Science: the Enterprise of Pattern Recognition

Tianxiao Shen

"The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions ..."

—*Pattern Recognition and Machine Learning*, Christopher M. Bishop

"There is ... a rhythm and a pattern between the phenomena of nature which is not apparent to the eye, but only to the eye of analysis; and it is these rhythms and patterns which we call Physical Laws"

—*The Character of Physical Law*, Richard Feynman

If we regard science as the human enterprise of understanding the natural world, we might regard computer science as the computer's enterprise of understanding the artificial world. We have language, memory and structure for our knowledge representation. Computers have their own language, memory, and structure, which comprise their "understanding". Careful investigation into epistemology and methodology in computer science will help us to apprehend scientific methods. By analyzing machine learning, this paper studies human learning in retrospect, including topics of inductivism, falsificationism, scientific realism, paradigms and incommensurability. We shall see that all these problems are linked up by patterns, both for machines and for humans.

Following is an outline of my thesis:

1. Induction and falsification resemble the training and testing process in machine learning. But Hume asserted that induction cannot be logically justified, and Popper even discarded it. As for falsification, Popper's justification is unsatisfactory to me. In machine learning, by assuming that seen and unseen observations are distributed according to a certain probability distribution, the generalization error caused by training and

testing is computable. Borrowing this idea, I will argue that if we presume patterns, i.e. recognizable regularities, induction and falsification are logically justifiable. [section 1]

2. Patterns are recognizable, but may be recognized in different forms. The patterns of digits are shapes of strokes in human eyes, or hyperplanes in SVMs, or weighted networks in NNs. They look totally different, but they are describing the same regularity. [section 2]

3. Computational model is like Kuhn's scientific paradigm, which has "incommensurability". When scientific theories become incommensurable, their reality and objectivity become questionable. Many philosophers explain the development of science by the theory of evolution, which is also unsatisfactory to me. In contrast, I think patterns could give a more clear explanation for scientific realism. [section 3,4]

4. In the end, by making an analogy between a scientific paradigm and an adaptive model, I want to argue that though paradigms' architectures are incommensurable, their adaptability to patterns is mensurable. [section 5]

1 Inductivism and Falsificationism

According to inductivism, which is one of the most influential methodologies of science, scientific inquiry has four stages: first we observe facts; we next draw inductive generalizations from them; we then try to find laws of nature from which the generalizations follow; we finally try to organize these laws into scientific theories. Theories should ideally be universal, which means that they can be applied to related situations, explaining known phenomena and predicting future outcomes. Extracting universal laws from empirical phenomena is the principle of science.

This inductive method has been considered logically unjustifiable since Hume, who held that there is no valid logical argument supporting that "instances of which we have had no experience resemble those of which we have had experience", and therefore "even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience." In the light of this view, theories can never be inferred from observations, because their universality or applicability about what we have had no experience of cannot be rationally justified.

Approving of Hume's rejection of inductivism, Popper proposed that falsifiability plays the central role in the logic of science, serving as "the criterion of demarcation between science and pseudoscience". In his view, science begins with hypotheses which are then tested by deriving particular observable consequences. Theories in principle can never be justified or verified. They can only be tentatively accepted, and they are rejected as long as falsifying evidence has been found.

Falsificationism is an important methodology in science, but I think Popper's denial of the function of inductivism is unacceptable. Popper claimed that "There is no need even to mention 'induction' ". The first two steps of science—observing facts and making inductive generalizations from them—to him do not exist. Scientists just *jump* from an observation statement to a theory. But how to jump to a good theory? Popper's response was "by jumping first to any theory ... repeatedly applying the critical method, eliminating many bad theories, and inventing many new ones." It seems to me that Popper's attitude towards rationality was vague and dubious. He thought that it is reasonable to believe that the future will be very different from the past, but it is also reasonable to act on the assumption that it will be like the past, since we can have no better assumption to act upon.

Popper's criterion for a good theory itself suffers from a form of the problem of induction, for there is no valid logical argument supporting that a theory which has withstood tests will continue to pass further tests or outperform falsified theories. That is to say, the best theory so far does not necessarily lead to a better prediction than a falsified theory when they are applied in a new situation. Consequently it is not reasonable to act upon the so-called "best theory so far". The connection between the known past and the unknown future is severed. No theory can guide our future action, whether it has been successful in the past or not. When he said we can have no better assumption to act upon, Popper failed to realize that inductively grounded action becomes, on his view, equivalent to acting randomly and irrationally.

A nice example—a common hoax of the internet era—will show that falsificationism fails in a random world. During the World Cup, many websites claim that they can predict the results of games based on big data analysis. To convince you, they tell you which teams will win in the next three matches—a falsifiable hypothesis. After you have been surprised by their magical foresight, they ask you for money in return for further predictions. You may luckily win if you bet according to their prediction. However, you will probably find that the big data analysis fails. What went wrong? Well, the answer is plain. These websites just enumerate all possible outcomes and show different predictions to different people. They will lose 7/8 of their potential customers, but nevertheless they can make money from the remaining 1/8. From this example we see that falsificationism does not give us a good theory. By first jumping to any theories and eliminating the falsified ones, the remaining theories can still be useless for future applications. If theories are uniformly random, the verified ones and falsified ones have no difference in any sense. Conversely, if the world is totally random, all theories are eqaully powerless. Say a person guessed right about the last ten coin tosses; will he or she be more sure about the next toss than other people?

Unlike predicting the World Cup, there are many tasks that computers have accomplished successfully, for instance spam filtering. As a human learns from experience, a machine learns from data. Here is big data analysis at its best[1]. It starts with a large dataset of N emails $\{x_1, \cdots, x_N\}$ called a *training set*. The categories (spam or not) of the emails in the training set are known in advance, typically by collecting labeling results from users. We can express the category of an email using variable t (it is called the target vector in machine learning terminology), so we get $\{t_1, \cdots, t_N\}$. The aim of machine learning is to get a function $y(x)$ which takes an email x as input and generates an output t. The precise form of the function $y(x)$ is determined during the training phase, also known as the learning phase, on the basis of the training data. Namely, $y(x)$ is an adaptive model whose parameters are calculated so that $\{y(x_1), \cdots, y(x_N)\}$ and $\{t_1, \cdots, t_N\}$ are matched in the best way. (Machine learning develops theories and methods on models, learning algorithms and criteria of matching degree. We omit technical details here and focus on methodology.)

[1]See Christopher M. Bishop, *Pattern Recognition and Machine Learning* (springer, 2006).

Compared to batch training, an online training procedure in which data becomes available in a sequential order is more like the process of induction. It starts from an initial model $y_0(x)$ which can take an email x as input and output whether it is spam or not. After receiving an email x_1 and its category t_1, the machine adjusts its model to $y_1(x)$ to fit this instance. Then it receives (x_2, t_2) and adjusts to $y_2(x)$ accordingly. It continues with $(x_3, t_3) \cdots (x_N, t_N)$ until it gets the final form $y_N(x)$. I think this is also the way a human makes an induction. He or she has an adaptive model in mind and modifies it according to experience step by step. Hume explained our engagement in induction by a custom or habit originating from the repeated observation that things of a certain kind are constantly conjoined with things of another kind. I think this successive and adaptive view of induction is more adequate: it allows that human can react before getting a vast number of repeated observations; also, it makes room for novelty.

Once the model is trained it can then determine the category of new emails, which are said to comprise a *test set*. The ability to categorize correctly new examples that differ from those used for training is known as *generalization*. Hence the test set is used to check the model's capability of generalization.

We can see both inductivism and falsificationism in the above machine learning process. Learning the model $y(x)$ from the training set is trying to inductively find some kind of regularity. Testing the model on a new dataset is a more sophisticated form of falsification—rather than rejecting a model so long as a counter-example emerges, a comprehensive evaluation is adopted. The model is not an arbitrary one got from jumping as Popper contended. As a matter of fact, jumping is not feasible: just imagine how hard it would be to jump to an effective model. The machine cannot merely guess what the parameters should be in the absence of the training set. On the contrary, induction does play a key role in its success. Meanwhile, the test set on behalf of falsificationism is indispensable as well. Otherwise a simple exhaustive model specifying t_i for each x_i will perform perfectly on the training set, but have no ability to classify other emails, and thus have no value in practical use.

Why are inductivism and falsificationism significant in scientific practice? Or in the context of machine learning, how can the training set lead to a good model whose success rate is much higher than a blind one, and how can the test set guarantee a verified model's further success in new applications? Theoretical machine learning analysis of *generalization error*, which is a measure of how accurately a model is able to predict outcomes for previously unseen data, is based on the assumption that all data pairs, including training data $\{(x_1, t_1), \cdots, (x_N, t_N)\}$, test data and any potential data, are independent and identically distributed with a certain probability distribution on space $\mathcal{X} \times \mathcal{T}$. All formulae, performance evaluations and error analyses are derived from and computed against that probability distribution. Namely, a *pattern* is presumed in a task, which is the origin and criterion of everything. The pattern is contained in training data and thus discovered and characterized by the trained model. The test set, as a set of samples from all potential data sharing the intrinsic pattern, verifies the captured pattern in the model.

I have argued that patterns help to justify our scientific methods, without which induction and falsification may degenerate to random behaviors. Now we need to call on an elucidation of patterns: what are they? Do they really exist?

2 Patterns

Hume held that the premise that nature has underlying patterns cannot be established by reasoning. It cannot be proved deductively, because underlying patterns are not logically necessary, so long as we need to find them by scientists' empirical efforts besides sitting and making trivial deductions. Nor can it be proved inductively, because induction is based upon the principle "that instances of which we have had no experience, must resemble those of which we have had experience", "that the course of nature continues always uniformly the same"; and then we get a *petitio principii*. So how can we justify the existence of patterns or the principle of uniformity of nature?

The assumption that patterns exist is vital in scientific practice. To me it should be clear that there are patterns, and this is my position of scientific realism. Regardless of how people would describe it as, a belief or a dogma or whatever, that there are patterns is the most reasonable belief and the most empirical dogma, so to speak. In contrast, the claim that there are no patterns sounds unreasonable and nonempirical at any rate. We are thrown into a world full of all sorts of phenomena; some have underlying patterns and some may not.

Sciences study the ones whose patterns are more apparent and easier to discover, and this is the reason why sciences have made some but not all (perhaps never all) progress. Isn't the success of science persuasive enough to demonstrate the empirical and reasonable existence of patterns?

(People may think it absurd to argue the existence of some entity without revealing it, or claim the knowledge of some information without demonstrating it. This is an outdated view. In cryptography, computer scientists come up with a method called *zero-knowledge proof*[2], by which the prover can prove to the verifier that a given statement is true, without conveying any information apart from the fact that the statement is indeed true. Proving the existence of an entity by showing it, or proving one has knowledge of certain information by revealing that information, is a way, but not the only way. If I could, I would like to give a zero-knowledge proof of patterns.)

I need to clarify the notion of pattern by giving an explicit definition, but before that let me give another machine learning example to provide some insight into it. Consider the widely applicable task of recognizing handwritten digits (it can be used in a postal system to recognize zip codes, for example). The machine will take an image as input and produce the identity of the digit $0 \cdots 9$ in it as output. Although handwriting can vary greatly, one may think that the task is tractable, as the patterns of digits are there distinctly. Yet apparently distinct patterns to human eyes could be unintelligible to computers, which have a totally different architecture of understanding. Computer scientists used to tackle this problem from a human perspective, analyzing and formalizing handcrafted rules and heuristics to distinguish the digits based on the shapes of the strokes. But such an approach leads to a proliferation of complicated rules and infinite exceptions, and invariably gives poor results in practice.

Emancipated from human patterns, the machine views the image as a pixel vector. Pre-processing may be exploited on original input vectors to help to refine the pattern. For instance, we can translate and rescale images so that each digit is contained within a box of a fixed size. After the location and scale

[2] See Uriel Feige, Amos Fiat and Adi Shamir, "Zero-knowledge Proofs of Identity", *Journal of cryptology*, (1988); Jean-Jacques Quisquater and Louis Guillou, "How to Explain Zero-knowledge Protocols to Your Children", In *Advances in Cryptology-CRYPTO'89 Proceedings*, (1990).

of all handwritten digits have been normalized, the subsequent recognition will become much easier, focusing on different essential patterns. Computer scientists have been coming up with various pre-processing methods, which are also called *feature* extraction. We get strokes and lines as features in our eyes, and computers have their own features in a quite different sense—new transformed vectors in the feature space.

Next is the main learning stage. There are diverse models whose theories and criteria are completely different from each other, or we may say they are "incommensurable". We shall come back to the incommensurability topic later in detail. Here I would like to briefly introduce two popular models: support vector machines (SVMs)[3] and neural networks (NNs)[4].

SVMs are primarily used for classification between two categories, like the previous spam filtering problem (though the handwritten digits recognition is a multiple-category classification problem, the core theory is the same.) We have already learnt to process input data into vectors in the feature space, and an SVM constructs a hyperplane separating the whole space into two subspaces, each of which corresponds to a category. By telling which subspace an input item falls into, it gives out the corresponding category as output. Consider a simple case in which all training data can be perfectly separated, i.e. there exist hyperplanes such that all vectors in one category fall into one side, and all vectors in the other category fall into the other side. Then which hyperplane is the best among many such candidates? Intuitively, a good one has the largest distance to the nearest training-data point of either category, and this distance is called the margin. The SVM algorithm presumes that the larger the margin the lower the generalization error of the classifier. SVMs have variants to deal with complex situations when no perfect separation exists, but let me skip them as we already get the key point. Hyperplane and margin are essential components in SVMs, and the parameters describing them constitute a pattern to be learnt.

Things are very different when it comes to neural networks. The idea

[3]See Corinna Cortes and Vladimir Vapnik, "Support-vector networks", *Machine Learning*, (1995).

[4]See Warren McCulloch and Walter Pitts, "A Logical Calculus of Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, (1943); Frank Rosenblatt, "The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review*, (1958).

of artificial neural networks in machine learning comes from biological neural networks of animals. Neural networks generally take the form of interconnected "neurons", i.e. a graph of linked nodes. Front neurons receive the input vector, back neurons produce the output, and intermediate neurons are hidden nodes. Each neuron performs some transforming function determined by the network designer, and exchanges information with other neurons through its links. Each link has a weight parameter on it which is determined based on training data. So the graph structure and transforming functions are designated by computer scientists as *prior architecture*. And as a pattern, the set of weight parameters is learnt from specific tasks.

Hopefully the above two models can shed some light on the definition of pattern, at least for machines. Since recognition has an inherent psychological dimension, *recognized patterns* cannot be mind-independent ("psychological" and "mind" here also refer to computers in a broad sense). For a machine, a pattern is a set of parameters which rests on the architecture of its model, *viz.* the hyperplane and margin in SVMs and the network of neurons in NNs. In other words, a pattern is internally formed in an adaptive model according to external phenomena. Likewise for humans, laws (of physics, biology, or engineering etc.) are patterns of nature recognized and depicted in our mind. This is reminiscent of the neo-Kantian view that the world investigated by the sciences is dependent on scientists' theoretical assumptions and perceptual training. But more than that, I would like to introduce an objective and external dimension of patterns, which has a close relation to paradigms, incommensurability and scientific realism.

3 Paradigms and Scientific Evolution

Kuhn's concept of a scientific paradigm (or a disciplinary matrix which he later suggested to use as a clarification of paradigm) is an ideal counterpart of a model in computer science. A paradigm consists of metaphysical assumptions, values, key theories, instruments, methods and problem-field. Normal scientific practice is a coherent tradition guided under a paradigm, through which puzzle-solutions are cumulatively generated and the scope and precision of scientific knowledge are steadily extended. Just as there are different world views, there are different paradigms. Alternation from one paradigm to another results in a shift in almost everything, including the scientific imagination, the problems available for scientific scrutiny and the standards of a legitimate problem-solution. Such an alternation is subversive to the previous paradigm's basic commitments, leading Kuhn to call it a scientific revolution, a transformation of the world within which scientific work was done. Transformations from Ptolemaic astronomy to Copernican astronomy, from Aristotelian dynamics to Newtonian mechanics to Einsteinian dynamics are all typical scientific revolutions.

The concept of a paradigm challenges scientific realism and inevitably leads to a pessimistic and skeptical attitude towards sciences. Kuhn held that we have no higher standard from which to evaluate different paradigms. He even drew a parallel between political and scientific revolutions that when they occur, political or scientific recourse fails. Different political parties acknowledge no supra-institutional framework for their adjudication, and as a final result, they resort to force. Similarly, the issue of paradigm choice can never be unequivocally settled by logic and experiment alone but has to appeal to factors external to science.

We should not simply accuse Kuhn of subjectivity and irrationality. A thorough discussion of his genuine intention is necessary. Though relinquishing the view that scientific development is a piecemeal process by which knowledge has been ever growing closer and closer to the truth, Kuhn held the alternative view that scientific development is a process of evolution which has no ultimate goal in advance. He did not believe in a full, objective, true account of nature by which paradigms are evaluated in terms of the closeness to it. But Kuhn did not reject that science is making progress, instead he considered this progress to be evolutionary, from present knowledge rather than toward future truth.

It is interesting that many philosophers of science ultimately resorted to Darwin's theory of evolution to explain the success or justify the development of science, from Popper to Kuhn to van Fraassen. Popper asserted that his falsificationism is "the conscious attempt to make our theories, our conjectures, suffer in our stead in the struggle for the survival of the fittest", and "not in need of any further rational justification". A similar conclusion was made by van Fraassen, who asserted that "science is a biological phenomenon, an activity by one kind of organism which facilitates its interaction with the environment". He made a further assertion that the scientific mind is Darwinist mind: only successful scientific theories survive, and no other explanation is required.

I think the theory of evolution cannot be employed as a fundamental explanation. It tells us why many inadequate creatures have been eliminated, but does not provide any guarantee for the survivors' future. If what we have obtained from the past environment has nothing to do with the future environment, if nature turns its behavior sharply and frequently, what is the use of evolution? One must admit that nature has some stability and regularity continuing for some period of time at least. Then what is the demarcation between present knowledge and future truth? Specifically, what is the distinction between evolution from present knowledge and development toward future truth?

Moreover, the theory of evolution only tells us that some theory will be successful (at present, actually), but does not tell us why a particular theory is successful. Besides external factors, let us admit that there should be specific scientific features rendering it successful and seek a resolution from the theory itself.

To answer these questions introduces a more profound aspect of the theory of evolution—it presupposes patterns of nature. The evolution of biological populations depends on the external environment, and those who successfully capture patterns of nature survive. For example, the theory of evolution explains the long neck of a giraffe by food shortage in the lower reaches of trees. If we regard the former—long neck—as an internal pattern, then we should regard the latter—food shortage in the lower reaches of trees—as an external pattern. To put it another way, we may say evolution is the natural selection of all kinds of inhabitants by keeping the ones whose internal patterns match with natural patterns and eliminating the rest.

Once again, we meet with patterns from evolution. Next I will argue that, as the world is comprised of observable phenomena with underlying patterns, the scientific enterprise should aspire to the latter rather than be limited to the former.

4 Observable Phenomena/Underlying Patterns

Undue reliance on the theory of evolution stemming from biology may result in an unwarranted judgement that patterns of nature are volatile and temporary. Nature has patterns of varying degrees, and correspondingly we have theories of different depths. If one can only see what is apparent to the eye and only believe in observations on the surface, he or she will probably come to instrumentalism. van Fraassen held that the value of science is to "save the phenomena"—the aim and only aim of science is to describe observable phenomena. As for the scientific terms concerning unobservables, they are merely instruments for predicting or integrating observables, and we should hold an agnostic attitude, not believe or disbelieve them.

In the first place, the line between what is observable and what is unobservable is quite vague. It has been recognized from primitive times that a look through eyes is an observation. Nowadays it is widely accepted that a look through a telescope is an observation. But some people deny the observation of particles in a cloud chamber. To me it is simply because they have not got used to our extended sense brought by technology yet. Think about computers: raw data are definite observations, statistical data such as the mean and variance are reliable observations, so how could the parameters after complicated calculations no longer be observations? Since computers only have the ability of computation, it is improbable that someone would argue that the number of steps of computation matters. Fortunately we human beings have many abilities, and observation is only one form of perception. It is the most direct form, and unfortunately most people stay at this level. Why do we discriminate in this way, trusting observations from our eyes and doubting theories from our thoughts? All these perceptions are beneficial and important for our recognition of natural patterns!

The exact form of a pattern is not that essential or crucial. Recall the handwritten digits recognition. From our point of view, patterns of digits are visually perceived. In primary school, teachers just write down digits $0 \cdots 9$ on the blackboard and ask pupils to copy these patterns. Nobody will question what the exact form of a digit is or which font is the standard. Students immediately learn digits after such a demonstration. As Schlick considered that a color is defined by showing it, which is an irreducible sense activity, recognition of digits is carried out in the same way. Yet this is not the way computers learn. We can ask computers to output digits, but always in a fixed form regardless of how many fonts they know. Computers cannot write digits freely as humans. In turn, it becomes extremely difficult for them to recognize handwritten digits. Computers have to train intricate models such as SVMs and NNs based on a large dataset, whereas humans get the point after taking a look. And the pattern formed in an SVM or a neural network can hardly be visual anymore. What do these parameters stand for? What connection do they have with the shape of digits? No human knows, but a computer does. The parameters describing the hyperplane and margin in an SVM, the weight parameters in a neural network, and the visual image in a human's eyes share the pattern of digits despite they differ superficially. They reflect different aspects of that pattern which links them altogether.

The above discussion may encourge us to try to review our sciences from God's perspective. Nature has patterns which are clear in God's eyes, but limited to our poor capability, we cannot perceive them easily in "the right form". We invent scientific theories and conceive physical entities such as particle, force, string and so on. Perhaps God is confused by us, just like we are confused by computers. Nonetheless we learn some characteristics of patterns in our tortuous way. We lose the original appearance of patterns and get various incompatible paradigms such that even God cannot tell which one is more accurate. But the patterns we recognize could reflect different aspects of God's patterns, and they are linked altogether.

I absolutely do not want to make science a religion by envisaging God's perspective. Instead, I think this envision may help us to discern scientific realism metaphysically and semantically. We have good reasons to believe the mind-independent existence of the world, although the world of our experience, i.e. the world investigated by sciences is mind-dependent. Reality and objectivity come from the connection between external patterns of the world itself and internal patterns of the world we perceive (including but not limited to observation), where the latter captures certain aspects of the former. I agree that scientific practice should be committed to a literal interpretation of theories, for only literally construed, a paradigm could guide scientists' researches clearly and effectively, which ensures the orderly and rapid development of normal science. But this does not mean that the entities postulated by a theory—particle, wave, energy, etc.—exist. Metaphysically, I recommend the view that theoretical entities, which are internal patterns in the mind, are reflections of external patterns of nature. Again, there is no need to draw a demarcation between "observational terms" and "theoretical terms". They are all perceptual terms as different levels of reflections obtained by different forms of perception.

I must concede that not every phenomenon, or every group of phenomena, has underlying patterns. Spam and handwritten digits are clear positive cases, whereas the result of the World Cup may be a negative case, or at least a hard case which requires a lot of further efforts. I believe science should be the enterprise of pattern recognition. Any theory which claims itself as a science without showing us the patterns it has constructed should be regarded as a pseudoscience, no matter how many instances from which it alleges to make an induction, or how many tests it somehow passes, as in the case of the internet hoax. Of course, there are other enterprises whose aim is not to find out the pattern but to enlighten us in other ways, such as literature, arts and religion, and they deserve our same respect.

5 Incommensurability

Now we come back to the haunting incommensurability. Kuhn maintained that different paradigms are incommensurable. Different computational models are incommensurable in the same way. Viewing data as vectors in a high-dimensional space, SVMs capture patterns by dividing the space into homogeneous subspaces, while NNs capture patterns by organizing a network. They have been competing against each other for over half a century. The neural network theory had been overwhelmingly popular since the 1950s. Experimental psychologists firmly believed in its practical validity, but a group of statisticians headed by Vladimir Vapnik severely criticized its theoretical foundation. Vapnik and Alexey Chervonenkis developed the Vapnik-Chervonenkis theory during 1960-1990, which is a computational learning theory from a statistical point of view. Based on the Vapnik-Chervonenkis theory, Vapnik invented the SVM method in 1995, and it gradually overtook neural networks in machine learning popularity. (Ironically, in order to publish his work in the era of neural networks, Vapnik had to choose the word "network" which he hated to entitle his paper—"Support-vector networks". Later, it was renamed "Support Vector Machine".) In the 2000s, the breakthrough of deep learning architectures revived neural networks, which seized the initiative from SVMs again, but the doubts of their theoretical foundation never go down.

I conceive a scientific paradigm as an adaptive model. Their architectures are incommensurable (it is hard to say which one is more basic, a hyperplane or a network?), but their adaptabilities—the capacity to capture patterns—are commensurable. In computer science, an acknowledged measure is the accuracy on the test set: whether the model classifies the category correctly or fits the value closely on new data. The measure of accuracy is not the final judgement. Holding a paradigm or a model's basic commitments, scientists can develop it flexibly to improve its adaptability and accuracy, e.g. deep learning promotes neural networks' revival. In natural sciences, adaptability becomes much more perplexing and difficult to measure among different paradigms. A pragmatic concern, or technology, should be a feasible measure. Occasionally there are apparent situations where one paradigm's adaptability has it all over another's. The replacement of Newtonian dynamics by relativistic dynamics is the best-known example. Their different architectures—the different meanings of Newton's concepts of space, time, mass and Einstein's concepts of space, time, mass—are not the issue here. The theory of relativity reduces Newton's theory to a special case with a low speed, and its superiority in adaptability is solidly justified.

6 Conclusion

So far, we have discussed scientific methods and scientific realism from a machine's as well as a human's point of view. Pattern recognition is at bottom the same as machine learning: historically, one originated from engineering and one grew out of computer science; actually, they are two faces of the same field. As regards human learning, I would describe science as the enterprise of pattern recognition too. Patterns establish the goal, the reality, and the methods of science. We should activate all our capacities to capture them, including induction, deduction, observation, hypothesis, imagination, or whatever else one can think of. But patterns' volume always exceeds our capacity. For a deaf person, sounds do not exist. What if all human beings lack some kind of sense? But things are not absolute. Deaf people can perceive sounds by the visualization of acoustic waves. To discover the patterns beyond our senses, we must develop sophisticated methods to perceive them. And this is the honorable mission of sciences.

Although I have not fully argued for patterns, I hope to have shown that they are plausible. I wish my work may start a first step towards a bigger goal, to think about real patterns in nature.