5-7-2022

Relating Machine Learning to the Real-World: Analogies to Enhance Learning Comprehension

Vishnu S. Pendyala
San Jose State University, vishnu.pendyala@sjsu.edu

Recommended Citation

Vishnu S. Pendyala. "Relating Machine Learning to the Real-World: Analogies to Enhance Learning Comprehension" *Soft Computing and its Engineering Applications. icSoftComp 2021* (2022): 127-139.
https://doi.org/10.1007/978-3-031-05767-0_11

Relating Machine Learning to the Real-World: Analogies to Enhance Learning Comprehension

Vishnu S. Pendyala[1[0000-0001-6494-7832]]

[1] San Jose State University, San Jose, CA 95192
vishnu.pendyala@sjsu.edu

Abstract. Machine learning is an exciting field for many, but the rigor, math, and its rapid evolution are often found to be formidable, keeping them away from studying and pursuing a career in this area. Similarity has been substantially explored in machine learning algorithms such as in the K-nearest neighbors, Kernel methods, Support Vector Machines, but not so much in human learning, particularly when it comes to teaching machine learning. In the course of teaching the subject to undergraduate, graduate, and general pool of students, the author found that relating the concepts to real-world examples greatly enhances student comprehension and makes the topics much more approachable despite the math and the methods involved. This paper relates some of the concepts, artifacts, and algorithms in machine learning such as overfitting, regularization, and Generative Adversarial Networks to the real world using illustrative examples. Most of the analogies included in the paper were well appreciated by the students in the course of the author's teaching and acknowledged as enhancing comprehension. It is hoped that the material presented in this paper will benefit larger audiences, drawing more learners to the field, resulting in enhanced contributions to the area. The paper concludes by suggesting deep learning for automatically generating similarities and analogies as a future direction.

Keywords: Machine Learning, Nearest Neighbors, Learning by Analogy.

1 Introduction

Machine learning is rapidly evolving as the mortar of modernization. The diverse applications of machine learning are making it ubiquitous in the areas of modernization. Detailed studies such as by the World Economic Forum [1] confirm the increased need for manpower skilled in machine learning and related areas to drive the fourth industrial revolution [1]. The study brings out the need to retrain the existing workforce in addition to training the budding engineers in the new skills. However, the study of machine learning continues to be a daunting prospect for young students and experienced professionals alike. There is a serious need to simplify the discourse on the subject and make it interesting to wider communities. Human beings and machines learn by analogy. Human beings relate new knowledge to what they already know to help in the assimilation of the new knowledge. It is not possible to easily comprehend an abstract concept, completely new from thin air if it does not resemble or relate to any known metaphor. This is similar to how machines learn to classify the test data, unseen till then, by relating it to the training data that is used to build a model. Analogies, therefore, play a critical role in comprehending complex topics.

The fact that machine learning can be applied to fundamental aspects of our civilization and social fabric, such as the truthfulness of information [2-5] clearly shows the impact it can have on transforming our lives and how intertwined it is with the world we live in. It should therefore not be too difficult to relate the machine learning topics to real-world phenomena in ways that enquiring minds can easily understand. As an experiment, the author took upon himself to teach children and adults ages 8 and up, two areas that have a profound impact on life: the ancient wisdom of the Bhagavad Gita and the cutting-edge technology of machine learning, drawing parallels between the two [6-7]. Both of them comprise philosophically deep and intellectually challenging concepts. It turns out that there are quite a few underlying principles that are common to both and apply to the real world as well. The talks on the topics were well understood and appreciated by the participants, including two 8-year-old children. It proves that machine learning can be made interesting to many and not intimidating.

The concepts of similarity, data proximity, and nearest neighbors have found tremendous applications in machine learning. Given that the dot product of vectors inherently has the semantics of similarity and matrix multiplications are essentially a series of dot products, it can be concluded that a significant part of machine learning is inherently, learning by analogy. While classification algorithms such as K-Nearest Neighbors and Non-linear Support Vector Machines using Kernel Methods use the concept of data proximity directly, any algorithm that uses the dot product or matrix multiplication is directly or indirectly leveraging the notion of similarity. In that sense, dot product can be interpreted, using a real-world analogy, as a "doting" product that measures how much a vector is like (or fond of) another vector. Matrix multiplication can be explained as a series of "doting" products.

A doctorate in philosophy is often a required or preferred qualification for many teaching positions for a valid reason. Teachers are expected to analyze and explain the underlying philosophy of the subject they teach. This is best done in small increments or deltas, building upon what is already known. Analogies aid in this delta learning. This paper presents some of the analogies that can simplify the concepts and enhance the comprehension of the topics in machine learning. From a literature survey based on searches using relevant keywords, there is no evidence of enough work done in this direction. The next section summarizes the current relevant literature. The subsequent sections discuss some key concepts and artifacts of machine learning, relating them to real-world phenomena and philosophies. The paper ends with a conclusion and future directions.

2 Literature Survey

The literature survey was carried out by using search queries such as '*teaching "machine learning" "real world*",' '*understanding "machine learning*"' and '*"machine learning" analogy "real world"*'. The search results include textbooks such as [8], which explain how machine learning can be applied to the real world, but not how the machine learning concepts are similar to the real-world notions, indicating that the work in this paper is unique and novel. In an extensive 124-page write-up, Mehta et al [9] draw some parallels between Physics and machine learning to explain various concepts. Using what they call a "physics-inspired pedagogical approach," they point out that similar to Physics, machine learning emphasizes empirical results and intuition. They compare the cost function to "energy," some of the steps in Stochastic Gradient Descent to the momentum-based methods, the pooling step in Convolutional Neural Networks to the decimation step of Renormalization Group (RG).

The human brain and machine learning algorithms both take high-dimensional data as input and perform classification tasks. In [10], the author touches upon multiple areas of intersection of neuroscience and machine learning when giving insights into his laboratory's research program to decipher the algorithms in biological computations that go on in human and animal brains. Analogies accelerate the pace of innovation. Demonstrating this important hypothesis, the authors of [11] use recurrent neural networks to mine idea repositories, specifically, an online crowdsourced product innovation website, Quirky, to generate analogies. The inspiration from these analogies caused the participants in their experiments to generate better ideas. Drawing an analogy to human learning in a teacher-student setup, authors of [12] propose a machine learning framework for various DNN models that uses far lesser training data and executes faster, in fewer iterations, but still achieves similar performance. Acknowledging the difficulty in teaching machine learning to design students, the authors of [13] describe their attempts at using the Lego Mindstorm NXT platform that the students are more at home with, to explain the concepts.

Machine learning models, to a substantial extent, are opaque, lacking explainability. Addressing this issue, the author of [14] describes how this is a problem for socially significant applications of machine learning. The opacity not only makes it difficult to interpret the results but makes it harder for the students to get deeper insights. Relating the algorithms to real-world experiences, as we describe in this paper alleviates the comprehension difficulties. There is hardly any published research that delves into the problem of teaching machine learning effectively to any population [15]. Detailing the learning objectives and strategies for helping students achieve those objectives, the author of [15] presents insights into educating creative practitioners such as musicians on machine learning topics. The functioning of the human brain has inspired the development of neural networks and deep learning frameworks. The authors of [16] argue that relating infant and toddler psychology to algorithms used in computer vision such as Convolutional Neural Networks may result in newer principles of learning.

Teaching machine learning in K-12 schools is increasingly gaining attention for a good reason. Software skills were introduced in K-12 schools a few years ago to prepare the students for the social environment that then was increasingly becoming computer savvy. A similar need is felt today to prepare the young students for the revolution that machine learning and Artificial Intelligence are bringing in. It is therefore imperative that we invent better ways to teach machine learning to all ages. In their detailed work mapping visual tools to ten years of education in K-12 schools, von Wangenheim et al [17] present an extensive survey of the tools that can be used to teach machine learning to K-12 school students. Google's Teaching Machine (TM2) is one of them, which is already popular with students. Authors of [18] compare the deep learning that happens in human learning to the deep learning that is part of machine learning to explain the corresponding similarities and differences between the two.

2.1 Contribution

This paper presents a number of real-world analogies to simplify and explain machine learning concepts and paradigms. From a literature survey and to the best of the author's knowledge, this work is the first of its kind. A future direction the author plans for this work is to explore similarity further and using deep learning, automatically generate analogies for difficult topics in any area of study from a corpus of real-world information such as Wikipedia.

3 Machine Learning Parallels to Real World

Learning, whether in humans or machines follows similar principles. A newborn learns to recognize his father and mother using features and labels. When the newborn is pointed to a person labeled "dad," the baby collects the features of this person through her sensory organs. The features may be that the person labeled "dad" is taller, has facial hair, wears shirts, with a deeper, low-pitch voice, has short hair on head, and so on, whereas the person labeled "mom" may be a bit shorter, has a high-pitch voice, does not have any facial hair, but has long hair on head and so on. As the baby sees more persons, her brain realizes that these features are not constant. They vary. The variability of features, such as the length of the hair or the pitch of the voice is not predictable. A new aunt walks in with a randomly varying voice pitch. The feature varies randomly. In mathematical terms, the baby starts to model each feature as a random variable as she examines more and more people. The label also varies and is the target of the learning exercise, so is called the target variable.

3.1 Logistic Regression

In the above analogy, people are rows of data, their features are columns in the table, and the feature is modeled as a random variable. As more people walk in, the baby realizes that not all features are equal. One of the uncles had long hair too, implying the weight for the feature, 'length of the hair' may not be as much, say as the weight for the feature, 'has facial hair'. Once the analogy is laid out, math automatically follows. A feature, modeled as a random variable is associated with a Probability Density Function (PDF). The 'label' depends on the features, each of which is associated with a weight. The simplest dependency is when the label varies linearly with each feature. The label can therefore be expressed as a weighted sum of features, as we studied in high school. It must be noted that a weighted sum is a dot product of the transpose of the vector of weights and the vector of features. Thus similarity, which the dot product is often a measure for, is fundamental to machine learning. If the label can take only two classes such as 'man' and 'woman,' the label, also called the target variable, can be modeled as a probability of a person being a man or a woman. The weighted sum, therefore, needs to be converted into a value between 0 and 1 to represent a probability and an established way to do it is by passing it through a logistic function. The machine learning model in this case is appropriately called Logistic Regression. The label in this case is a logistic function of the weighted sum of the features.

3.2 Loss Function

The above analogies map the material world in the problem domain to mathematical artifacts. Therefore, all the features involved in the domain need to be converted to numbers so that we can do math with them. When the conversion is done, we need an apples-to-apples comparison. For instance, if our problem is to classify a person as a man or a woman based on height and weight, a difference of 5 inches in height and difference of 30 lb in weight are numerically very different. If Euclidean distance is considered as a measure of similarity, both 5 and 30 are squared, giving both the differences the same importance, and the square root of their sum is used to compare the respective persons. This is erroneous because the weight difference is unduly influencing similarity. The numbers therefore need to be normalized.

A reputed engineering school, which accepts students from various states of the country based on their scores given in their respective states normalizes the scores. This is because some states may be liberal whereas some are more specific. The numbers representing the various features of the data similarly need to be scaled appropriately using normalization techniques. Once the features are all normalized numbers, the next problem is to compute the weights. From the discussion so far, it is clear that we cannot always predict the label accurately, given the nature of random variables and probability. Therefore, a machine learning algorithm such as Logistic Regression comes with a cost. In real world, from a business perspective, the loss is computed as the difference between the sale price and the cost price. In the case of machine learning algorithms, a simple way to think of loss is to view it as the difference between the actual label and the label predicted by the algorithm. It is more a measure of how erroneous the model functioned.

Loss can be seen as a feedback to the system. A business views loss as an important lesson to learn from and takes measures to avoid or minimize the loss. In the case of machine learning algorithms, loss avoidance is not possible because 100% accuracy is never possible. Loss can only be minimized. Loss in machine learning is also a feedback or a lesson back to the system. Just like a mild rebuke may not always work on a student, a simple loss does not have enough effect. The loss, therefore, needs to be a bit more involved than a simple difference between the actual label and the predicted label. A slightly more sophisticated loss is when the difference is squared. There are more sophisticated loss "functions," that can be used to make the algorithm learn faster, but the fundamental factor in all these loss functions is the basic difference between the actual label and the predicted label. It must be noted that the loss function cannot be made complex beyond a level, just like excessively harsh feedback beyond a certain point can demoralize students and have a negative effect. Well-balanced feedback, on the other hand, can motivate the student to learn better. Same applies to the loss functions in machine learning. We need to use a loss function of appropriate complexity and semantics.

The goal of a machine learning algorithm is to minimize the loss, similar to that of a business. We know that the predicted label is a function of the weighted sum of the features. By the time we compute the value of the loss function, we know the values of the features and the actual labels. The only unknown in the loss function is therefore the weights. Loss function becomes an equation with weights as the unknown variables. Finding the weights can therefore be treated as a multivariate maxima-minima problem in partial differential calculus, with the goal being to minimize the loss. This can be accomplished by equating the partial derivatives of the loss function with respect to the weights for each feature, to 0, just like in any other minima problem in calculus.

3.3 Universal Approximation Theorem

We often encounter self-starters and serial entrepreneurs, who have a goal in mind and achieve that goal or at least come close to it over some time against all odds. Such go-getters rise quickly in career and are widely accepted. Once they master the art of succeeding, it is just a matter of applying the learned model to different scenarios. Determined students prepare for a certification or competitive exam over a period and ace the exams. In such cases, a goal can be seen as one or more outputs. The student aiming at the goal takes inputs such as textbooks, courses, blogs, and videos, works with them repeatedly over some time, possibly going back and forth reading books, watching videos, and eventually masters the concepts.

Artificial Neural Networks (ANN) can be thought of similarly. They accept inputs, work with them for some time and produce the desired results, approximately. Computer Science is mostly applied math. The artifact that takes inputs and produces outputs is called a function in math. Since ANNs can approximately do that for any combination of inputs and outputs, they are considered universal function approximators. Universal because this applies to any combination of the inputs and outputs. Function because they behaved like a math function that takes inputs and produces outputs. We know that the results are approximate. The self-starters master the art of succeeding. ANNs master the art of learning to produce any desired output from the given inputs over many back and forth iterations over the entire dataset called epochs.

3.4 Bias

Logistic Regression helps determine the label or class of the given data. This is just one way to come up with the label and as described above, is not free of errors or 100% accurate. Sometimes, we tend to judge a person's behavior and 'label' the person without giving enough thought and consideration. Such simplistic assumptions are called biases. For instance, we often find a bias against a particular age group, gender, or ethnicity arising out of simplistic assumptions. The fact is that the behavior of human beings is far beyond the simplistic assumptions made based on age group, gender, or ethnicity. Therefore, our preconceived hypothesis based on simplistic assumptions is prone to error. Similarly, when a machine learning algorithm makes a simplistic assumption, it suffers from bias, which is one type of error. For instance, Logistic Regression makes a simplistic assumption that the label is linearly dependent on the features and classes can be determined by a predefined function. This results in an error attributable to the bias. Table 1 below lists a few common assumptions in machine learning and the real world, which result in bias.

Machine learning Models	Real World
One line or hyperplane fits all data points to use linear regression for prediction	People of a particular gender or ethnicity or religion are all behaviorally the same
Features of the data domain are all independent of each other	Judging a person, place, or thing by the first or the best or the worst impression
Each data point is independent of all other data points or observations	Overconfidence in oneself and the decisions one makes

Table 1 Simplifying assumptions in ML models and the real world which result in bias (error)

3.5 Overfitting, Variance, and Occam's Razor

There is another type of error in machine learning that is attributable to overfitting. This occurs when the machine learning algorithm closely models the data. It is similar to when people give undue importance to their personal beliefs, values, and perceptions without seeing the big picture. In such cases, two people starting their lives together do not get along after a few days if their value systems are different. In machine learning parlance, the pair has overfitted to the limited value system and behavior they are used to. Algorithms do a similar mistake when they attach undue importance to the limited data they are given. This could manifest as large weights or complex mathematical functions to model the data. Occam's razor, which is widely applicable to many phenomena applies to machine learning as well, because of which, we need to consider model complexity when evaluating a machine learning model and prefer simpler models, which is one way to avoid overfitting.

We often find persons who are highly admired or liked in their respective circles in which they grew up but find it difficult to adjust and get acceptance when they join a newer circle such as by way of a marriage. Similarly, machine learning models, which perform well with the training data based on which they are modeled, but do not perform as accurately with new data, called test data, are said to suffer from "variance." Variance is the other type of error that impacts machine learning models. Variance is typically a result of overfitting the machine learning model to the training data. When the variance component of the error is absent or minimum, the model is said to generalize well, similar to the case of a progressive upbringing of a person.

3.6 Regularization

Spirituality suggests that one should not be too attached to anything in this world. That is, the weights we attach to physical happenings and the impact we perceive should not be high. The Sanskrit word, “Hari” that is often used for meditation and spirituality means “reduce”. Spirituality reduces the weights we attach to the physical happenings around us. In machine learning, this functionality is achieved by what is called “regularization.” Regularization reduces weights by adding a smaller function, called the regularizer, to the loss function at the time of finding its minima. When this new loss function with an added regularizer function is minimized using partial derivatives, the resulting weights are smaller compared to the weights obtained without adding the regularizer. This is similar to how a person depressed because of intense worry is suggested to develop a minor avocation to distract herself from deep, depressing thoughts, which are weighing heavy on the mind. Depression is a result of attaching huge weights to one or more happenings or features of life. The weights can be reduced by adding an activity that lightens up the day.

4 Machine Learning Types and Algorithms

Machine learning algorithms are associated with considerable math that can intimidate new learners. Drawing parallels with the real world is important for improving comprehension in this area. Table 2 lists a few analogies that help simplify the intuition behind machine learning concepts, types, and algorithms.

4.1 Supervised and Unsupervised Learning

Learning is said to be supervised, when the data is already labeled, much akin to a child learning to recognize the world using the labels that the parents attach to the entities that they want their children to know about. A child is supervised to learn about the world around her through the labels such as “good” and “bad” for a behavior or entity. Using these labels, the child develops a habit of attaching relative importance or weights to the features of these entities or behaviors she observes. These weights remain in the child’s mind as impressions or models. A person can be modeled and predicted based on the weights or impressions in his mind. When the weights or impressions in a person’s mind are fully known, the person is mostly predictable. The advertisement world on the Internet runs on this premise. Internet companies such as Google and Facebook predict what advertisements a visitor to their websites may like based on the information they can gather about the visitor. This information is a result of the impressions in the web surfer's mind and the weights she attaches to the various entities she encounters surfing the web.

Similarly, a Supervised machine learning model typically comprises the weights it learns from the labeled data, also called the training data. A machine learning algorithm has learned the model from the training data just like a person learns the relative importance or weights from the labeled entities and behaviors he has been trained on since childhood. When a child grows up to college age and starts to stay in a dorm by herself, parents are no longer available to label the entities such as her classmates or their behavior as good or bad. That is when the child's learning becomes unsupervised. A simple unsupervised task for the child may be to form groups of friends. In machine learning, this unsupervised process of forming groups of data is called clustering. Unlabeled data is grouped into clusters based on similarity.

Most of the learning in a human being's life happens unsupervised. It took many years for the researchers to realize that unsupervised learning holds the true promise of human-level machine intelligence. Recent Turing Award winner, Yann LeCun [19] in 2020, called a form of unsupervised learning, termed self-supervised learning as the future of machine learning. A simple analogy with human learning may have revealed this much earlier. This is one simple example to indicate the power of similarity and analogies in learning and discovery.

Machine Learning Concept	Real-World / Simpler Analogy	Similarity
Supervised Learning	Parents labeling what is good and bad for children	The world is already classified
Unsupervised Learning	Students forming groups without any supervision	Lack of labels
Matrix multiplication	Series of "doting" products	Similarity of two vectors/entities
Maximum Margin Classifier	Arbiter	Equidistant from either class
Artificial Neural Network	Self-starters, Achievers	Mastering the art of achieving targets
Hidden Layers in Artificial Neural Networks	Departments in an organization	Division of labor
GAN	Akinator game	Adversarial nature
Lazy Learning	Last-minute exam preparation	No preparation in advance
Boosting	Student improving exam over exam	Focus on past failures
Overfitting	Narrow-minded	Undue importance to limited artifacts
Regularization	Spirituality	Reduce weights attached to features
PCA	Caricature	Capturing the variance

Table 2 A few parallels between machine learning and the Real-World

4.2 Support Vector Machines

An arbiter or a judge of a court of law is required to be equally unattached to the disputing parties. This concept is imbibed in the philosophy of Linear Support Vector

Machines. A decision boundary in binary classification is essentially acting as an arbiter between two classes. The decision boundary in Support Vector Machines (SVM) is the Maximum Margin Hyperplane (MMH), which is equidistant from the data points at the boundaries of both classes.

4.3 Lazy and Eager Learners

Students tending to procrastinate and put off preparations until the last minute before the exam are lazy learners. In machine learning, algorithms such as the k-nearest neighbors (KNN) do not process the data until the test data arrives, qualifying them as lazy learners. On the other hand, most machine learning models such as logistic regression, SVM, and ANNs do most of the work before the test data arrives, much like the proactive students who prepare for the exams well in advance. In the latter scenario, the models in the mind in the case of the student or some serialized fashion in the case of the machine learning model are ready to be used when it is time to test. Also, when the training data is noisy, if using the KNN algorithm, the value of k needs to be high. This is analogous to when rumors and inaccurate information is floating around such as on the Web, one needs to depend on more information sources than when the information is all true and accurate.

4.4 Ensemble Methods

Ensemble methods combine the efforts of multiple models to cover their individual lacunae, much like in the civilized world where multiple people come together to overcome their individual weaknesses and leverage group synergies. It is an established finding in management that diversity improves team dynamics and the overall quality of the output. Much in the same way, the trees in a random forest are designed to be as different as possible. In sequential ensemble methods like boosting, there is increased focus on the past mistakes, much akin to the way students focus on the mistakes they committed in their previous exams to score better in the forthcoming exams.

4.5 Principal Component Analysis

When drawing a caricature of a celebrity, a cartoonist exaggerates the unique features of the celebrity, while downplaying the common aspects. For instance, if the celebrity has a relatively long nose, the cartoonist exaggerates the nose so that the celebrity can be easily identified from the caricature even if it is not a true depiction of the celebrity. It is much easier and faster to draw a caricature than a portrait. It still serves the purpose of identifying the celebrity in the context. What the cartoonist did in her mind is a Principal Component Analysis (PCA) of the celebrity's face. She identified where the celebrity varies the most and predominantly captured those principal components of the celebrity's face. In machine learning, principal components are determined by computing the eigenvectors of the covariance matrix of the attributes.

Many cultures and people have favorite directions. For instance, some pray facing east, while some others pray facing in the direction of their holiest shrine. Similarly, square matrices have favorite directions. This is the direction of their eigenvectors. When a square matrix is multiplied by a vector, the result is usually a vector with a changed direction and magnitude. The square matrix spends its transformational energy in both rotating and scaling the original vector to produce a new vector. However, when the same square matrix is multiplied by the vectors in its favored directions, all of the square matrix's transformational energy is spent in scaling the vector, increasing its magnitude, while doing nothing to rotate it. The resultant vector is stretched in the favorite direction but its direction is the same as the original vector. The first principal component of a given dataset is in the direction of the most variance in the data. That direction is naturally the most favored direction of the covariance matrix of the dataset, which is the direction of its first eigenvector.

4.6 Artificial Neural Networks and Deep Learning

Artificial Neural Networks (ANN) are biologically inspired and offer many analogies from the real world. For instance, each hidden layer in an ANN accomplishes a minor task in the overall solution to a problem, reminiscent of the division of labor in the industry. The hidden layers can be thought of as divisions in an organization, each responsible for a chunk of the overall mission of the organization. Each neuron in an ANN can be compared to an ant. A single neuron, like a single ant, may not be able to achieve much, but a collection of neurons can do wonders much like a colony of ants can kill the strongest serpent, when they work in tandem.

Interspersing the ReLU activation function in ANNs provides the much-needed non-linearity and boosts the power of ANNs significantly, much similar to how short breaks and context switches can help refresh minds and energize thoughts. Quite a few deep learning frameworks can be compared to parts of our brain that have a similar function. For instance, Convolutional Neural Networks (CNN) used for computer vision are comparable to the occipital lobe of the brain and Recurrent Neural Networks (RNN) have a similar function as the frontal lobe of the brain.

More advanced deep learning frameworks such as Generative Adversarial Networks (GAN) can be explained using analogies too. The discriminator acts like a parent teaching a toddler to write the letters of the alphabet. The generator, like the toddler, generates noise or gibberish at first. The discriminator gives feedback for the generator to better itself much like the parent does to the toddler. After many iterations, both the generator and the toddler learn to produce almost accurate output. Another analogy is the game of Akinator, sometimes called Bulls-eye, where one player or the system in the case of Akinator, thinks of a person and the other player keeps guessing who the person in the first player's mind by asking questions that can only be answered in either a yes or a no. After many iterations of questions and yes/no answers, the second player can guess the character correctly.

5 Conclusion and Future Directions

Time and again, software concepts continue to draw inspiration from the real world. Deliberately or unconsciously, many computer science artifacts bear a striking resemblance to the happenings in the real world. Similarity is probably the single most important underlying principle of machine learning. From a linear predictor to advanced deep learning frameworks, all use dot products. Dot product that is ubiquitously used in machine learning is a measure for similarity. It can therefore be concluded that machines predominantly learn by way of similarity. However, this fundamental way of learning remains unexplored to a significant extent in human learning of difficult topics like machine learning itself. This paper attempted to start filling that gap. When the subject is challenging as is the case with machine learning, it helps to draw parallels to the concepts that the students are already familiar with to help explain the underlying philosophy.

Accordingly, a number of real-world analogies for machine learning concepts have been discussed in this paper. All the analogies are based on human intuition and ingenuity. A future direction for this work is to evaluate the feasibility of automatic generation of analogies, not just for machine learning topics, but for any advanced subject with hard-to-understand concepts. The similarity is a fundamental notion in machine learning. Using the right type of topic, language modeling, and NLP techniques, it may be possible to discover similarities automatically between topics using deep learning, particularly given the universal approximation theorem. It is hoped that this first of its kind work will open up exploring the similarity angle of human learning using both automation and human ingenuity.

References

1. World Economic Forum: Data Science in the New Economy. http://www3.weforum.org/docs/WEF_Data_Science_In_the_New_Economy.pdf, last accessed 2021/09/20.
2. Pendyala, V.: Veracity of big data: machine learning and other approaches to verifying truthfulness. 1st edition. Apress, USA (2018).
3. Pendyala VS, Figueira S.: Towards a truthful world wide web from a humanitarian perspective. In: 2015 IEEE Global Humanitarian Technology Conference (GHTC) Oct 8 (pp. 137-143). IEEE.(2015).
4. Pendyala VS: Securing Trust in Online Social Networks. In: Sahay S., Goel N., Patil V., Jadliwala M. (eds) Secure Knowledge Management In Artificial Intelligence Era. SKM 2019. Communications in Computer and Information Science, vol 1186. Springer, Singapore (2020).
5. Pendyala VS: Evolving a Truthful Humanitarian World Wide Web. https://scholarcommons.scu.edu/eng_phd_theses/18 last accessed 2021/09/20.
6. India West: Free Classes on Bhagavad Gita and Other Ancient Scriptures with Technology Interludes. http://www.indiawest.com/calendar/free-classes-on-bhagavad-gita-and-other-ancient-scriptures-with-technology-interludes/event_a549b5b8-bef7-11ea-a4bf-308d99b28daf.html last accessed 2021/09/20.

7. Pendyala VS: [Video Playlist]. Bhagavad Gita 101 with Technology Interludes for ages 7 and up. https://www.youtube.com/playlist?list=PLrbG5zg_L7VKj79UBsBAXbpT1hn4RkG9a, last accessed 2021/09/20.
8. Brink H, Richards JW, Fetherolf M, Cronin B. Real-world machine learning. Shelter Island, Manning, USA (2017).
9. Mehta P, Bukov M, Wang CH, Day AG, Richardson C, Fisher CK, Schwab DJ. A high-bias, low-variance introduction to machine learning for physicists, pp. 810:1-24. Physics reports. (2019)
10. Helmstaedter M. The mutual inspirations of machine learning and neuroscience. Neuron. (2015).
11. Hope T, Chan J, Kittur A, Shahaf D. Accelerating innovation through analogy mining. InProceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 235-243 (2017).
12. Fan Y, Tian F, Qin T, Li XY, Liu TY. Learning to teach. arXiv preprint arXiv:1805.03643. (2018).
13. Van Der Vlist B, Van De Westelaken R, Bartneck C, Hu J, Ahn R, Barakova E, Delbressine F, Feijs L. Teaching machine learning to design students. In International Conference on Technologies for E-Learning and Digital Entertainment, pp. 206-217. Springer, Berlin, Heidelberg. (2008).
14. Burrell J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. Big Data & Society. (2016).
15. Fiebrink R. machine learning education for artists, musicians, and other creative practitioners. ACM Transactions on Computing Education (TOCE). (2019).
16. Smith LB, Slone LK. A developmental approach to machine learning?. Frontiers in psychology. (2017).
17. von Wangenheim CG, Hauck JC, Pacheco FS, Bueno MF. Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. Education and Information Technologies. (2021).
18. Webb ME, Fluck A, Magenheim J, Malyn-Smith J, Waters J, Deschênes M, Zagami J.: Machine learning for human learners: opportunities, issues, tensions and threats. Educational Technology Research and Development. (2020).
19. Hinton G, LeCunn Y, Bengio Y. AAAI’2020 keynotes Turing award winners event. https://www.youtube.com/watch?v=UX8OubxsY8w (Access: August 10, 2021)