Basic Mathematics behind AI learning

Mert TATLISU, Grade 10, Vienna International School, March 2026

1. Introduction

Artificial intelligence (AI) has become one of the most used technology of the 21st century and increasingly integrated into our daily lives, especially for current generation, who uses it in learning more than books. It is a thriving field with many practical applications and active research topics. While AI is able to defeat even the best human chess player, it only recently has begun matching some of the abilities of average human beings to recognize objects or speech. This article explores the basic math behind how AI systems learn and respond to human. This study focuses on the interrelationships between various branches of mathematics (particularly algebra, linear algebra, calculus, probability and statistics) and their roles in the operational principles of AI.

As with everything in the universe, mathematics is the foundation of modern computers and AI. The first mathematician to spark developments in the field of AI was Alan Turing, who posed the question 'Can a machine think?' and, in 1950, invented the Turing machine. The first mathematician pushed the button on AI learning is Alan Turing, who asked a question "Can a machine think?' and invented the Turing machine in 1950 as the foundation for modern computers and artificial intelligence. Benoit Mandelbrot followed him and developed fractal geometry in 1975, which studies complex and self-similar shapes found in nature and is widely used in computer graphics, and then Judea Pearl developed in 2011 probabilistic reasoning techniques for dealing with uncertainties and decision theory in AI, and finally Yoshua Bengio developed multilayer neural networks (deep learning) in 2019 as a critical part of AI learning. All these allowed a computer speech recognition to perceive the world and perform tasks.

2. Diagnosing a patient

Mathematics is the language of AI and AI uses mathematical tools to learn. AI learning may sound complicated, but it is not a code, it is an applied math. The key mathematical concepts used in AI are linear algebra, calculus, probability and statistics. A robot can be programmed how to walk, or how to learn to walk. While the first one is a simple computer code, the second one is a learning program. The learning program is based on features of an example that have been quantitatively measured.

Whilst a human can recognize all dogs after seeing just one example, AI requires thousands of examples. AI defines the concept of a" dog' from scratch using mathematical numbers and learns by adjusting the numerical weights it has defined based on thousands of examples. During the training process, AI uses thousands of examples from books, articles, websites and conversations, and refines the models. Once training is complete, the AI's 'brain' remains frozen unless a new training cycle begins. The linear algebra being the most important mathematical concept in Al is used to represent and manipulate the data, and it works with vectors, matrices and tensors. It is foundations of how AI represents and processes information. Everything in AI, e.g., images, text or sound, is converted into vectors, matrices or multi-dimensional arrays.

A vector is an array of numbers, and a matrix is a 2-dimentional array of numbers, while a tensor is an array with more than two axes.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

$$X=\begin{bmatrix} X_{1,1} & X_{2,1} \\ X_{2,1} & X_{2,2} \\ X_{3,1} & X_{2,3} \end{bmatrix}$$

Training an AI involves massive matrix multiplications to pass information between layers of AI networks (matrix/tensors), so-called neurons. Mathematical technique like principal component analysis (PCA) is a dimension reduction technique to keep the most important features and use eigenvalues and eigenvectors to simplify massive datasets.

AI does not need a database because it doesn't "look up" information; instead, it uses a mathematical representation of the knowledge it learned during training. During training, it processes massive amounts of data and converts that information into trillions of numerical values called weights or parameters. These weights are stored directly within the model's file. Instead of saving the original data, the AI encodes the relationships and patterns between concepts into these weights.

Linear algebra enables to handle lots of data, while algebra helps creating a formula or to predict, and calculus helps improving the formula or model by minimizing errors.

When you ask a question, the AI passes your input through its layers of weights. Using probability and calculus, it calculates the most likely sequence of words to follow your prompt. It is essentially "reconstructing" an answer on the fly based on the logic it absorbed during training. It responds spontaneously because it is calculating the answer rather than searching for it. The probability is the logic to situations with uncertain information, and the formalization of probability, combined with availability of data, led to the emergence of statistic as a field.

Let's explain how AI works on a simple example helping doctors diagnose a disease "Pneumonia". An AI system turns these medical data into numerical numbers and the dataset trains the AI in diagnosing.

Table 1: Patient medical data used to create a matrix

Patient	**Fever** **(Yes: 1, No: 0)**	**Cough** **(Yes: 1, No: 0)**	**CRP value from blood analysis >5 is 1 (sick) and <5 is 0 (not sick)**	**Chest pain** **(0: no pain, 10: severe pain)**	**Diagnosis**
Patient A	1	1	1 for CRP=100	1 for level=9	Pneumonia
Patient B	0	1	0 for CRP=5	0 for level=3	Healthy
Patient C	1	1	1 for CRP=20	1 for level=7	Pneumonia

Each patient becomes a vector of numbers, e.g., [1,1,1,1], [0,1,1,0], [1,1,1,1], and the weight factors of each symptom is another vector of numbers, then the AI forms a matrix from several patient's data using linear algebra, and creates a formula from the data matrix for three patients and weight vector of each symptom in diagnosing Pneumonia disease:

$$y=X.W+b, \text{ where}$$

X: matrix of three patients' data using their symptoms,

$$X=\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

W=weight factor used for each symptom according to its importance in diagnosing Pneumonia sickness,

W= [0.5, 0.2, 0.6, 0.4], and

b=bias (error).

Matrix multiplication allows the AI to compute prediction score (y) for each patient's status, based on an algebraic formula by:

$$y_1=w_1.x_1+w_2.x_2+w_3.x_3+w_4.x_4+b$$

$$y_1=1x0.5+1x0.2+1x0.6+1x0.4=1.7+\mathbf{b}$$

When the prediction scores of other two patients are calculated using the same formula, $y_2=0.8+\mathbf{b}$ and $y_3=1.7+\mathbf{b}$.

If the prediction score is above a threshold (probability), it predicts pneumonia. The bias (b) in its predictions improves with more patients, which means the more data sets better predictions and learning.

Calculus in AI is mainly used to improve the model automatically. It measures how a function changes as its input changes and determines how to adjust model parameters to reduce prediction bias (b) or errors. It uses optimization algorithms like derivatives, gradient descent to find the best model parameters. The AI changes the weighing factor of symptoms in diagnosing phenomena and checks whether bias is increasing or decreasing. Calculus answers this using the derivatives to determine exactly how much to adjust a model's internal weights to minimize errors (error becomes smaller or bigger when weighing factor increases), which is called gradient descent. After several iterations, the gradient becomes nearly zero, meaning the AI has "found" the best possible answer and stopped changing.

$$\text{Error}=1/n\Sigma^n_{i=1}(y_{actual,i}-y_{predicted,i})^2,$$

where AI uses derivatives from calculus to adjust weights to reduce error.

The probability predicts likelihood of outcome and helps AI express predictions with uncertainty, while graphs visualize relationship and models. Statistics helps AI learn patterns from past data. Statistics provides the tools to verify if a model is actually accurate or just "memorizing" the training data. Statistics helps AI to detect patterns in this data. The average CRP value for sick and healthy patients provides an initial reference value and helps to quantify these patterns. Without statistics, the AI would not able to detect meaningful trends. Instead of a simple "yes" or "no", AI outputs probability distributions, e.g., 85% is pneumonia. A fundamental rule used to update the probability of a hypothesis as new evidence appears.

By optimization, AI compares the results between maximum and minimum iteratively, and changes its parameters in every iteration until it reaches the optimum results, where errors are minimized. The "learning" happens as the algorithm uses the derivative (calculus) to step toward the bottom of the curve. In mathematics, the learning curve is the quantifiable relationship between experience (data/training time) and performance (error reduction). AI learning is typically represented by a Power Law ($b(x)=ax^{(-n)}$). As the amount of data (x) increases, the error (b) decreases, but at a diminishing rate.

3. Language recognition

So far, the above example of diagnosing a patient shows the roles of linear algebra, algebra and calculus however, the role of probability and statistics are not explained clearly with that example. The role of probability can be better explained in language recognition: answering the fundamental question that how AI understands the human language. AI does not understand text the way humans do. Instead, it processes text using mathematical patterns learned during training. A word is a sequence of characters, which are not independent but constrained by the words of the language. This has the advantage that even if a human being can not recognize a character, it can still read t**?**e word. AI learns sequences and model such dependencies, and use many different algorithms to interpret (e.g., standardization and tokenization) and statistical methods to make decisions on the patterns.

Standardization is a process of converting a text to the basic form removing differences, punctuations, upper cases etc., e.g., “Is not it?” and “isn’t it?” are two different words for a machine, so that converting them to basic form bring them closer: “is not it?” versus “is nt it?”.

Tokenization is the process of breaking down a large body of text into smaller units, like words or sentences. By splitting text into tokes, machine can start analyzing each element (token) individually. Let’s start with “How do you understand text?”, it is converted into a vector of tokens [“How”, “do”, “you”, “understand”, ‘?’], and each token has a probability or index factor according to its proximity [0.12, -0.45, 0.88,..].

“by”: 0.35

“using”: 0.28

“through”: 0.15

The model predicts probabilities for the next token and selects a likely token and continues generating the answer. Token indexes are adjusted using gradient descent which uses derivatives from calculus. Tokens are not created according to the human logic, it is derived from likelihood, for example:

The “believ+e” has two tokens “believ” and “e” and each has its own weights, while “believ+ing” or “believ+able” have closer 5-dimentional vectors:

Believe=[0.82,0.44,0.67,0.21,0.59]

Believable=[0.79,0.47,0.63,0.25,0.61]

Believing=[0.84,0.42,0.69,0.20,0.58]

When the AI processes text of “believe”, it multiplies the vector (x) which is token indexes by weight matrix of each token, W=[0.5,0.2,0.1,0.3,0.4], to give a meaning (y) to the word:

y=W.x

y=0.5x(0.82)+0.2x(0.44)+0.1x(0.67)+0.3x(0.21)+0.4x(0.59)=0.864

Repeating same calculations for “believable” and “believing”, they will have 0.873 and 0.858. Since these values are very close, the AI knows that these words are related.

AI learning relies heavily on powerful hardware, as a deep learning algorithm is running on specialized hardware which can do 10^{14} and 10^{17} operations per second, while a standard computer CPU can do 10^{9} or 10^{10} operations per second.

4. Conclusion

In conclusion, mathematics being fundamental of AI learning are essential for understanding how algorithms function and make predictions. AI relies on mathematics: statistics detects patterns, linear algebra represents and processes data, algebra creates predictive formulas, probability quantifies uncertainty, calculus optimizes weights, and graphs provide visual understanding. These mathematical foundations reveal that AI is not magic, it is the rigorous application of mathematical principles.

5. References

1. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press, 2010.
2. Alexander Jung, Machine learning: the basics, Springer, Singapore, 2022.
3. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press book (online), 2016.
4. Stuart Russell, Peter Norvig, Artificial Intelligence: A Modern Approach, 4th Edition, Pearson series, 2020.
5. David C. Lay, et. al., Linear Algebra and its applications, 5th Edition, Pearson series, 2016.
6. Francois Chollet and Matthew Watson, Deep learning with Phyton, 3rd Edition, Manning, 2025.
7. Preethi Senthilkumar, the fundamentals of NLP: How machines understand human language, Payoda technologies, 2025, https://payodatechnologyinc.medium.com/the-fundamentals-of-nlp-how-machines-understand-human-language-549a1f1698fb.