DEEP LEARNING WITH MATHEMATICS

Duy Khoa Pham and Nadezda Sukhorukova

Swinburne University of Technology School of science, computing, and engineering technologies

Abstract – Deep learning (DL) is key tool of the modern area of machine learning (ML). It plays a pivotal role in the daily lives, with many practical applications including, but not limited to, automatic system, language translation, signal processing and data analysis. The foundation of it is built on mathematics with many theorems. This report with illustrate the Kolmogorov–Arnold Representation Theorem and Universal Approximation Theorem which is of read it paramount importance to the development of deep learning. Furthermore, the report provides a demonstration of Artificial Neural Networks (ANN), which is a typical model of deep learning.

Keyword: Deep learning, Kolmogorov–Arnold Representation Theorem, Universal Approximation Theorem, ANN

1. Introduction

In November 2016, Google Neural Machine Translation (GNMT) is introduced, which is a turning point for the successful implementation of deep ANN in large scale production. Since then, an increasing number of companies have embraced artificial intelligence (AI) in general and DL in particular as powerful tools for resolving complex challenges [1]. Recently, there has been a marked increase in the availability of high computational resources and a wealth of diverse data, leading to a surge in growth and development in the field. The mathematical theorem plays a crucial role in the proliferation of deep learning. This report aims to provide a mathematical explanation of Kolmogorov-Arnold representation theorem, the Universal Approximation theorem, and their connection to neural network. Additionally, a practical demonstration of these theorems will be presented through the training of a neural network for Handwritten Digit Recognition.

2. Kolmogorov-Arnold representation theorem

2.1. Mathematics background

The Kolmogorov – Arnold representation (KA representation in the following) was published in 1957 solving the Hilbert's 13th problem concerning the algebraic equations. It states that every multivariate continuous function can be represented as a composition of continuous univariate functions [2]. Particularly, for any continuous function $f : [0,1]^d \rightarrow$ R, the $(2d + 1)\,(d+1)$ univariate functions $g_q, \psi_{p,q}$ could extract the representation of a d-variate function such that:

$$f(x_1, \cdots, x_d) = \sum_{q=1}^{2d+1} g_q(\sum_{p=1}^{d} \psi_{p,q}(x_p))$$

This theorem is based on the idea of using infinite sum of basis functions to represent any continuous and complex functions. Each basis function is a bump function which is continuously differentiable and compactly supported. The theorem lays a foundation for decomposing a system into separate parts and studying them individually and independently. It has a wide range of applications in mathematical fields and beyond.

2.2. Deep learning architectures with the theorem

There are some arguments that the KA representation theorem is irrelevant to the context of neural network for learning, which is clearly demonstrated by Girosi & Poggio (1989) [3] in the "Representation properties of networks: Kolmogorov's theorem is irrelevant" paper. Nevertheless, the later research illustrates the mathematics background of theorem in deep learning, particularly the context of ReLU networks. It is a type of neural network which is composed of multiple layers of artificial neurons. Each layer is connected to the next layer forming a hierarchical structure. The first layer is the input 1layer connected to the hidden layer in the middle and the last layer is the output layer. The network can be considered as a special case of KA representation theorem that ReLU is used as activation function applied to the output of each neuron. Specifically, Schmidt-Hieber (2021) provides a specific analysis of the theorem including its mathematical foundations, history, and the connection to deep ReLU networks [4]. The paper demonstrates that the construction of ReLU networks imitate the framework provided by KA representation theorem.

3. Universal Approximation theorem and Neural Network

The mathematical goal of neural network is to find a function $y = f(x)$ that maps input attributes (x) to output (y). The function $f(x)$ complexity depends on the dataset's distribution and the network architecture, and it can be arbitrarily complex [5]. The Universal Approximation Theorem states that a feedforward neural network with a single hidden layer can approximate any continuous function for inputs within a specific range to arbitrary precision [6]. The theorem was first formally stated by Hornik, K., Stinchcombe, M., & White, H. (1989) which is of paramount importance to the development of neural networks in solving practical problems [7]. With the activation function applied after a linear layer to provide non-linearities, it is possible for a neural network to appropriate any continuous function to an arbitrary degree of accuracy by adjusting suitable parameters (weights, biases, etc). The theorem has a profound influence in various fields such as natural language processing, robotics, and computer vision. The subsequent section of the report showcases one of the applications of computer vision: image classification.

4. Neural Network Training for Handwritten Digit Recognition

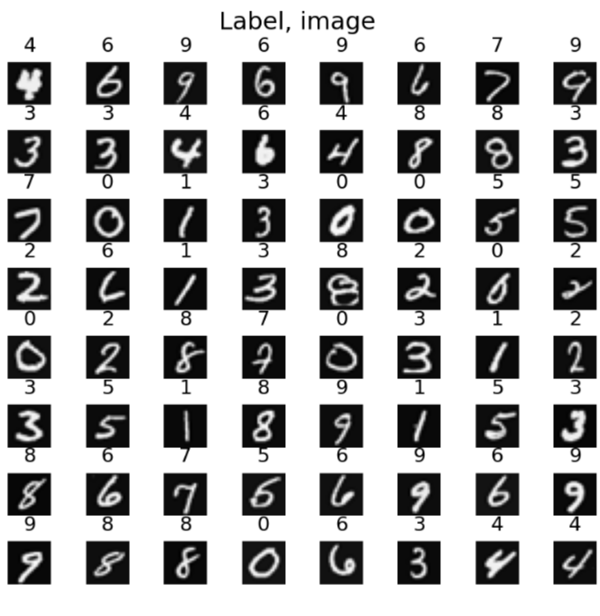

Figure 1. Data visualisation for hand-written digit

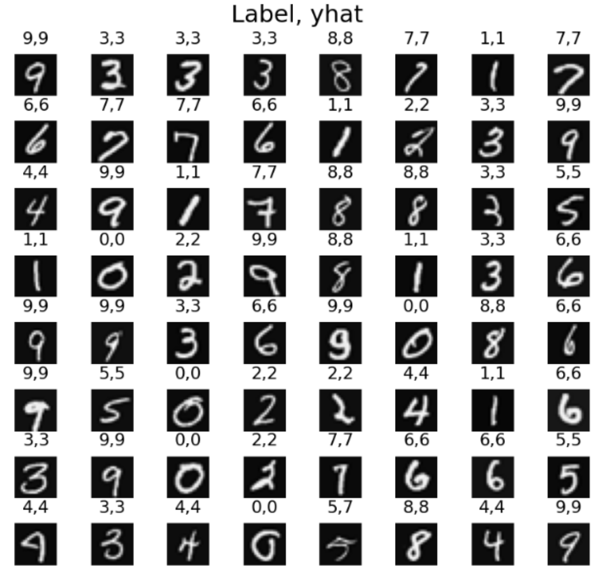

Figure 2. Predictions of the model for each image

For illustration purposes, the author trains ANN model for recognising the hand-written digits 0-9. The activation function employed in this sample is ReLU due to its continuous linear relationship, which is well-suited for classifying a continuous range of values such as 0-9. The raw data for training, depicted as images of handwritten digits, is displayed in Figure 1. The model's objective is to make predictions for each image, recognizing the digit it represents. Upon completion of the training process, the model's predictions are displayed above each image in Figure 2. It can be observed that as the number of epochs increases, the loss function decreases, reflecting the improved accuracy of the model's predictions which is presented in Figure 3.

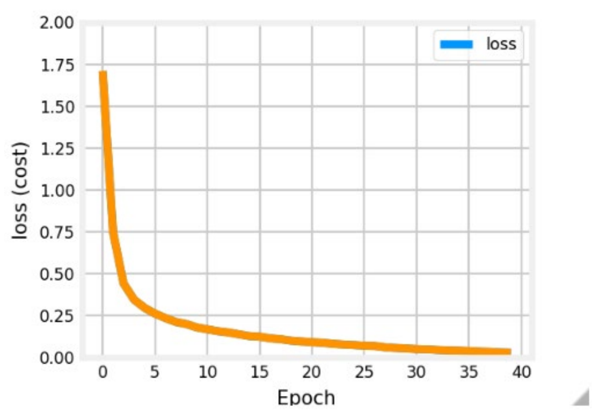

Figure 3. Training Loss vs. Number of Epochs

5. Conclusion

Deep learning is key tool of the modern area of machine learning with many practical applications including automatic system, language translation, signal processing and data analysis. The foundation of it is built on mathematics with many theorems especially the Kolmogorov-Arnold representation theorem and Universal Approximation theorem. Despite computational limitation, Kolmogorov-Arnold representation has an influence in neural network with its framework. With arbitrary precision, the formulation of Universal Approximation theorem has solved the limitations of the prior theorem. With those foundational theorem, neural network training model can deal with many problems including handwritten digit recognition.

6. Acknowledgment

The authors extend their gratitude to DeepLearning.AI and Stanford University for their valuable contributions through the material provided in the Machine Learning Specialization.

7. Reference

[1] Y. Wu. *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.* https://arxiv.org/pdf/1609.08144.

[2] A. N. Kolmogorov. *On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition.* Dokl. Akad. Nauk SSSR, 114:953-956, 1957.

[3] Girosi, F., & Poggio, T. (1989). *Representation properties of networks: Kolmogorov's theorem is irrelevant.* Neural Computation, 1(4), 465–469.

[4] Schmidt-Hieber, J. (2021). *The Kolmogorov–Arnold representation theorem revisited.* Neural Networks, *137*, 119–126. https://doi.org/10.1016/j.neunet.2021.01.020

[5] Sahay, M. (2021). *Neural Networks and the Universal Approximation Theorem*. Medium. https://towardsdatascience.com/neural-networks-and-the-universal-approximation-theorem-8a389a33d30a

[6] Fortuner, B. (2017, March 11). *Can neural networks solve any problem?* Medium. https://towardsdatascience.com/can-neural-networks-really-learn-any-function-65e106617fc6

[7] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

8. Appendix

The code for the demonstration in this paper can be found below:

https://github.com/EspiusEdwards/handwritten_digit_recognition.git.