

Machine learning algorithms, how they work, and their uses.

Seth Messier

University of Victoria, Department of Software Engineering, August 26, 2022

Abstract

Machine learning is a methodology that enables computers to predict outcomes based on specific input data. It is a broad field that is used to help identify trends in data and allow everything from personalized recommendations for things like movies and ads to self-driving cars. While there are many different machine learning algorithms, four of the most commonly used ones include, linear regression, logistic regression, neural networks, and collaborative filtering. Each of these algorithms has its different strengths, weaknesses, and primary uses. Machine learning algorithms have created a large number of useful tools. The application of these different tools has and can be used across a wide variety of fields ranging from healthcare to climate change. My summer research project involved understanding the underlying algorithms used in machine learning and applying them to small-scale projects to gain a deeper understanding of how machines learn.

Materials & Methods

Over the course of my research, I took a machine learning course taught by Andrew Ng, a professor at Stanford University and one of the most influential minds in the machine learning world [1]. Through this course I learned many of the fundamentals of machine learning, I then applied what I had learned by completing small projects during multiple phases of the course. One of the more notable projects was using a neural network to recognize the handwritten digits 0-9. This project reinforced the content of the course while giving a challenging project to work on. Through this course and these projects, I was able to learn about the fundamentals of machine learning, its most common algorithms, and how to use them.

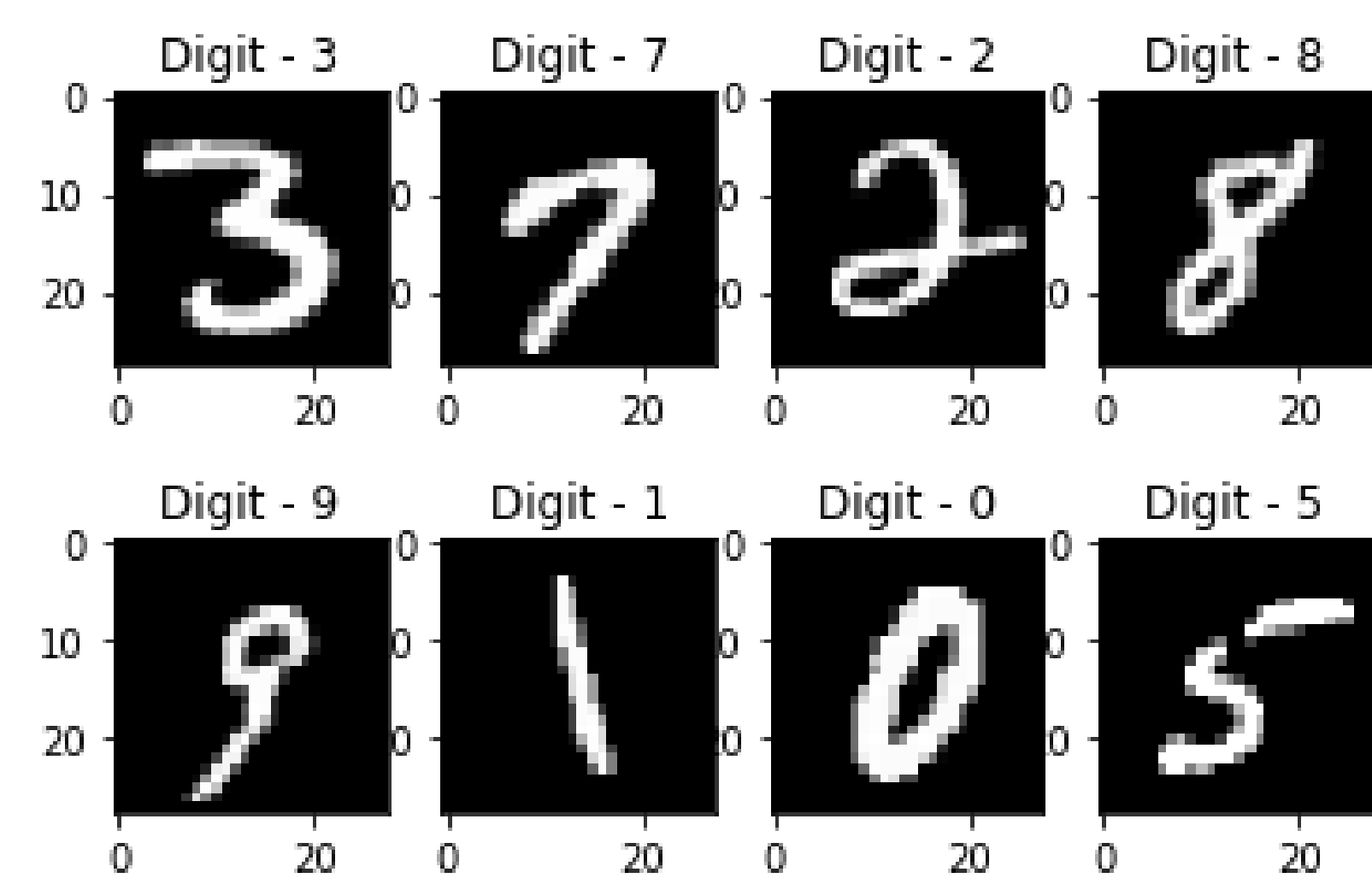


Figure 1: Sample Digits from MNIST dataset
Source: [2]

Results

I studied the implementation and applications of a wide variety of machine learning algorithms. I gained an understanding of the strengths and weaknesses of each model. Using these algorithms it is possible to solve a variety of different problems. In particular, I used the following algorithms: linear regression, logistic regression, neural networks, and collaborative filtering.

Linear regression is an algorithm that develops a model to fit given labeled data such that it can predict the label of a new example it hasn't been trained on. An example of this is estimating the price of a house given its size in square meters. The algorithm would build a model based on a collection of examples with the size of a house and its associated cost. The model would then be able to predict the cost of a unknown house given its size [Figure 2].

Logistic regression is similar to linear regression where the algorithm builds a model based on labeled data but instead of being a range of values the output is binary. An example of this would be estimating a student's chances of passing an exam based on the amount of time they spent studying. The output of the model is a prediction of whether a student would get a passing grade on the test [Figure 3].

Neural networks are machine learning algorithms that mimic the structure of neurons in the human brain which consists of multiple layers of neurons. These networks work by taking input from data in the Input layer of the network and passing the result of its computation through a series of hidden layers and until the output layer [Figure 4]. This algorithm is very versatile and can be modified by changing the structure of the neurons and the number of layers. An example of this algorithm is photo recognition. A common example of photo recognition is classifying the handwriting.

Collaborative filtering is an algorithm that uses what other users rate to match you with items that you might like [Figure 5]. This is the kind of algorithm that is used on many websites and apps such as Netflix and YouTube to recommend content that you may like based on your previous views.

Photos



Figure 2: Price vs Square Footage Linear Regression
Source: [3]

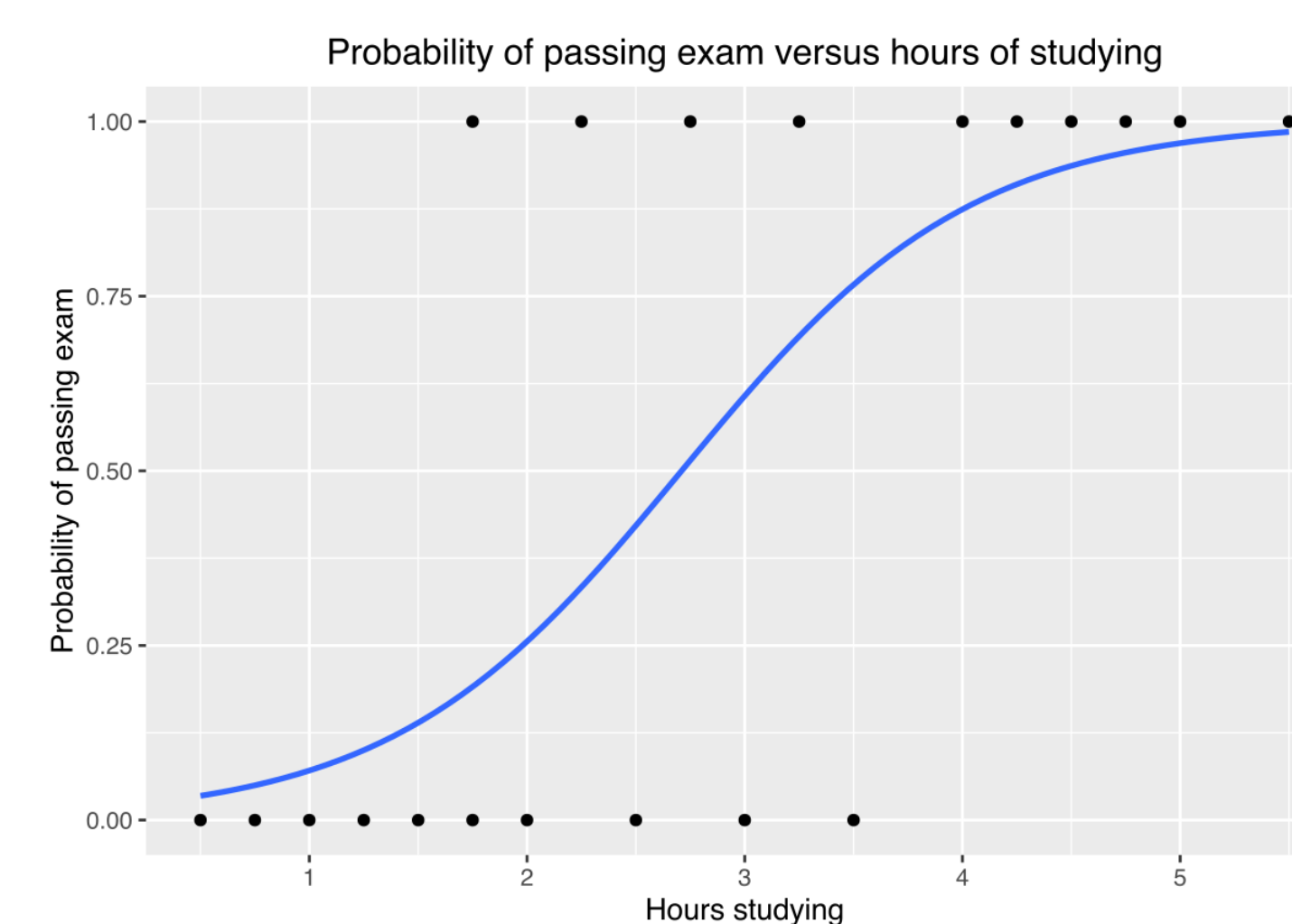


Figure 3: Exam pass logistic curve
Source: [4]

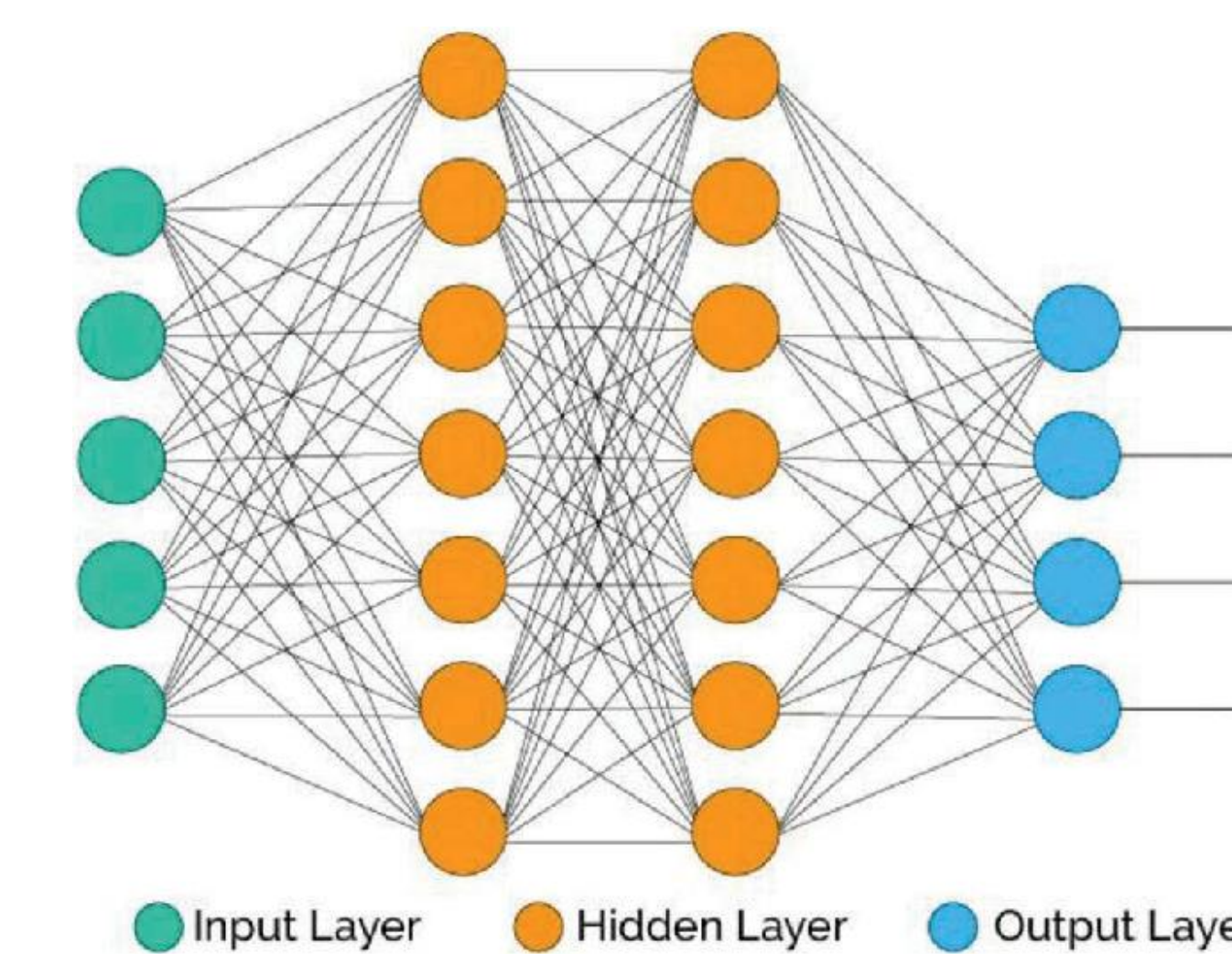


Figure 4: Neural network structure
Source: [5]

COLLABORATIVE FILTERING

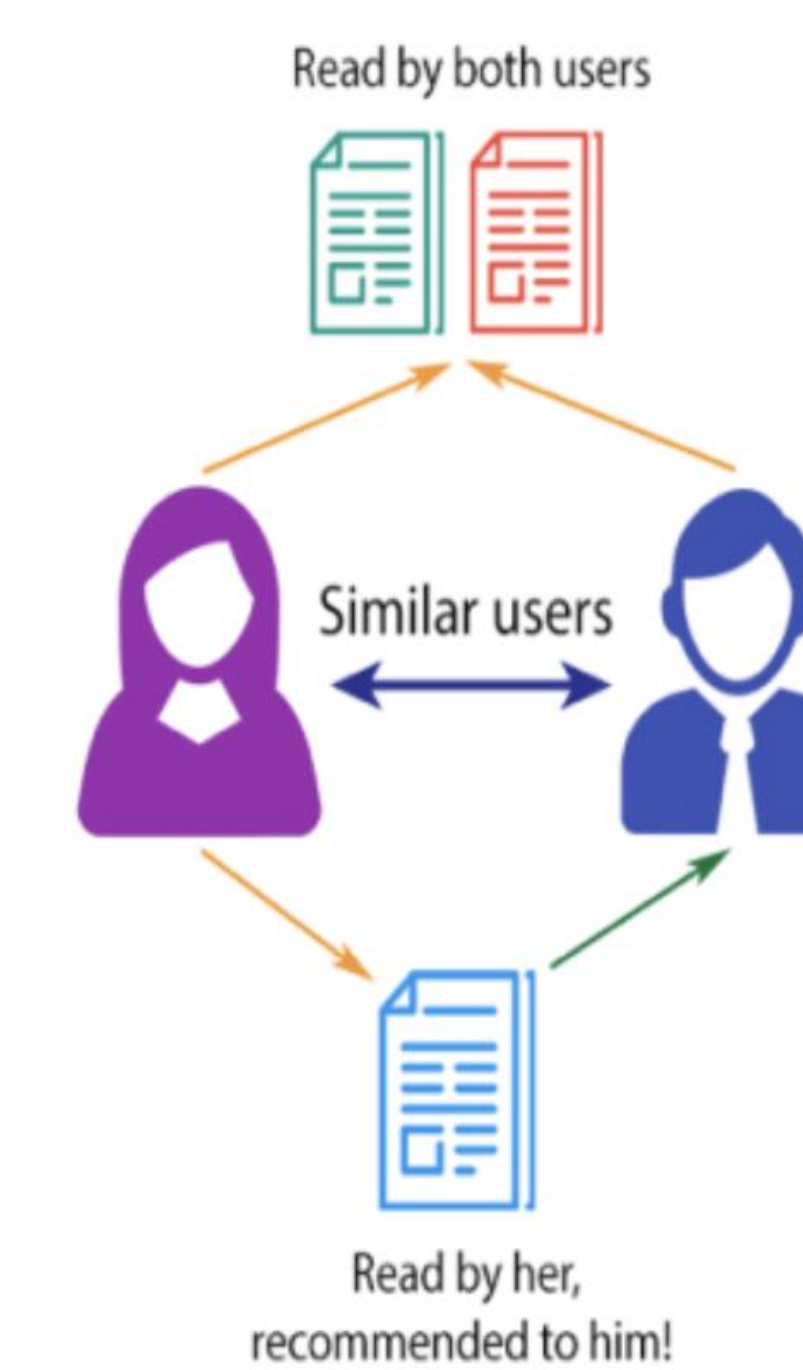


Figure 5: Collaborative filtering
Source: [6]

Discussion/Conclusion

Machine learning is a powerful tool that can be applied to many problems in almost any field. This makes machine learning a popular field to study and apply. Machine learning is at its most useful when it can learn from accurate unbiased data. Although machine learning tools can be extremely useful, a critical component is the quality of the input data. Many datasets contain systematic biases that are often hard to detect resulting in unreliable outputs. As applications such as self-driving cars become common it is important to understand the fundamentals of how these algorithms work along with their limitations. My research gave me a solid understanding of these fundamentals of machine learning.

References

- [1] Ng, Andrew. "Online Courses & Credentials from Top Educators. Join for Free." Coursera, 2011, <https://www.coursera.org/learn/machine-learning-course/home/info>.
- [2] LeCun, Yann. "The Mnist Database." *MNIST Handwritten Digit Database*, Yann LeCun, Corinna Cortes and Chris Burges, Nov. 1998, <http://yann.lecun.com/exdb/mnist/>.
- [3] Svoboda, Ali. *Applied Regression Analysis: Project 2*, 19 Mar. 2015, https://rstudio-pubs-static.s3.amazonaws.com/64900_3f884598a6e84ea3aacef364cbfc54ba.html.
- [4] "Logistic Regression." Wikipedia, Wikimedia Foundation, 22 Aug. 2022, https://en.wikipedia.org/wiki/Logistic_regression.
- [5] Heipke, Christian. "Deep Learning for Geometric and Semantic Tasks in Photogrammetry and Remote Sensing." ResearchGate, Feb. 2020, https://www.researchgate.net/publication/339014914_Deep_learning_for_geometric_and_semantic_tasks_in_photogrammetry_and_remote_sensing.
- [6] Simsek, Gokhan. "Recommendation Systems." *Software Engineering Daily*, 5 Dec. 2018, <https://softwareengineeringdaily.com/2018/10/24/recommendation-systems-by-gokhan-simsek/>.

Acknowledgements

This research was supported by the Valerie Kuehne Undergraduate Research Awards, University of Victoria.

Supervised by Dr. Xiaodai Dong from the Department of Electrical and Computer Engineering.