Motivating and Engaging Students in Active Learning Mathematics and AI

Weiqun Zhang
Wright State University Lake Campus

Artificial Intelligence (AI) is math! This paper demonstrates how mathematics works in the development of AI from modeling, training, learning to decision making. A key focus of the paper is to show how AI not only presents a new approach to teaching mathematical applications in the classroom, but also serve as a learning tool to solve real world problems. Using mathematical tools, an artificial neural network is built and optimized to predict an individual's gender. The evolving AI tools create active learning environments for students to better absorb abstract mathematical concepts and formulas. They have been proven to be an effective and efficient way to motivate students and keep them engaged, even more so than traditional approaches to higher-level education.

0. Introduction

In the time of ChatGPT, both educators and students are talking about Artificial Intelligence (AI) in and out of classrooms. AI makes predictions via a machine learning process using artificial neural networks, which mimic the brain, (Boden 2018). This paper explains the building of a simple neural network from scratch. Terms in neural networks including data mining, machine learning and decision making are introduced with matching mathematical terms. The neural network is designed to predict an individual's gender based on two factors: weight and height. Examples with complete calculation are presented for students to learn more actively and stay more engaged.

While the use of AI within classrooms has been widely debated, it is proven to have positive effects on the learning experiences of students, making it a beneficial and even integral part of education, (Wang et al., 2024). After learning derivates are behind the natural language processing, students show stronger motivation to learn calculus. When using AI, students are enabled to create a model to reduce the pollution of a local lake and they can connect the model to a differential equation. AI helps significantly keep students engaged in learning differential equations. More importantly, both examples offer hands-on experience for students to handle the major challenges of the world.

1. Function and Neuron

An artificial neural network is composed of interconnected neurons in a layered structure. Mathematically, a neuron is a function which produces outputs based on inputs. A neuron with inputs x_1 and x_2 and one output y is shown in Figure 1.

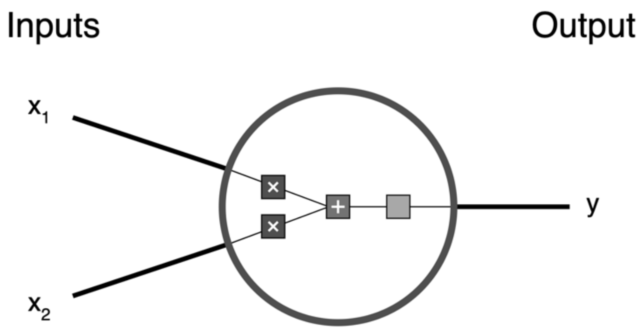

Figure 1. A Neuron

The inputs x_1 and x_2 may inference the output y at different levels. Accordingly, a weight is assigned for each input, w_1 for x_1 and w_2 for x_2. A model is usually not perfect. There might be missing information from the data collected, or there are hidden features of the model itself. Thus, a new variable *b* is introduced to provide extra leverage over the model. The variable *b* is known as a bias in the AI field. Finally, the output y is generated from an activation function with a variable as the sum of weighted inputs x_1, x_2 and a bias b,

$$y = f(w_1x_1 + w_2x_2 + b).$$

Note that the weights and the bias can be learned in the machine training process, (Aggarwal, 2023).

Every neuron has an activation function. A neuron with a linear activation function is essentially a linear regression model. In most studies, a neuron requires a nonlinear activation function. There are different nonlinear activation functions for different purposes. For example, the logistic function is used for binary classification and the soft-max function is used for multiclass problems. The following logistic function,

$$f(x) = \frac{1}{1+e^{-x}},$$

is used as an activation function for every neuron.

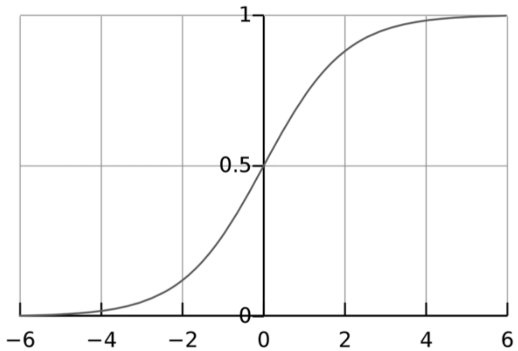

Figure 2. The Logistic/Sigmoid Function

The logistic function is originally derived from the logistic growth model, (Stewart, 2010). Its domain is for all real numbers, and its range is $(0, 1)$. It is feasible to generate one binary output from the logistic function. If the function is close enough to 0, the output is assigned 0; if it is close enough to 1, the output is 1. Due to its S shape in the graph shown in Figure 2, the logistic function is considered a sigmoid function. A sigmoid function is a function with an S shaped graph, such as the logistic function and the hyperbolic tangent function, both of which are commonly used activation functions in artificial neural networks.

The process of passing inputs forward to get an output is known as ***feedforward*** in a neural network. For example, if there is a neuron with two inputs,

$$x_1 = 2 \text{ and } x_2 = 3,$$

with two corresponding weights,

$$w_1 = 1 \text{ and } w_2 = 0,$$

and a bias,

$$b = 4,$$

and a logistic activation function,

$$f(x) = \frac{1}{1+e^{-x}},$$

then the neuron's output is,

$$\begin{aligned} y &= f(w_1x_1 + w_2x_2 + b) \\ &= f(1(2) + 0(3) + 4) \\ &= f(6) \\ &= \frac{1}{1+e^{-6}} \\ &= 0.9975. \end{aligned}$$

As displayed above, ***feedforward*** is a mathematical process of evaluating function values.

2. Functional Composition and Neural Network

An artificial neural network is a bunch of interconnected neurons. In the network shown in Figure 3, there are two inputs, one hidden layer with two hidden neurons, h_1 and h_2,

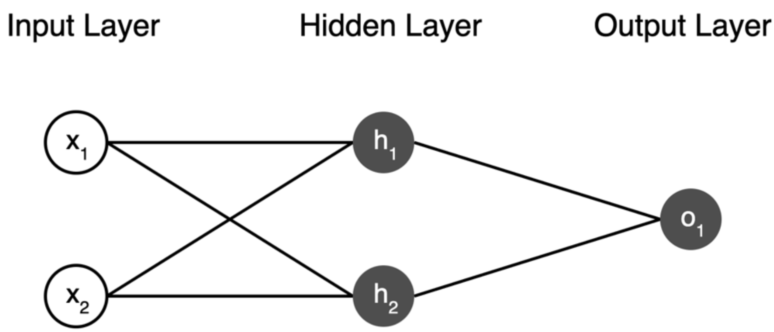

Figure 3. An Artificial Neural Network

and an output layer with one neuron o_1.

The neuron h_1 has the inputs x_1 and x_2 with corresponding weights w_1, w_2 and a bias b_1. The neuron h_2 also has the inputs x_1 and x_2 with corresponding weights w_3, w_4 and a bias b_2. The neuron o_1 has the inputs h_1 and h_2 with corresponding weights w_5, w_6 and a bias b_1. Indeed, the network is a chain of function compositions,

$$\begin{aligned} h_1 &= f_1(w_1x_1 + w_2x_2 + b_1), \\ h_2 &= f_2(w_3x_1 + w_4x_2 + b_2), \end{aligned}$$

and

$$o_1 = f_3(w_5h_1 + w_6h_2 + b_3),$$

where f_1, f_2 and f_3 are activation functions. Notably, the output is not directly from the inputs x_1 and x_2, rather from the hidden neurons h_1 and h_2. Such an arrangement is a typical implementation of an artificial neural network. There can be multiple hidden layers in a network and there can be multiple neurons on each layer. Artificial neural networks are highly efficient because they can solve large scale problems with only two or three layers.

Feedforward in a network is all about function evaluations, certainly a lot more than that of a single neuron. For example, assume there are the following weights,

$$w_1 = w_3 = w_5 = 1,$$
$$w_2 = w_4 = w_6 = 0,$$

and the biases,

$$b_1 = 2,\ b_2 = 1,\text{ and } b_3 = 0.$$

For simplicity, the same logistic activation function is used for all three neurons,

$$f_1 = f_2 = f_3 = f(x) = \frac{1}{1+e^{-x}}.$$

Given the same inputs $x_1 = 2$ and $x_2 = 3$ for a single neuron, one round of ***feedforward*** is done through the network,

$$\begin{aligned} h_1 &= f(w_1x_1 + w_2x_2 + b_1) \\ &= f(1(2) + 0(3) + 2) \\ &= f(4) \\ &= \frac{1}{1+e^{-4}} \\ &= 0.9820, \end{aligned}$$

$$\begin{aligned} h_2 &= f(w_3x_1 + w_4x_2 + b_2) \\ &= f(1(2) + 0(3) + 1) \\ &= f(3) \\ &= \frac{1}{1+e^{-3}} \\ &= 0.9526, \end{aligned}$$

and

$$\begin{aligned} o_1 &= f(w_5 h_1 + w_6 h_2 + b_3) \\ &= f(1(0.9820) + 0(0.9526) + 0) \\ &= f(0.9820) \\ &= 0.2725. \end{aligned}$$

As expected, the output from a neural network is different from a single neuron. AI is built on neural networks. In the examples shown above, weights and biases are randomly chosen. How to choose weights and biases? As a simple answer: no, they are not chosen; however, a machine will learn by itself. Machine learning is the essence of artificial intelligence. It's driven by math!

3. Statistics and Loss Function

Suppose one is presented with a data set on weight, height and gender in Table 1. Can it be discovered how gender corresponds with weight and height? In other words, is there a direct correlation between the two factors?

Name	Weight(lb)	Height(in)	Gender
Adam	152	69	M
Beth	122	63	F
Carol	138	65	F
Dan	148	67	M

Table 1. Weight, Height and Gender

Note in the table, the means are 140 pounds and 66 inches for weight and height respectively. As a standard statistical procedure, the data set is normalized to center around the mean 0 for both weight and height. For gender, male is denoted by 0, female by 1. The normalized data set is shown in Table 2.

Name	Normalized Weight	Normalized Height	Gender 0 male, 1 female
Adam	12	3	0
Beth	-18	-3	1
Carol	-2	-1	1
Dan	8	1	0

Table 2. Normalized Weight, Height and Gender

Now, similar to Figure 3, an artificial neural network can be set up. As a quick review for the network, there is one hidden layer with two hidden neurons, h_1 and h_2, and an output layer with one neuron o_1 for gender. The neuron h_1 has the inputs x_1 and x_2 with corresponding weights w_1, w_2 and a bias b_1. The neuron h_2 also has the inputs x_1 and x_2 with corresponding weights w_3, w_4 and a bias b_2. The neuron o_1 has the inputs h_1 and h_2 with corresponding weights w_5, w_6 and a bias b_1. The network is a set of functions:

$$h_1 = f_1(w_1x_1 + w_2x_2 + b_1),$$
$$h_2 = f_2(w_3x_1 + w_4x_2 + b_2),$$

and

$$y_p = o_1 = f_3(w_5h_1 + w_6h_2 + b_3),$$

where y_p is the predicted gender value, 0 for male or 1 for female. Let y_t denote the true gender value. The input x_1 represents someone's weight, x_2 the height.

The quality of an artificial neural network is calculated with a loss function. The lower the loss, the better the network! There are different loss functions, such as mean squared error, mean absolute error and logarithmic loss. The logarithmic loss is more suitable for binary classification problems. However, for the given example, the mean squared error will be chosen as the loss function. In statistics, the **M**ean **S**quared **E**rror (MSE) is the mean of the sum of squared errors,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_t - y_p\right)^2,$$

where n is the sample size. It is to measure the accuracy of an estimator.

Here is an example on how the loss would be when given a poorly designed network, which will only output the gender value 0 for whatever the inputs are. The following equation is used to calculate this network's MSE,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_t - y_p\right)^2$$
$$= \frac{1}{4}((0-1)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2)$$

$$= 0.5.$$

Such a loss is significant when the outcome is only 0 or 1. In this situation, the network must be trained to provide more accurate results. Calculus makes up these fundamental themes of machine training and learning.

4. Calculus and Machine Learning

In machine learning, the goal is to minimize the loss of accuracy of an artificial neural network. By re-examining the neural network, there are two inputs, each carrying a weight, going through a hidden layer with two neurons, and an output neuron from which the predicted gender value y_p is generated. The network is a chain of functions:

$$h_1 = f_1(w_1 x_1 + w_2 x_2 + b_1),$$
$$h_2 = f_2(w_3 x_1 + w_4 x_2 + b_2),$$

and

$$y_p = o_1 = f_3(w_5 h_1 + w_6 h_2 + b_3).$$

The predicted gender value y_p is a function of multiple variables: $w_1, w_2, w_3, w_4, w_5, w_6, b_1, b_2$ and b_3. So is the loss function *L*,

$$\begin{aligned} L &= MSE \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(y_t - y_p\right)^2 \\ &= L(w_1, w_2, w_3, w_4, w_5, w_6, b_1, b_2, b_3). \end{aligned}$$

The loss function is influenced by weights and biases. One can find the minimum loss by choosing weights and biases wisely. From calculus, the optimized weights and biases can be found with the gradient descent method. To further simplify the discussion, assume that the dataset has only one individual, Carol. Then, the loss function is simplified to

$$L = MSE = \frac{1}{1}\sum_{i=1}^{1}\left(y_t - y_p\right)^2 = \left(1 - y_p\right)^2.$$

The rate of change of the loss function with respect to w_1 is the partial derivative $\frac{\partial L}{\partial w_1}$. By the chain rule, this can be simplified into,

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_p} \cdot \frac{\partial y_p}{\partial w_1}.$$

From the simplified loss function, the result is,

$$\frac{\partial L}{\partial y_p} = \frac{\partial (1-y_p)^2}{\partial y_p} = -2(1-y_p).$$

Meanwhile from the function set of the network, the partial derivatives can be found as,

$$\frac{\partial y_p}{\partial w_1} = \frac{\partial y_p}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$$

$$\frac{\partial y_p}{\partial h_1} = w_5 f'(w_5 h_1 + w_6 h_2 + b_3),$$

and

$$\frac{\partial h_1}{\partial w_1} = x_1 f'(w_1 x_1 + w_2 x_2 + b_1).$$

Keep in mind that f is the logistic function,

$$f(x) = \frac{1}{1+e^{-x}}.$$

Its derivative is from the chain rule again,

$$\begin{aligned} f'(x) &= -1(1+e^{-x})^{-2}(-e^{-x}) \\ &= \frac{e^{-x}}{(1+e^{-x})^2}. \end{aligned}$$

To summarize, the final equation is,

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_p} \cdot \frac{\partial y_p}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}.$$

This process of calculating partial derivatives by working backward is known as ***backpropagation***, or simply ***backprop***. Backprop is all about calculus!

Finally, the network can be trained to learn from the dataset with only one individual, Carol. To get started, all weights are simply initialized as 1, $w_1 = w_2 = w_3 = w_4 = w_5 = w_6 = 1$, and biases, $b_1 = 2$, $b_2 = 1$ and $b_3 = 0$. After one round of ***feedforward***, the consequential numbers are,

$$\begin{aligned} h_1 &= f(w_1 x_1 + w_2 x_2 + b_1) \\ &= f(-2 + (-1) + 2) \\ &= 0.2689, \\ h_2 &= f(w_3 x_1 + w_4 x_2 + b_2) \\ &= f(-2 + (-1) + 1) \\ &= 0.1192, \end{aligned}$$

and

$$\begin{aligned} y_p = o_1 &= f(w_5 h_1 + w_6 h_2 + b_3) \\ &= f(0.2689 + 0.1192 + 0) \\ &= 0.5668. \end{aligned}$$

The network output $y_p = 0.5668$, does not favor either 0 for male or 1 for female. To train the network, first round of ***backprop*** starts with the partial derivatives:

$$\frac{\partial L}{\partial y_p} = -2\left(1 - y_p\right)$$
$$= -2(1 - 0.5568)$$
$$= -0.8864,$$
$$\frac{\partial y_p}{\partial h_1} = w_5 f'(w_5 h_1 + w_6 h_2 + b_3)$$
$$= 1 \cdot f'(0.2689 + 0.1172 + 0)$$
$$= f'(0.3861)$$
$$= 0.2409,$$
$$\frac{\partial y_{h_1}}{\partial w_1} = x_1 f'(w_1 x_1 + w_2 x_2 + b_1)$$
$$= -2f'(-2 + (-1) + 2)$$
$$= -2f'(-1)$$
$$= -0.3932,$$

and

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_p} \cdot \frac{\partial y_p}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$$
$$= (-0.8864)(0.2409)(-0.3932)$$
$$= 0.0837.$$

Because $\frac{\partial L}{\partial w_1} > 0$, the loss function L increases as w_1 increase. To reduce the loss, w_1 must be subsequently reduced. The gradient descent method enables the neural network to learn the optimal value of w_1,

$$w_1^{(1)} = w_1^{(0)} - \eta \frac{\partial L}{\partial w_1}.$$

In the update equation above, η is a positive constant, known as relaxation factor in math, or learning rate in AI. Given a large dataset, an artificial neural network can set its own learning rate that controls the pace of machine learning. Let $w_1^{(0)}$ denote the original value of w_1, and $w_1^{(1)}$ to denote the updated value of w_1. It's important to note that w_1 may increase or decrease in either direction depending on the sign of $\frac{\partial L}{\partial w_1}$. If $\frac{\partial L}{\partial w_1}$ is positive, w_1 decreases, which makes the loss L decreases. If $\frac{\partial L}{\partial w_1}$ is negative, w_1 increases, which makes the loss L decreases again. The same process can be applied iteratively for every weight and bias on one sample and for every sample in the dataset. Ultimately, the neural network improves steadily as the loss decreases.

5. AI and Active Learning

Machine learning is an iterative process optimized with mathematical tools. In turn, AI-driven tools can provide an active learning environment in ways that other tools could not. AI tools offer high quality services continuously and consistently (Gardner, 2021). Their skills have been proven to facilitate learning and improve upon already advanced methods of learning. Within the classroom and from personal experience, students have successfully used AI as an effective study tool to boost their academic performance, increase comprehension of content and gain confidence in using technology to solve complex problems.

While mathematics is not often considered an exciting subject for many students, the rapidly growing use of AI within classrooms connects students with real world technology, and simultaneously increasing their participation in classroom via hands-on learning. For instance, after learning derivatives and their applications in AI, students used their skills to identify the various elements of the mathematical tools for natural language processing. To form a sentence, AI selects the optimal word each round by calculating the minimum of the loss function, which is obtained by setting derivatives equal to zero. From derivatives to sentence generating, students demonstrated a stronger motivation to learn calculus as they realized its power in AI.

AI can be used to solve challenging problems. With help of AI, students built a model to predict the timeline to reduce pollution from a local lake. They were able to identify the various elements of the model, determine the values of specific components, and connect the model to a first order linear differential equation. Additionally, they were able to interpret and draw conclusions from graphs and numerical values. Moreover, students were able to summarize and justify analyses of the mathematical model for the lake pollution problem. And ultimately express those solutions by using an appropriate combination of words, symbols, tables and graphs. AI certainly keeps students more engaged in learning complex subjects, such as differential equations.

For both classes involved, calculus and differential equation classes, with the use of AI, students demonstrated higher levels of motivation and better academic performance and engagement, compared to what was reported during previous classes.

AI can generate personalized learning experiences to meet the individual demands of students. They can produce multiple explanations for math concepts and formulas to students with different backgrounds, (Shemshack et al., 2020). Meanwhile, AI tools present challenges to both educators and students. Depending on data collected and algorithm used, AI may provide inaccurate, or even false answers, (Khosravi et al., 2022). Educators need to be more knowledgeable so they can detect anomalies in AI. While AI excels in solving computational problems, educators need to emphasize more on problem solving skills, (The National Research Council, 2013). As technology advances, it's time for educators to reimagine teaching and assessment. The use of AI within classrooms not only exposes students to the world of technology, but also builds upon their current education in mathematics.

At the time this paper was prepared, two AI scientists were awarded the Nobel Prize in October 2024, (The Nobel Prize Press Release October 2024). Geoffrey Hinton, known as the godfather of AI, won in physics for inventing a method that can autonomously find properties in data. Demis Hassbis, cofounder of Google DeepMind, won in chemistry for his work in developing an AI model for predicting proteins' complex structure. The continuous success of AI will inspire a growing interest in mathematics for more and more students.

References

Aggarwal, Charu C. (2023). Neural Network and Deep Learning, 2nd ed. ISBN: 978-3031296413

Boden, M.A. (2018). Artificial intelligence: A Very Short Introduction, *Oxford University Press*. ISBN: 978-0199602919

Khosravi, H., Shum, S.B., Chen, G, Conati, C., Tsai,Y-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., Gašević, D. (2022). Explainable Artificial Intelligence in Education, *Computers and Education: Artificial Intelligence.* https://doi.org/10.1016/j.caeai.2022.100074

Gardner, J., O'Leary, M. & Yuan, L. (2021). Artificial Intelligence in Educational Assessment: "Breakthrough? Or buncombe and ballyhoo?" *Journal of Computer Assisted Learning*, 37(5), 1207–1216. https://doi.org/10.1111/jcal.12577

Shemshack, A., Spector, J.M. (2020) A Systematic Literature Review of Personalized Learning Terms, *Smart Learning Environments*, 7(33). https://doi.org/10.1186/s40561-020-00140-9

Stewart, J. (2023). Calculus – Concepts & Contexts, 5th ed., *Brooks/Cole Cengage Learning*. ISBN: 978-0357632499

The National Research Council, (2013). The Mathematical Sciences in 2025, *The National Academies Press*.

The Noble Prize Press Release October 2024. https://www.nobelprize.org

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., and Du, Z. (2024). Artificial Intelligence in Education: A Systematic Literature Review, *Expert Systems with Applications*, Vol 2525, Part A. https://doi.org/10.1016/j.eswa.2024.124167