finding the law, as Holmes saw it, is the process of finding a pattern strong enough to support such “highest generalizations.” Not all “existing dogma” lends itself, however, to ready description as law; one does not always locate in the body of experience a “crystal, transparent.” Likewise, not all datasets have a well-fitting curve; the $y$ coordinates may simply be too noisy.

Some writers refer to machine learning systems as “inferring rules from data,” “deriving rules from data,” and the like. We recommend the
phrase “finding patterns in data,” because it is better here to avoid any suggestion of clean precise law-like rules. The patterns found by machine learning are not laws of nature like Newton’s laws of motion, and they are not precise stipulative rules in the sense of directives laid down in statutes. They are simply fitted curves; and if the data is noisy then the curves will not fit well.

While we have noted here that pattern finding is an element shared by machine learning and law, we should also note a difference. Law as Holmes saw it, and as it must be seen regardless of one’s legal philosophy, is an activity carried out by human beings. Law involves intelligence and thought. Machine learning is not thought. Once the human programmer has decided which class of curves to fit, the machine “learning” process is nothing more than a mechanical method for finding the best fitting curve within this class. Caution about anthropomorphizing machine learning is timely because there is so much of it, not just in popular culture, but in technical writing as well—and it obscures what machine learning really is. Machine learning is not thought. It is not intelligence. It is not brain activity. Pearl described it as curve fitting to emphasize this point, to make clear it is nothing more than a modern incarnation of the draftsman’s spline curve. That description does not entail any modesty at all about what machine learning can do. It only serves to illustrate how it is that machine learning does it.

**Notes**

1. Reprinted in *The Atlantic* (May 19, 2018).
2. 10 Harv. L. Rev. at 476.
3. Holmes, *Law in Science and Science in Law*, 12 Harv. L. Rev. 443, 444 (1898–1899).
4. Id. at 445.
5. Davis v. Mills, 194 U.S. 451, 457, 24 S.Ct. 692, 695 (Holmes, J., 1904).
6. *Lochner v. New York*, 198 U.S. 45, 75–76, 25 S.Ct. 539, 547 (Holmes, J., dissenting, 1905).
7. Towne v. Eisner, 245 U.S. 418, 425 (1918).
8. New York Trust Co. v. Eisner, 256 U.S. 345, 349 (1921).
9. 10 Harv. L. Rev. at 461.
10. The idea, which in the movie is read by a student from a fictive text entitled *Understanding Poetry* as the teacher draws a chart representing Byron and Shakespeare, is this:
If the poem’s score for perfection is plotted along the horizontal of a graph, and its importance is plotted on the vertical, then calculating the total area of the poem yields the measure of its greatness.

*Dead Poets Society* (release date: June 9, 1989): http://www.dailyscript.com/scripts/dead_poets_final.html. For a critique, see Kevin J.H. Dettmar, *Dead Poets Society Is a Terrible Defense of the Humanities*, *The Atlantic* (Feb. 19, 2014).

11. Taken from Livermore & Rockmore (eds.), *Law As Data. Computation, Text & the Future of Legal Analysis* (2019). The examples are from Vlad Eidelman, Anastassia Kornilova & Daniel Argyle, *Predicting Legislative Floor Action* (id. 117–50); Hannah Laqueur & Anna Venancio, *A Computational Analysis of California Parole Suitability Hearings* (id. 191–230); Charlotte S. Alexander, Khalifeh al Jadda, Mohammad Javad Feizhollahi & Anne M. Tucker, *Using Text Analytics to Predict Litigation Outcomes* (id. 271–308); Elliott Ash & Daniel Chen, *Case Vectors: Spatial Representations of the Law Using Document Embeddings* (id. 309–34); Daniel Chen, *Machine Learning and the Rule of Law* (id. 429–38). The latter application is a thought experiment, and the first two are implemented case studies.

12. As at GDPR Arts. 13(2)(f), 14(2)(g), and 15(1)(h).

13. For examples of how jurists have applied Holmes’s aphorism in disputable ways, see Bosmajian, 38 J. CHURCH STATE 397–409 (1996).

14. See for example Kroll et al. at 638.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.