CAS420 Signature Work

Chris Callender

May 2, 2023

1 Abstract

Artificial intelligence (AI) refers to the development of computer systems that can perform tasks similar to a human. Self-driving vehicles are a well-known example of AI. In this paper, we will explore how mathematics is used in AI development. We will also take a more general look at AI development and how it works.

2 Introduction

I think that the future will be greatly impacted by AI. In this course, I learned valuable knowledge about AI, including ChatGPT. AI differs from programming in that programming uses algorithms to produce results from data. That is,

$$\text{data} + \text{algorithms} = \text{results}.$$

Whereas machine learning creates algorithms from data and results. That is,

$$\text{data} + \text{results} = \text{algorithms}.$$

In order to create the algorithms, mathematics must be used, but many people do not realize that math is involved in AI development. Artificial intelligence uses mathematics to emulate human actions. The branches of mathematics that are most used in AI development include linear algebra, statistics and probability, calculus, and mathematical logic. Additionally, neural networks are another interesting component of AI development and will also be explored in this paper.

3 Linear Algebra

Linear algebra is probably the most important area of mathematics in regards to AI development. Linear algebra covers topics such as matrices and vectors, both of which are crucial to AI. Understanding linear algebra, vectors and vector spaces, as well as matrices are crucial for AI developers.

Vectors

A vector is something that can be scaled and added. To be scaled means to be multiplied by a scalar, i.e., constant. Vectors and vector spaces are key parts of AI development. Computers use vectors to manipulate and perform calculations with data. A vector looks like this:

$$A = \begin{bmatrix} 5 \\ 1 \\ -3 \end{bmatrix}$$

Here how vectors are scaled and added.

Scaling: To multiply a vector by a scalar, just multiply each vector component by the scalar.

$$2 \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

Adding: To add vectors, just add the components of each vector. That is, to get the first component of the new vector, add the first components of the given vectors.

$$\begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 12 \\ 28 \end{bmatrix} = \begin{bmatrix} 16 \\ 34 \end{bmatrix}$$

A vector may not seem very useful, but they have many applications. One very important concept is vector spaces.

A *vector space* is similar to a field, but the elements of a vector space are vectors, not numbers.

The vector axioms are:

- Associativity of vector addition
- Commutativity of vector addition
- Identity element of vector addition
- Inverse element of vector addition
- Compatability of scalar multiplication with field multiplication
- Identity element of scalar multiplication
- Distributivity of scalar multiplication w.r.t. vector addition
- Distributivity of scalar multiplication w.r.t. field addition

Some key concepts related to vector spaces include linear combinations, linear independence, linear span, and basis and dimension. Although scaling and adding vectors may seem simple, they form the foundation of these concepts, so it is necessary to know how to perform operations on vectors.

For example, a *linear combination* of some vectors is a new vector that can be achieved by scaling and adding the given vectors. For example, $\begin{bmatrix} 16 \\ 34 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 7 \end{bmatrix}$ since $2\begin{bmatrix} 2 \\ 3 \end{bmatrix} + 4\begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 12 \\ 28 \end{bmatrix} = \begin{bmatrix} 16 \\ 34 \end{bmatrix}$. So we can get $\begin{bmatrix} 16 \\ 34 \end{bmatrix}$ by scaling and adding $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 7 \end{bmatrix}$. That makes $\begin{bmatrix} 16 \\ 34 \end{bmatrix}$ a linear combination of those vectors.

Vector spaces are a crucial part of linear algebra and AI development. AI algorithms operate in high-dimensional vector spaces. We need to use linear algebra to understand these vector spaces and perform operations on vectors.

Matrices

A *matrix* is a grid of values used for storing data. It looks like this:

$$A = \begin{bmatrix} 5 & -4 & -3 \\ 1 & 1 & 0 \\ -3 & 8 & 1 \end{bmatrix}$$

We would say that matrix A is a 3×3 matrix. More generally a matrix is expressed as a $m \times n$ matrix, where m is the number of rows and n is the number of columns. A matrix can also be written using parenthesis, but it is more common to use square brackets.

Linear algebra helps us perform operations on matrices, such as multiplication, inversion, and eigenvalue decomposition. This is helpful for algorithms including Principle Component Analysis (PCA), Singular Value Decomposition (SVD), and Neural Network Training. Let's take a look at a simple example of multiplying two matrices.

If A is a $m \times k$ matrix and B is a $k \times n$ matrix, then AB is a $m \times n$ matrix.

$$\text{Let } A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} -2 & 4 \\ 7 & 1 \end{bmatrix}$$

We want to find the product of matrix A and matrix B, i.e. matrix AB. The entry in the ith row and jth column of AB is the dot product of vectors row i of A and column j of B. So to find the element in row 1 and column 1 of AB we must take the dot product of row 1 of A and column 1 of B. This process is performed for all elements of AB.

$$AB = \begin{bmatrix} 1 * -2 + 0 * 7 & 1 * 4 + 0 * 1 \\ 2 * -2 + 3 * -7 & 2 * 4 + 3 * 1 \end{bmatrix} = \begin{bmatrix} -2 & 4 \\ 17 & 11 \end{bmatrix}$$

Matrix inversion is another important operation. A matrix A is *invertible* if there is a matrix B s.t. $BA = I$, where I is the identity matrix, i.e. a matrix whose components are all 0. Then B is the inverse of A and we write $B = A^{-1}$.

Matrices are also used to solve systems of linear equations. This is a very important application of matrices. For example, let's look at a system of equations.

$$\begin{aligned} x + 3y &= 17 \\ 5x - 2y &= 0 \end{aligned}$$

We can make a matrix where the components are the coefficients and constants of the system, namely $A = \begin{bmatrix} 1 & 3 & 17 \\ 5 & -2 & 0 \end{bmatrix}$

Then, we can row reduce that matrix, which can be done on a calculator, and we get: $RREF(A) = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 5 \end{bmatrix}$

This tells us that $x = 2$ and $y = 5$. You can verify this solution if you would like.

4 Statistics and Probability

Statistics is the study of data and probability is the study of uncertainty. More precisely, statistics is the study of collecting, organizing, analyzing, and interpreting numerical information from data. Data is information that is used to make inferences and decisions. There are two branches of statistics: descriptive statistics and inferential statistics. Descriptive statistics describe observations and include mean, median, mode, and standard deviation. Graphs such as histograms and box plots are also used. Inferential statistics uses sample data to predict outcomes for a population. This is important since sometimes it is difficult, or even impossible, to look at an entire population, so we need to look at a sample. A population is the set of all objects that are of interest and *could* be observed, whereas a sample is a subset of the population that *was* observed. Many statistical tests such as z-test, t-test, and ANOVA are useful for AI.

AI uses statistics to find patterns in data. This makes knowledge of statistics and probability necessary for AI developers.

Let's take a look at some data and descriptive statistics.

$$1 \quad 4 \quad 5 \quad 6 \quad 7 \quad 7 \quad 9$$

Mean: *Mean* is the average value of a sample. It is the sum of all observations divided by the number of observations. For the above data, the mean would be $\frac{1+4+5+6+7+7+9}{7} = 5.57$. This tells us that the average value of this sample data is 5.57.

Median: *Median* is the value where half of the sample is smaller and half is larger than that value. If the number of observations is odd, the median is easy to calculate since it is just the value in the middle position. If the number of observations is even, then you must find the average of the two middle positions. For the above data, the median would be 6.

Mode: *Mode* is the most frequently occuring value. For the above data, the mode would be 7 since it occurs twice, and every other value only occurs once.

5 Calculus

Calculus is the study of change. There are two branches of calculus: differential calculus and integral calculus. Differential calculus uses gradient descent to find the minimum value of a function. This is applied to AI to minimize an error function. We want AI algorithms to be as accurate as possible, so we want the error to be minimized, i.e. as small as possible. Gradient descent is also used to minimize a cost function.

Let's take a look at how to find the minimum value of a function.

$$f(x) = 3x^2 - 4x + 7$$
$$f'(x) = 6x - 4$$

To find the minimum of a function, we need to calculate when the derivative is 0. So, to find where $f(x)$ attains a minimum, we need to find where $f'(x) = 6x - 4 = 0$. Using algebra, we see that $f'(x) = 0$ when $x = \frac{2}{3}$. This tells us that f attains a minimum value when $x = \frac{2}{3}$. This would be very valuable information if f represented an error or cost function since we would want error and cost to be as low as possible.

6 Mathematical Logic

Logical reasoning is the backbone of mathematics, since logic is necessary for proofs. Logical reasoning and attention to detail are skills that all mathematicians must possess. Mathematicians must be able to think under pressure and have great mental endurance. A key component of AI development is producing correct arguments and recognizing incorrect ones. Mathematics is about problem solving, so mathematical skills are necessary for AI developers.

7 Neural Networks

A neural network is a programming technique that replicates the human brain. When neurons try to solve a problem, it strengthens the connections that lead to success and weakens the connections that lead to failure. A perceptron is a single neuron with only one input layer. On the other hand, neural networks are multi-layered perceptrons. There are 3 layers: the input layer, the hidden layer, and the output layer.

Decisions are made by perceptrons in the input layer, then sent to the hidden layer. The hidden layer then weighs results from the input layer. The hidden layer makes more complex decisions than the input layer. Then, the final output is produced.

A deep neural network has multiple hidden layers, so they can perform even more complex operations.

8 Genetic Breeding Model

Here is a useful metaphor, explained in CGP Grey's YouTube video, for understanding how AI learns. AI uses bots. A human can't build bots that understand everything at first. So, the human builds bots that build and teach other bots. The builder bot builds bots that are not very smart at first, so they are sent to the teacher bot. The teacher bot tests the students, then sends the ones that perform well back to the builder. The other student bots are eliminated. The builder then readjusts the connections in the bots' brains. The builder strengthens the successful connections and weakens the connections that lead to failure. Then, the students are sent back to the teacher and retake the test. This process is repeated until there is one student that outperforms the other students and can perform similarly to a human. There are millions of bots and millions of questions on the test. The test is performed as many times as necessary. The grade needed to pass each test gets higher and higher each round.

You may not realize that you actually contribute to AI learning. For example, if you get an online test that tells you to click the squares with cars, that is building a test for self-driving cars, where you are creating the answer key for the test. This helps the bots learn to identify other cars on the road. Humans can not tell a bot what is a car and what is not a car, but they can create more questions and make the test longer to include questions that the bots previously got wrong.

This algorithm is also used by social media and movie companies. For example, YouTube uses an algorithm to maximize viewer watch time. The bots learn what types of videos you like, then recommend similar videos. The bots that get the most watch time advance to the next round and the other bots are eliminated. This eventually leads to one bot that outperforms the others at a high level.

9 Conclusion

In conclusion, mathematics is necessary for AI development since it is needed to create algorithms. I think that AI is very interesting and may consider AI development as a possible career choice.

10 References

1. W3Schools. "Machine Learning." https://www.w3schools.com/ai/default.asp

2. Wikipedia. "Vector Space." https://en.wikipedia.org/wiki/Vector_space

3. Dsouza, Jason. "All The Math You Need to Know in Artificial Intelligence." *FreeCodeCamp*, November 24, 2021. https://www.freecodecamp.org/news/all-the-math-you-need-in-artificial-intelligence/

4. Kung, Inge. "How is Math Used in Artificial Intelligence?" *GoStudent*, March 18, 2022. https://insights.gostudent.org/ca/how-is-math-used-in-artificial-intelligence

5. Parbhakar, Abhishek. "Mathematics for AI: All the essential math topics you need." *TowardsDataScience*, August 9, 2018. https://towardsdatascience.com/mathematics-for-ai-all-the-essential-math-topics-you-need-ed1d9c910baf

6. Jamal, Madiha. "Why is Mathematics Vital to Thrive in Your AI Career?" *TowardsDataScience*, January 15, 2020. https://towardsdatascience.com/why-is-mathematics-vital-to-thrive-in-your-ai-career-c11bd8446ddc

7. Deaett, Louis. "Linear Algebra." Class notes, Linear Algebra, Quinnipiac University, Hamden, CT.

8. Soda, James. "Applied Statistics." Class notes, Applied Statistics, Quinnipiac University, Hamden, CT.