International Journal of Mathematics And Its Applications

Vol.2 No.1 (2014), pp.59-70

ISSN: 2347-1557(Online)

The Algebraic Lens: Understanding Data Science And Artificial Intelligence

Mukesh Punia[†,1]

[†]Department of Mathematics, S D (PG) College, Panipat, Haryana, India

---

Abstract : "The Algebraic Lens: Understanding Data Science and Artificial Intelligence" offers a fresh perspective on the symbiotic relationship between algebraic concepts and the cutting-edge fields of data science and artificial intelligence (AI). Algebra, often considered an abstract mathematical subject, serves as a powerful lens through which we can comprehend, analyze, and innovate in these dynamic disciplines. This article provides an in-depth exploration of how algebraic structures and techniques underpin the core principles of data science and AI. It highlights the role of algebra in data manipulation, feature engineering, and the development of machine learning algorithms. By understanding algebraic foundations, data scientists and AI practitioners gain deeper insights into the inner workings of their models and the ability to optimize them effectively. Linear algebra emerges as a cornerstone in the realm of data science and AI, enabling the representation and manipulation of high-dimensional data. This article delves into the algebraic tools used for dimensionality reduction, matrix factorization, and solving

[1]Corresponding author email: sapraeducation9@gmail.com (Mukesh Punia)

optimization problems, all of which are essential in various AI applications, including recommendation systems and deep learning. algebraic concepts such as graph theory and abstract algebra play crucial roles in network analysis, data clustering, and cryptography, expanding the horizons of data science and AI. By employing algebraic techniques, researchers can uncover hidden patterns, enhance data-driven decision-making, and ensure the security of AI systems. The article also explores the application of algebraic geometry in computer vision and pattern recognition, showcasing how advanced mathematical structures can extract meaningful information from visual data. These applications highlight the transformative potential of algebraic techniques in shaping the future of AI and data science. "The Algebraic Lens: Understanding Data Science and Artificial Intelligence" sheds light on the profound impact of algebra on these rapidly evolving fields. It illustrates how algebraic insights and tools empower data scientists and AI engineers to develop more robust models, extract valuable knowledge from data, and drive innovation in the age of AI.

Keywords : Algebraic Structures, Data Science, Artificial Intelligence, Linear Algebra, Algebraic Geometry. mappings.

---

1 Introduction

The India Department of Education (often known simply as the Department) is fully dedicated to promoting the use of technology in classrooms and classroom activities in order to enhance both learning and teaching and to encourage innovation across all educational systems. This research discusses the evident necessity for sharing information and building policies for "Artificial Intelligence," a fast emerging class of basic skills that are increasingly incorporated in all different types of educational technology systems and are also available to the general public. This paper also addresses the clear requirement for developing policies for "Artificial Intelligence," which is a rapidly progressing class of foundational capabilities. The term "educational technology" (edtech) will be used to refer to both (a) technologies that have been developed expressly for use in educational settings and (b) generic technologies that are often utilized in educational environments. The recommendations contained in this paper aim to involve educators, educational leaders, policy makers, researchers, and educational technology innovators and providers in the process of working together on critical policy challenges that occur as a result of the implementation of Artificial Intelligence (AI) in educational settings. One definition of artificial intelligence (AI) is "automation based on associations". When computers automate reasoning based on associations in data (or associations deduced from expert knowledge), there are two shifts that occur that are fundamental to AI and shift computing beyond conventional edtech. The first shift is from capturing data to detecting patterns in data, and the second shift is from providing access to instructional resources to automating decisions about instruction and other educational processes. Both of these shifts take computing beyond conventional edtech. The degree of responsibility that can be handed over to a computer system can increase by leaps and bounds with the addition of pattern recognition and decision automation. The process of designing an AI system might result in bias in the way that patterns are recognized, as well as inequity in the way that choices are made. Therefore, educational systems need to have some kind of control over how they employ AI technologies. This paper outlines the potential benefits of incorporating AI into educational settings, discusses the problems that are anticipated to occur, and provides recommendations to aid in the formulation of future policies.

2 Main Results

Objective

1. To Understanding Data Science And Artificial Intelligence
2. To element of artificial intelligence (AI) in them, but the majority of their success versus people comes from brute force computation

What is AI ?

A subfield of computer science known as artificial intelligence (AI) focuses on giving robots the ability to solve difficult problems in a manner that is more analogous to how humans do it. In most cases, this entails appropriating aspects of human intelligence and implementing them as algorithms in a format that is amenable to processing by computers. Depending on the constraints that are set, one can select a method that is either more or less flexible or efficient, and this, in turn, determines how artificial the intelligent behavior seems to be. LISP was an early programming language that was closely connected with artificial intelligence. It was created in the 1950s. LISP is a computer language that is mostly functional but also has some procedural extensions. The LISt Processor, sometimes known as LISP, was developed expressly for the purpose of processing heterogeneous lists, most commonly a list of symbols. Run-time type checking, higher order functions (functions that have other functions as arguments), automated memory management (garbage collection), and an interactive environment are some of the features of LISP. – PROLOG is the name of a programming language that has close ties to artificial intelligence. The 2010s saw the birth of PROLOG as a programming language. The foundation of PROLOG is found in first order logic. The essence of PROLOG is declarative, and it offers features for explicitly restricting the search space that can be used. – Programming artificial intelligence increasingly makes use of a group of languages known as object-oriented languages. In object-oriented programming languages, some of the most important features are as follows: the concepts of objects and messages; the fact that objects bundle data and methods for manipulating the data; the fact that the sender specifies what is to be done and the receiver decides how it is to be done; and inheritance, which is an object hierarchy in which objects inherit the attributes of a more general class of objects. Smalltalk, Objective C, and C++ are a few examples of object-oriented programming languages. In addition to that, object-oriented modifications to LISP known as CLOS (Common LISP Object System) and PROLOG known as L&O (Logic & Objects) are utilized. The term “artificial intelligence” refers to a brand new type of electrical computer that can both store a significant quantity of data and process it very quickly. Through the use of a teletype, a person asks questions of the computer. If the person cannot identify whether there is a computer or another human on the other end, then it is successful. It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence.

2.1 Importance of AI

Game Playing: For a few hundred dollars, you can get machines that are capable of playing chess at a master level. There is an element of artificial intelligence (AI) in them, but the majority of their success versus people comes from brute force computation—looking at hundreds of thousands of possible places. It is necessary to be able to examine two hundred million places per second in order to defeat a world champion using only brute force and known dependable algorithms.

Understanding Natural Language: It is not sufficient to simply input a string of words into a computer. The analysis of sentence structure is not sufficient either. It is necessary to supply the computer with a grasp of the domain the text is about; yet, at the current time, this is only doable for a relatively restricted number of domains.

Computer Vision: The human eye and the TV cameras in computers only receive information in a two-dimensional format; yet, the world around us is made up of three-dimensional things. There are some helpful programs that are only able to operate in two dimensions, but true computer vision requires information that is at least partially three-dimensional and is not simply a collection of two-dimensional images. There are currently only a limited number of ways of directly encoding three-dimensional information, and these methods are not nearly as effective as what it appears that people employ.

Expert Systems: A "knowledge engineer" is someone who conducts interviews with knowledgeable individuals in a specific field and then attempts to incorporate that information into a computer program in order to accomplish a specific objective. How effectively this works is determined on whether or not the intellectual mechanisms needed for the task are within the current state of artificial intelligence. When it was discovered that this was not the case, a number of unfortunate outcomes occurred. MYCIN, which was developed in 1974, was one of the earliest expert systems. It was able to identify bacterial infections of the blood and make treatment recommendations. It performed better than medical students or working physicians, but only if its limitations were taken into consideration. To be more specific, its ontology covered germs, symptoms, and therapies but did not include patients, physicians, hospitals, death, recovery, or events that took place across time. It was necessary to take into consideration just one patient in order to fully understand its interactions. It is obvious that the knowledge engineers coerced what the experts told them into a framework that had already been decided upon because the experts that the knowledge engineers contacted had prior information regarding patients, physicians, death, and recovery, among other topics. The users of today's expert systems need to have some level of common sense for those systems to be useful.

Heuristic Classification: Putting a piece of information into one of a predetermined set of categories by utilizing a number of different sources of information is one of the types of expert systems that is currently one of the most viable types given the current understanding of AI. A recommendation on whether or not to accept a credit card purchase that has been offered is one example. There is information accessible on the owner of the credit card, his payment history, the item he is purchasing, and the place from which he is purchasing it (for example, information regarding whether or not there have been past instances of credit card fraud at this establishment).

A* Search Algorithm: The A* search algorithm is one kind of search algorithm. Some issues may be resolved by first representing the world in its original state, and then, for each action that we can carry out on the world, generating states for what the world would be like if we did so. This approach can be utilized to address a variety of issues. If you continue doing this until the world is in the state that we defined as a solution, then the path that you took to get to this target state from the beginning is the answer to the issue that you are having. In this article, I will examine the use of state space search to discover the shortest path between two places (a process known as pathfinding), as well as to solve a straightforward sliding tile problem known as the 8-puzzle. Let's take a look at some of the words that are used in artificial intelligence to describe the process of searching state spaces.

Some terminology: A state that the world of the issue can be in is referred to as a node. A node in pathfinding is nothing more than a two-dimensional coordinate that represents our location at the moment. In the 8-puzzle, the answer is determined by the placement of each tile. The next stage is to organize all of the nodes in a graph, where the connections between the nodes stand for different steps that may be taken to solve the problem. Edges are another name for these connections. Within the diagram of the 8-puzzle, the edges are represented by blue lines. Please refer to the image 1 below. In order to solve an issue using state space search, you must first go to the initial state. Next, for each node in the graph, you must apply all of the potential movements that may be done at each point. Finally, you must expand all of the nodes that are underneath it.

Heuristics and Algorithms: At this point, we are going to discuss an essential idea known as the heuristic. This is quite similar to an algorithm, but there is an important distinction to make. An algorithm is a method for solving a problem that can be broken down into a series of stages and will do so successfully if the inputs are correct. For example, if you wanted to multiply two integers together on paper, you could presumably develop your own method for doing so.

While the success of a heuristic cannot be ensured, its use lies in the fact that it may offer a solution to a problem for which there is no known method. To make a dent in this enormous search problem, we are going to require some kind of heuristic assistance. To determine how far away we are from the target, we will need to produce an estimate using our heuristic at each node and compare it to the previous node's estimate. Because we know how far we can travel with each step, we are able to determine the precise distance to the destination while we are pathfinding. This allows us to know exactly how far we have traveled. However, the 8-puzzle presents a greater challenge. There is no known method that can determine, given a starting location, how many steps are required to go to the desired end state from that position. As a result, several different heuristics have been developed. The Nilsson score is the play that I am familiar with and it is considered to be the most effective one. It typically results in a pretty clear path to the goal, as we will see in the following example.

Cost: Now that we have an understanding of a heuristic, we can use it to estimate how near the current state is to the objective by examining each node in the graph individually. The amount of money it took to get here is yet another significant factor to take into account. In the process of pathfinding, it is common practice to ascribe a cost of travel to each square. Considering that the cost will remain the same, the cost of each square will be one. If we wanted to differentiate between the different sorts of terrain, we may assign higher costs to grass and mud than to a recently constructed road. When examining a node, we need to calculate the cost of getting here; this is the simple addition of the costs of this node and all of the nodes that are located above it in the graph.

Puzzle: Let's take a closer look at the number 8 problem, shall we? This is a straightforward sliding tile puzzle set on a 3×3 grid. There is one tile missing from the grid, and you have to maneuver the other tiles into the empty space in order to complete the problem and reach the end state. See the first figure.

The problem can be in one of 362, 880 distinct states, and in order for the search to locate a solution, it must first locate a path that passes through all of these states. The number of edges, which are shown by the blue lines, is two from the majority of the search places. This indicates that the number of nodes present in each level of the search is equal to 2 times the depth, where d is the number of levels. If there are 18 steps required to solve a certain state, then there are 262, 144 nodes alone at that level. Representing a list of the 9 squares and what

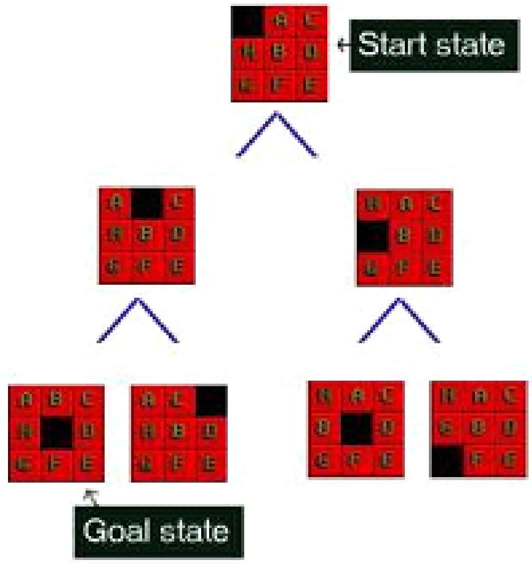

Figure 1: The 8-Puzzle state space for a very simple example

is located in them is all that is required to reach the state of the 8 puzzle game. Here are two examples of states; the final one is the GOAL state, which indicates that we have reached the point where the solution may be applied. The first is an example that has been messed up, and you may start from that.

Start state SPACE, A, C, H, B, D, G, F, E

Objective state The letters A, B, C, H, D, G, F, and E The guidelines that you can use to solve the problem are not overly complicated either. You are able to move a tile into an empty area if there is an empty tile above, below, to the left or to the right of the tile you want to move, respectively. In order to solve the problem, you will need to identify the route that leads from the initial state, through the graph, and finally into the target state.

Path finding: You want to search a state space and figure out how to travel from place you are to somewhere you want to be, without running into barriers or going too far, in a pathfinding scenario, such as a video game or another type of pathfinding scenario. The A* algorithm will not only locate a path, if there is one, if there is one, but it will find the shortest way. We shall learn why this is the case later on in this lesson. When it comes to pathfinding, a state is nothing more than a location in the world. In the case of a labyrinth game like Pac-Man, for instance, you might use a straightforward two-dimensional grid to depict where everything is located. The 2D coordinate of the location where the ghost is when the search begins would be considered the ”start state” for a ghost, for example. The destination point is where Pac-Man is located so that we can go there and devour him. On github, you can also find sample code that demonstrates how to do pathfinding.

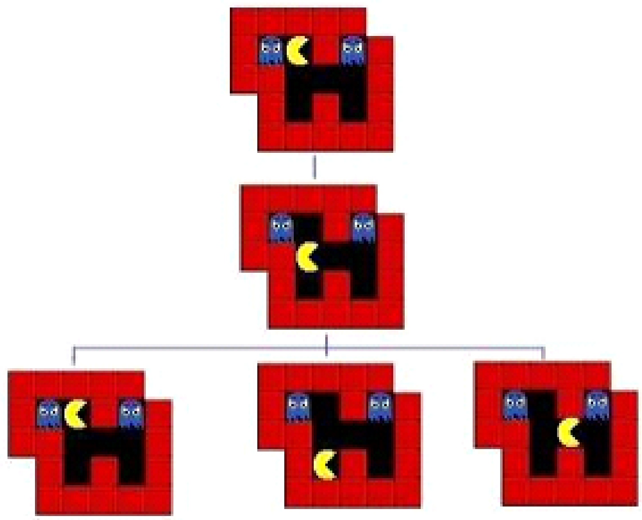

Figure 2: The first three steps of a pathfinding state space

What is data science?: Data science is an interdisciplinary field that uses mathematics and statistics, specialized programming, sophisticated analytics, artificial intelligence (AI), and machine learning in conjunction with in-depth knowledge of a particular subject area to mine an organization’s data for insights that can be put into action. These realizations can serve as a basis for decision-making and can be included into strategic planning. Because of the rapidly expanding number of data sources and, as a result, data, data science has become one of the fields that is seeing the most rapid growth across all industries. As a consequence of this, it should not come as a surprise that the position of data scientist has been referred to as the ”sexiest job of the 21st century” by Harvard Business Review (link is located outside of IBM). Their ability to understand data and make ideas that may be put into action to improve business results is becoming increasingly important to the organizations who hire them. It is possible for analysts to derive insights that may be put into action thanks to the data science lifecycle, which includes a variety of roles, tools, and procedures. The following are the typical steps that a project including data science will go through:

- **Data ingestion:** The first step in the lifecycle is the collecting of data, which may include raw structured data as well as unstructured data from any and all relevant sources utilizing a number of different ways. Manual input, online scraping, and the continuous streaming of data from systems and devices in real time are some examples of these approaches. The term "data sources" can refer to both structured data, such as information on customers, and unstructured data, which can include log files, video, audio, and image files, as well as information from the Internet of Things (IoT), social media, and other places.

- **Data storage and data processing:** Because the formats and architectures of data might vary, businesses need to take into consideration various storage systems depending on the type of data that has to be collected. The workflows associated with analytics, machine learning, and deep learning models are made easier because data management teams contribute to the establishment of standards on data storage and structure. During this step, the data is cleaned, deduplicated, transformed, and combined with the use of ETL processes (extract, transform, load) or other data integration technologies. Before putting data into a data warehouse, data lake, or any other repository of any kind, this data preparation is very necessary for improving data quality.

- **Data analysis:** In this step of the process, data scientists perform an exploratory data analysis in order to investigate the biases, patterns, ranges, and distributions of values that are included within the data. The creation of hypotheses for a/b testing is driven by the investigation of this data analytics. It also gives analysts the ability to decide whether or not the data is relevant for use in modeling attempts for predictive analytics, machine learning, or deep learning. To the extent that a model is accurate, businesses may learn to rely on the insights it provides for the purpose of making business decisions, which in turn enables them to achieve greater scalability.

- **Communicate:** In the end, insights are delivered in the form of reports and other data visualizations that make it simpler for business analysts and other decision-makers to comprehend the insights themselves as well as the impact the insights have on the organization. A data science programming language like R or Python has components for producing visuals; alternatively, data scientists might generate visualizations using tools that are specifically designed for that purpose.

3 Conclusion

It is very necessary to have a solid grasp of fundamental mathematical principles, particularly algebra, in order to succeed in the rapidly developing disciplines of data science and artificial intelligence. Even though "The Algebraic Lens: Understanding Data Science and Artificial Intelligence" is not a specific resource that I am familiar with, the title indicates that it is important to use algebraic concepts in order to get insights into these fields. Because having knowledge that focuses on the intersection of algebra and data science may equip individuals to flourish in the fast evolving area of artificial intelligence and data analysis, it is recommended to investigate books, courses, and resources that focus on this intersection in order to keep current and informed. Continue gaining knowledge and adjusting your strategies to account for any new information that comes your way in order to maintain your position of leadership in these fascinating areas.

References

[1] S. Akgun and C. Greenhow, *Artificial intelligence in education: Addressing ethical challenges in K-12 settings*, AI Ethics, 2(2008), 431-440.

[2] V. Aleven, E. A. McLaughlin, R. A. Glenn and K. R. Koedinger, *Instruction based on adaptive learning technologies*, Handbook of research on learning and instruction, (2008), 522-560.

[3] R. S. Baker, L. Esbenshade, J. Vitale and S. Karumbaiah, *Using demographic data as predictor variables: A questionable choice*, https://doi.org/10.35542/osf.io/y4wvj.

[4] P. Black and D. Wiliam, *Inside the black box: Raising standards through classroom assessment*, Phi Delta Kappan, 92(1)(2010), 81-90.

[5] P. Black and D. Wiliam, *Developing the theory of formative assessment*, Educational Assessment, Evaluation and Accountability, 21(1)(2009), 5-31.

[6] I. Celik, M. Dindar, H. Muukkonen and S. Järvelä, *The promises and challenges of artificial intelligence for teachers: A systematic review of research*, Tech Trends, 66(2011), 616-630.

[7] Center for Integrative Research in Computing and Learning Sciences (CIRCLS), From Broadening to empowering: Reflecting on the CIRCLS'21 Convening, https://circls.org/circls21report, (2010)

[8] A. Doewes and M. Pechenizkiy, *On the limitations of human-computer agreement in automated essay scoring*, In Proceedings of the 14th International Conference on Educational Data Mining (EDM21), (2012).

[9] Z. Ersozlu, S. Ledger, A. Ersozlu, F. Mayne and H. Wildy, *Mixed-reality learning environments in teacher education: An analysis of TeachLivETM Research*, SAGE Open, 11(3)(2012).

[10] S. Forsyth, B. Dalton, E. H. Foster, B. Walsh, J. Smilack and T. Yeh, *Imagine a more ethical AI: Using stories to develop teens' awareness and understanding of artificial intelligence and its societal impacts*, In 2011 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT), (2013).

[11] R. Godwin-Jones, *Big data and language learning: Opportunities and challenges*, Language Learning & Technology, 25(1)(2012), 4-19. http://hdl.handle.net/10125/44747

[12] K. Holstein, B. M. McLaren and V. Aleven, *Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity*, Journal of Learning Analytics, 6(2)(2010).

[13] R. M. Kaplan and D. P. Saccuzzo, *Psychological testing: Principles, applications, and issues*, Cengage Learning, (2008).

[14] S. Merrill, *In schools, are we measuring what matters?*, Edutopia. https://www.edutopia.org/article/schools-are-we-measuring-what-matters, (2012).

[15] E. Nentrup, *How Policymakers Can Support Educators and Technology Vendors Towards SAFE AI*, Ed SAFE AI Alliance, https://www.edsafeai.org/post/how-policymakers-can-supportaied, (2013).