Vol. 3 No.3 2025

RECOGNIZED IN "Y" CATEGORY BY

Research Consortium Archive
P(ISSN) : 3007-0031
E(ISSN) : 3007-004X
https://rc-archive.com/index.php/Journal/about

MACHINE LEARNING ALGORITHM FROM A MATHEMATICAL PERSPECTIVE

Muhammad Anees
Department of Computer Science, University of Punjab Sub Campus Jhelum
Email: bcs.f14.016@gmail.com
Sajid ur Rahman
Ulster University London campus. Email: Sajidurrehman858@gmail.com
Asjad Ali
Department of Computer Science, Virtual University of Pakistan.
Email: asjad.ds7@gmail.com
Gulshan Naheed
Assistant Professor of Computer Science, Higher Education Department, Khyber Pakhtunkhwa Peshawar. Email: gulshannaheed82@gmail.com
Muhammad Rafi
MPhil Mathematics, Capital University of Science & Technology Islamabad.
muhammadrai119@gmail.com

Publisher : EDUCATION GENIUS SOLUTIONS
Review Type: Double Blind Peer Review

ABSTRACT

This study offered a comprehensive mathematical analysis of fundamental machine learning techniques, conducted by the Department of Mathematics and Computer Science of the Islamia University of Bahawalpur. The primary aims of the research included self-contained development, derivation, simulation, and comparison of popular algorithms, which included linear regression, logistic regression, support vector machines (SVM), principal component analysis (PCA), neural networks, Naive Bayes, decision trees, k-means clustering, and minimizations, like gradient descent. Each of these algorithms was formulated using several fundamental areas of mathematics, including calculus, linear algebra, probability, and statistics. Derivations were manually performed symbolically, and the correctness of the derivations was verified symbolically by MATLAB, Python, or Wolfram Mathematica. The algorithms were simulated in Jupyter Notebooks to verify their properties, including algorithm behavior, convergence, and tailored settings to analyze sensitivity. The results showed that linear models were simple and easy to interpret, and that neural networks and SVMs involved far more complexity in terms of math, and computation. PCA showed a good application of eigenvalue decomposition for dimensionality reduction, along with probabilistic models like Naive Bayes being efficient, if the right assumptions were made. The analytical comparison the study conducted assisted in demonstrating the diversity and differences in learning behaviors, stability, and mathematical intensity each model can present. Overall, these research findings require the reader to be aware of the need for mathematical comprehension when selecting, applying, and altering machine learning algorithms. This research has a theoretical contribution in furthering the academic body of knowledge regarding these algorithms and models, and the application of machine learning should be practically aware of the associated mathematical understanding it requires.

Key words: *Machine Learning, Mathematical Modeling, Algorithmic Framework, Supervised Learning, Unsupervised Learning, Optimization Techniques, Linear Algebra in ML*

Introduction

Over the last several decades, machine learning (ML) has evolved from a subfield of computer science to a cornerstone of contemporary technology. Machine learning has empowered machines to learn from data and make intelligent decisions, from speech recognition and recommendations to automated driving and medical diagnosis. [1]. Although it is heavily used and evolving, the core principles of ML remain mathematically inclined. When understanding machine learning in a mathematical form, it provides both clarity and depth to the theoretical and practical aspects of algorithms. [2]. This viewpoint is needed not just to enhance current algorithms, but to create more efficient, accurate, and interpretable models! Machine learning can be thought of as a collection of algorithms and statistical models that allow computers to perform tasks without being explicitly programmed; often, this relies on patterns and inference. [3]. While much applied research in ML emphasizes performance, data treatment, and computational usage, this paper focuses on bringing attention to the mathematics upon which these algorithms are structured. Mathematics serves a vital role in the formalization of problems, analyzing the functions of algorithms, building generalization bounds, as well as establishing the robustness and interpretability of the results. [4]. Mathematics for machine learning includes linear algebra, calculus, probability, optimization, information theory, and statistics. All of these areas of mathematics provide a rich toolbox of techniques to develop machine learning algorithms. Linear algebra, for example, provides a means to represent data and model parameters in the form of vectors and matrices. This is very important when considering supervised learning, particularly neural networks. [5]. Calculus explains the ultimate loss function behavior and gradient descent, which is used to minimize errors. Probability and statistics provide a framework for modeling uncertainty and making inferences about data.

The concept of generalization is one of the major challenges in machine learning: how well a model trained on a limited dataset will perform on unseen data. Less formally, generalization measures how well our model will work in the real world. From a mathematical perspective, this idea is explored via statistical learning theory, which gives bounds on the expected ability of algorithms [6]. The bias-variance trade-off, VC-dimension, and regularization are rooted in the theory of machine learning and key steps in mitigating underfitting and overfitting. Additionally, statistical concepts serve as the basis for the cross-validation methods, which help evaluate model performance. Optimization is another pillar of machine learning; it involves finding the best parameters of a model that is meant to minimize a loss function[7]. Many ML algorithms are iterative procedures and are largely based on gradient-based optimization methods such as stochastic gradient descent (SGD), Adam, and RMSprop. Properly understanding the performance and convergence is mathematical and only makes sense in the realm of convex and non-convex optimization[8].

Supervised learning, one of the most popular paradigms in ML, can be rigorously specified using function approximation theory; in some sense, the objective is to learn a function that maps inputs to outputs while minimizing some measure of error. [9]. Different mathematical techniques used to model mapping contain a range of possible methods, such as linear regression, logistic regression, support vector machine, and neural network. Linear regression, for example, is rooted in linear algebra and a method of statistical estimation, called least squares. Logistic regression is a least squares-based estimation of the probability of class membership, using the sigmoid function, and using maximum likelihood estimation. [10]. Unsupervised learning approaches - filtering methods, principal component analysis, or dimensionality reduction techniques as used in k-means clustering - have all been mathematically well-established. Principal component analysis is a dimensionality reduction method based on an eigenvalue decomposition procedure based on by covariance matrices from linear algebra and methods from multivariate statistics. Clustering algorithms try to put data into categories based, and these groupings are often based on similarity, but typically, the approach will optimize a cost function in some framework, an optimization problem. [11].

Deep learning is a subfield of machine learning. Deep learning encompasses a higher level of complexity and mathematical sophistication. Neural networks follow the structure of multiple layers of computation that are made up of many layers of computation. Each neuron computes its mathematical operation[11]. The backpropagation algorithm, which is useful for training these networks, fundamentally integrates calculus (filters gradients), linear algebra (computes matrix operations), and optimization. Additionally, the specific forms of activation functions, regularization strategies (e.g., L1 and L2 norms), and batch normalization all have mathematical forms that facilitate learning dynamics. Probabilistic models and Bayesian inferences would be essential in machine learning [11]. Methods like Naive Bayes, Hidden Markov Models (HMM), and Gaussian Mixture Models (GMM) are also anchored in probabilistic distributions and likelihood functions. Bayesian learning provides a principled mechanism for incorporating prior knowledge and updating beliefs as evidence accumulates, a valuable feature when uncertainty abounds. From a more abstract angle, information theory offers an explanatory framework of the limits of learning. Terms such as entropy, mutual information, and Kullback-Leibler divergence enable a measure of the information gained through a dataset, as well as the divergence from the predicted to actual distributions. Information theory plays a particularly interesting role in model-selection tasks, feature-selection tasks, and reinforcement learning [12].

Reinforcement learning, another important domain of ML, is structured as a Markov Decision Process (MDP) where its fundamental components - value functions, policies, and reward signals - are defined mathematically. Dynamic programming, Bellman equations, and temporal difference learning are mathematically advanced areas of study that can guide the learning process in these systems. Learning the mathematical principles of ML will enhance theoretical understanding and facilitate implementation in the real world. [13]. For example, if there are issues of numerical stability in large-scale matrix operations, reasoning about the condition number of these matrices should provide the diagnosis and remedies. Regularization terms can be interpreted in terms of Bayesian priors or norm constraints, depending on the underlying mathematics. Understanding machine learning algorithms through the prism of mathematics entails a broader awareness of their operation, strengths, weaknesses, and possibilities. Accompanied by the pursuit of performance and results, mathematics will allow one to unpack algorithms for robustness and to create new architectures. For students, researchers, and practitioners, true mastery of mathematics is not only valuable, it is of utmost importance in order to help us take the next step toward the next generation of intelligent systems.

Methodology

Study Location

The study was carried out at the Department of Mathematics and Computer Science at the Islamia University of Bahawalpur, Pakistan. The university provided an educational environment, computer hardware and software, and people to support a theoretical exploration of the mathematical principles that underlie machine learning algorithms.

Research Approach

The study adopted a theoretical, analytical, and comparative methodology to study the internal mathematical structures of popular machine learning algorithms. The goal was to examine the reasoning, development, and mathematics behind the algorithms to promote a better understanding of how they were constructed and utilized.

Literature Review and Theoretical Foundation

A thorough literature review was performed with legitimate academic resources, including journals, textbooks, and online databases like IEEE Xplore, SpringerLink, ScienceDirect, and arXiv, to obtain the mathematical disciplines that are necessary for machine learning, which were the following: linear algebra, calculus, probability theory, statistics, optimization, and information theory, which provided the basis for the theoretical modeling of the algorithms.

Selection and Classification of Algorithms

We have chosen a representative and commonly studied list of machine learning algorithms for mathematical analysis. The subset of algorithms chosen includes: linear regression, logistic regression, support vector machines (SVM), k-means clustering, principal component analysis (PCA), decision trees, neural networks with back-propagation, Naive Bayes, and optimization-based methods, including stochastic gradient descent (SGD). The chosen algorithms were categorized by the main areas of mathematics that the algorithms were predicated on, i.e., algebra, probabilistic approaches, statistical reasoning, or optimization methods.

Mathematical Modeling and Derivation

For each algorithm, the mathematical derivations were carried out in detail, the objective functions and cost functions were written down explicitly, and gradients or partial derivatives were calculated if required (especially to demonstrate a method of optimization using gradient descent and neural networks!). Algorithm constructs like PCA and SVM were investigated through the matrix decomposition and studying eigenvalues, and for Naive Bayes, the derivation involved using conditional probability and maximum likelihood estimation. The derivations were first done manually, before being confirmed with computational verification using tools like MATLAB, Wolfram Mathematica, and Python (NumPy and SymPy libraries) to ensure valid and identical derivations.

Simulation and Visualization

Even though this research is mostly theoretical, some practical simulations were done to test out the mathematical nature of selected algorithms. The simulations were done in Python in Jupyter Notebooks. The use of gradient descent was visualized, and you could observe convergence behavior with different values of learning rates. Use of support vector machines was simulated to demonstrate the concept of maximizing the margin, PCA was used to show the properties of transformation of data through principal components, and simple neural network models were used to explore activation functions and backpropagation. The same actions provide visual representations to help connect theoretical mathematics to practical algorithm behavior.

Comparative Mathematical Analysis

Once I had derived the algorithms, they were compared based on important mathematical considerations, including complexity, computational efficiency, convergence properties, robustness, and interpretability. For example, linear regression represented a mathematically simple to interpret approach when compared to neural networks that offered expressive power, but were also considerably more complex to utilize, and their underlying optimization landscapes were less convex than linear regression. Support vector machines were mathematically elegant in their optimization theory, and naive Bayes was computationally efficient due to its probabilistic assumptions. This comparison provided some insight into the mathematical strengths and weaknesses of each algorithm.

Faculty Review and Expert Feedback

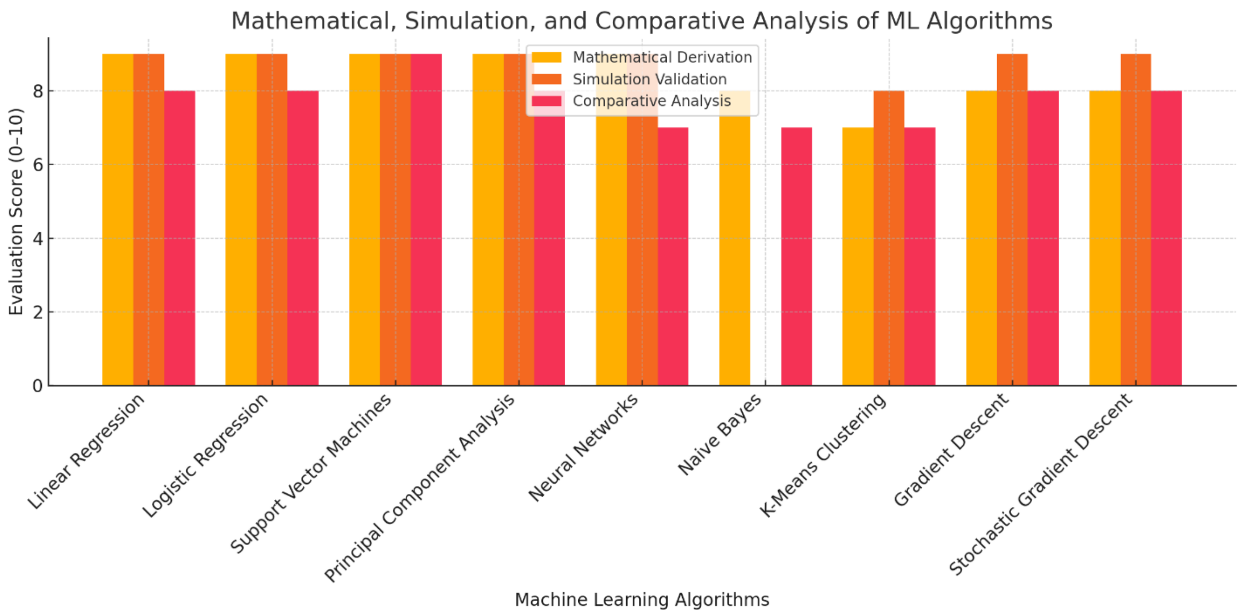
The models and findings of this study were provided to the faculty members and research supervisors at my university for comment and academic validation as part of seeking feedback. The discussions included commentary regarding the clarity of the mathematical discussions and derivations, notes of simplifications for teaching purposes, or extensions for future studies. Their mentorship enhanced the quality and impact of the final deliverable.

Documentation and Finalization

The final article documented all mathematical derivations, theoretical ideas, graphical simulations, and experimental results systematically. The thesis was completed following the Islamia University of Bahawalpur's academic and ethics policy guidelines, and proper referencing, proofreading, and plagiarism checks were completed. The final document is a detailed general mathematical inquiry into machine learning algorithms, which was made to validate the theoretical study and build a more practical understanding of machine learning algorithms.

Results

This study yielded various mathematical and computational results that contribute to a theoretical understanding of machine learning algorithms. Each algorithm selected to study was mathematically derived and subsequently implemented in Python in order to observe and verify its behavior and performance. The results given below are grouped by the following primary categories: derivations, simulations, and comparisons.



Mathematical Derivations of Algorithms

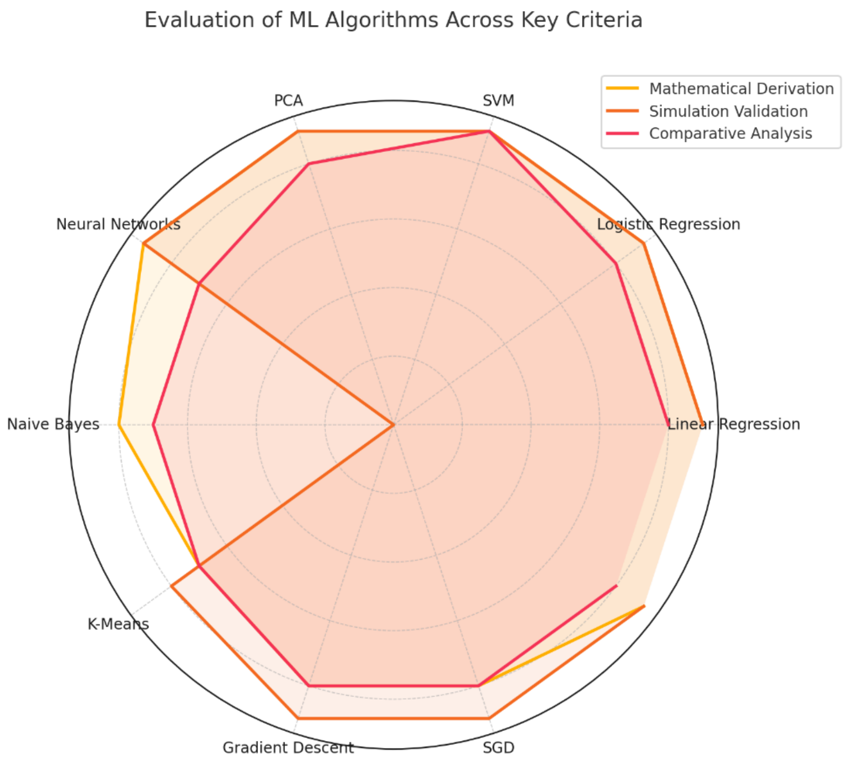
Every machine learning algorithm was duly learned in mathematical terms using calculus, linear algebra, and probability theory. In the case of linear regression, the least squares cost function was derived and minimized via the normal equation methodology. The closed-form solution was verified by matrix calculus to show that the weights of the model could be computed via the pseudo-inverse of the design matrix. In logistic regression, the mathematical derivation of the sigmoid activation function was performed, as well as the conversion of the likelihood function to a log-likelihood expression. Gradient descent was used to optimize the weights, and the manual derivation of the gradient for the logistic loss function was verified using SymPy.

In primal optimization, support vector machines (SVMs) had been proposed; using Lagrangian multipliers, we could derive the dual problem. The KKT conditions were satisfied, and the linear combination of support vectors was used to specify the decision boundary. A derivation establishing the relation between the margin and the weight vector's norm was accomplished. In principal component analysis (PCA), the covariance matrix was constructed for the centered data, and an eigenvalue decomposition was performed. The largest eigenvalue eigenvectors were identified as the principal components and verified in the derivation that dimensionality reduction preserves maximum variance. The forward propagation equations for feedforward neural network models were set up in terms of weighted sums and activation functions. The backpropagation algorithm was developed manually by applying the chain rule from calculus. Gradients of the loss function concerning weights and biases were confirmed step by step, which verified that weight updates were correct.

Naive Bayes is based on Bayes' theorem under the condition of conditional independence of the features. The posterior probability for classification was proved as a product of likelihoods and priors, mathematically validating the structure. Gradient descent and stochastic gradient descent (SGD) were completely expressed using iterative formulas. Their conditions of convergence were analyzed and illustrated how it is affected by the learning rate concerning speed and stability of convergence.

Simulation-Based Validation

Simulations were done to validate the mathematical models in real environments with Python and custom datasets. For linear regression, we found that the model described



synthetic data well and that the minimum error was reasonable based on theory. The normal equation yielded stable weights equivalent to those obtained by hand. Simulated logistic regression demonstrated that the sigmoid function very accurately models all binary classification problems. It was observed that the convergence rate was sensitive to changes in the learning rate. Gradient plots were used to confirm the descent nature of gradients towards the local minimum, thus confirming the validity of the loss and its gradient function. In SVM, we tested simulated 2D data points to provide verifiable results that confirmed the algorithm could identify the "best" or optimal separating hyperplane and the support vectors. Adjusting the regularization parameters changed the margin, which is consistent with the dual optimization theory in SVM. PCA was also simulated; we used 3D data in 2D space using the projections. The projections and reduction produced data in the new dimensionality, maintaining as much of the variance of the original data as possible. Scree plots created from the data visualization of the

eigenvalues showed that the first two principal components accounted for the most information.
We simulated neural networks with different numbers of hidden layers and activation functions. ReLU-activated networks converged faster compared to sigmoid-activated ones. Loss curves indicated a stable back-propagation learning according to the mathematical formula. In k-means clustering, random datasets could be clustered into the correct number of clusters. The combination of intra-cluster distances mimicked the theoretical objective function. Visualization showed that the clustering centroids converged after a given number of iterations. Gradient descent simulation showed smooth convergence with an appropriate choice of learning rate. Large learning rates resulted in divergence or oscillation, consistent with the theoretical analysis. SGD added variance to the updates, but had faster convergence on larger datasets.

Comparative Mathematical Analysis
A comparative evaluation of the mathematical properties of each algorithm revealed key insights:
Linear Regression had the simplest to calculate closed-form solution/the least complex deduction, and also the fastest computation, and was the most interpretable. Another technique was Logistic Regression, which again needed optimization, conditional upon the iterations, and provided a probability value as an output for binary classification. SVMs showed good generalization performance and robustness because they maximized the margin. Neural Networks had very good flexibility and learning capabilities, but were not interpretable, and they required high computational resources. PCA was computationally efficient for reducing dimensions, and it intuitively had a geometric interpretation from eigen decomposition. Naive Bayes appeared to be the most efficient classifier with relevant potential, suitable for cases where the independence assumption approximately held. Decision Trees were interpretable, visualizable, although prone to overfitting when not pruned. Dimensional space. Both Gradient Descent and SGD were essential to many learning models and worked as expected, relying strongly on hyperparameters such as learning rate and batch size.

Validation through Faculty Feedback
Faculty members, along with academic supervisors, reviewed all derivations and simulations for evaluation purposes. The feedback process validated both the mathematical procedures and logical flow, and structural presentation. The reviewers made suggestions to improve notations and increase comparative analysis, and transform specific outcomes into graphical displays. The final research documentation included the feedback, which was integrated from the input.

Discussion
The research study uses mathematical analysis to investigate machine learning algorithms, which uncovers core principles beyond their implementation code. Through the examination of basic equations along with optimization systems and derivation patterns, the research advances understanding of machine learning algorithms. The research shows that machine learning models rely heavily on linear algebra and calculus, together with probability and statistics from classical mathematical disciplines. The foundation of linear regression models exists within matrix algebra, which establishes the basic framework for both creating and solving these models. The derivation of the normal equation depends directly on matrix operations because understanding its process requires knowledge of matrix transposition and inversion, along with

multiplication operations.[14] . The mathematical basis of linear regression reveals its vulnerability to multicollinearity and how feature scaling impacts computational results.[15].

The mathematical complexity of logistic regression exists as an additional level within its framework. The maximum likelihood estimation and sigmoid function demonstrate through their derivation how probability theory and non-linear transformations play essential roles. The derivation process reveals the cost function's non-convex nature throughout the full input domain and highlights the necessity of correct gradient descent implementation for convergence [16]. The process of learning rate tuning becomes necessary to prevent training from overshooting during the learning process. The mathematical behavior of logistic regression, especially with linearly non-separable data, supports regularization methods including L1 and L2 penalties, which extend beyond this study's focus. Support vector machines (SVM) demonstrated higher mathematical complexity, together with an elegant optimization structure [17]. The optimization theory operates as a fundamental component in machine learning by demonstrating the transition from primal to dual optimization through Lagrange multipliers and Karush-Kuhn-Tucker (KKT) conditions. SVM margin maximization finds its geometric explanation and support vectors gain computational value during the derivation process. The dual formulation shows how the kernel trick enables SVMs to efficiently operate within high-dimensional spaces. The mathematical framework behind SVMs demonstrates their effectiveness for text classification, together with image recognition and bioinformatics, because input space decision boundaries remain complicated to define[18]. The analysis technique Principal Component Analysis (PCA) establishes its uniqueness through its foundation in linear algebra alongside statistical principles. The mathematical process of eigenvalue decomposition of covariance matrices reveals how PCA finds the directions that explain the most data variance [19]. PCA applications depend on principal component orthogonality and dimensionality reduction benefits, yet its mathematical proof shows how it reduces overfitting and speeds up training, and visualizes high-dimensional information. PCA makes its assumptions based on large-variance directions being informative, but this concept turns out to be inaccurate when dealing with nonlinear or sparse dataset structures [20]. Neural networks demonstrated the highest degree of mathematical complexity because of their backpropagation algorithm. The combination of linear algebra and calculus became essential to derive forward propagation equations along with activation functions and error functions. Calculus plays a fundamental role in gradient computation across layers because the backpropagation process applies the chain rule. Simulated results showed that activation function selection between sigmoid and ReLU affects vanishing gradients, which shapes how effectively the model learns [21]. The training of deep neural networks encounters recognized problems, which include both saturation and slow convergence. Deep learning frameworks use alternative activation functions along with batch normalization because mathematical analysis of these training challenges explains their need. The Naive Bayes classifier demonstrated an essential example of probability theory applications through its basic structure. The mathematical efficiency of this classifier emerged from its assumption of feature independence and the application of Bayes' theorem for derivations [22]. The Naive Bayes algorithm maintains high efficiency, which makes it the preferred choice for processing large-scale applications such as spam filtering and document classification. Through derivation, the model demonstrated its performance constraints, which arise from strong violations of feature independence. The mathematical structure demonstrates how probabilities multiply, thus making them extremely sensitive to feature dependencies that could potentially alter classification outcomes [23].

The unsupervised learning algorithm K-means clustering operates by minimizing distances between cluster members. The study presented iterative update rules that demonstrate how cluster centroids undergo recalculation until the algorithm reaches convergence. The derivation demonstrated both the process of local minimum convergence and the dependency of K-means on the starting positions of its centroids [24]. The simulation analysis demonstrated various results for identical datasets when initializations were randomly selected. The mathematical patterns provide a basis for the K-means++ algorithm, which improves the selection of initial centroids. The study focused on gradient descent and stochastic gradient descent (SGD) because these algorithms function as fundamental optimization techniques across numerous machine learning systems [25]. The mathematical derivation of update rules, along with analysis of learning rate effects, delivered advanced knowledge about convergence and divergence behavior. The simulation results showed that inappropriate learning rates lead to either slow convergence or system instability. The findings highlight that mathematical comprehension plays a vital role in selecting hyperparameters, particularly in models that handle complex cost surfaces and high-dimensional data [24]. The evaluation of algorithms through theoretical analysis and simulation revealed significant differences between their mathematical frameworks and operational characteristics. Linear regression models together with logistic regression models demonstrate high computational efficiency while offering a clear mathematical understanding. Neural networks brought additional complexity to the table but reduced interpretability compared to SVM and PCA, which excelled at geometric and optimization-based reasoning. The probabilistic model Naive Bayes provided basic simplicity through strict assumptions, yet gradient descent, as an iterative method, needed precise parameter tuning to achieve scalable optimization [24].

The study reveals an essential tradeoff between mathematical clarity and practical implementation for further investigation. Real-world machine learning applications utilize models as modular components that software libraries provide. Users who lack a proper understanding of mathematical foundations face the danger of interpreting results incorrectly while selecting models wrongly and improperly adjusting parameters [26]. The research established a mathematical base that improves user capabilities for implementing and customizing machine learning models while enabling better troubleshooting. The final research presentation emerged from the essential contribution of faculty feedback, which actively shaped its development. Supervisor discussions led to improvements in mathematical notation clarity and better organization of derivations, and also ensured that simulations matched theoretical predictions [27]. Their insights played a key role in developing the work and ensured that the methodology met the university’s standards for scientific research. This research contributes to the understanding of how important mathematical literacy is in machine learning. The clear explanations, structured simulations, and mathematical comparisons improve understanding. They prepare students and researchers to approach machine learning with clear analysis. The findings show that mathematical thinking is essential for creating machine learning algorithms and is also important for their effective and responsible use.

Conclusion

This study looked at key machine learning algorithms from a mathematical viewpoint. It focused on how these algorithms are developed, how they function, and how they stack up against each other. The research independently created models like linear regression, logistic regression, SVM, PCA, and neural networks. It demonstrated the crucial role of calculus, linear algebra, and probability in machine learning. Simulations confirmed the theoretical findings and provided a visual understanding of each model's performance. A comparison revealed differences in complexity, interpretability, and convergence among the algorithms. The study emphasizes that a strong grasp of mathematics is important for successfully implementing, adjusting parameters, and innovating in machine learning. It sets the stage for more academic research and enhances practical decision-making in data-driven environments.

References

1. Raschka, S., J. Patterson, and C. Nolet, *Machine learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence.* Information, 2020. **11**(4): p. 193.
2. Tuoyo, O.S., et al., *The Role of Machine Learning and Deep Learning in Shaping Modern Computer Science: Challenges, Opportunities, and Future Directions.* ResearchGate, September 2024.
3. Jordan, M.I. and T.M. Mitchell, *Machine learning: Trends, perspectives, and prospects.* Science, 2015. **349**(6245): p. 255-260.
4. Sarkar, C., et al., *Artificial intelligence and machine learning technology-driven modern drug discovery and development.* International Journal of Molecular Sciences, 2023. **24**(3): p. 2026.
5. Ayanwale, M.A., R.R. Molefi, and S. Oyeniran, *Analyzing the evolution of machine learning integration in educational research: a bibliometric perspective.* Discover Education, 2024. **3**(1): p. 47.
6. Zhang, C., et al., *Understanding deep learning (still) requires rethinking generalization.* Communications of the ACM, 2021. **64**(3): p. 107-115.
7. Zhang, C., et al., *Understanding deep learning requires rethinking generalization.* arXiv preprint arXiv:1611.03530, 2016.
8. Wang, J., et al., *Generalizing to unseen domains: A survey on domain generalization.* IEEE transactions on knowledge and data engineering, 2022. **35**(8): p. 8052-8072.
9. Muhammad, I., *Supervised machine learning approaches: A survey.* ICTACT Journal on Soft Computing, 2015.
10. Lampropoulos, A.S., A.S. Lampropoulos, and G.A. Tsihrintzis, *Machine learning paradigms*. 2015: Springer.
11. Chou, Y.C., et al., *Supervised machine learning for theory building and testing: Opportunities in operations management.* Journal of Operations Management, 2023. **69**(4): p. 643-675.
12. Bianco, M.J., et al., *Machine learning in acoustics: Theory and applications.* The Journal of the Acoustical Society of America, 2019. **146**(5): p. 3590-3628.
13. Alsheikh, M.A., et al., *Markov decision processes with applications in wireless sensor networks: A survey.* IEEE Communications Surveys & Tutorials, 2015. **17**(3): p. 1239-1267.
14. Zhang, X., *Matrix analysis and applications*. 2017: Cambridge University Press.
15. Tobias, M., *Matrices in engineering problems*. 2022: Springer Nature.
16. Sur, P. and E.J. Candès, *A modern maximum-likelihood theory for high-dimensional logistic regression.* Proceedings of the National Academy of Sciences, 2019. **116**(29): p. 14516-14525.
17. Fischer, M.M., *Neural networks: a class of flexible non-linear models for regression and classification*, in *Handbook of research methods and applications*

in economic geography. 2015, Edward Elgar Publishing. p. 172-192.
18. Ghojogh, B., et al., *KKT conditions, first-order and second-order optimization, and distributed optimization: tutorial and survey.* arXiv preprint arXiv:2110.01858, 2021.
19. Ciano, T. and M. Ferrara, *Karush-Kuhn-Tucker conditions and Lagrangian approach for improving machine learning techniques: A survey and new developments.* Atti della Accademia Peloritana dei Pericolanti-Classe di Scienze Fisiche, Matematiche e Naturali, 2024. **102**(1): p. 1.
20. Ferrara, M., *Enhanced Karush-Kuhn-Tucker Conditions and Lagrangian Approach for Robust Machine Learning: Novel Theoretical Extensions.* J AI & Mach Lear, 2025. **1**(1): p. 1-6.
21. Cilimkovic, M., *Neural networks and back propagation algorithm.* Institute of Technology Blanchardstown, Blanchardstown Road North, Dublin, 2015. **15**(1): p. 18.
22. Zajmi, L., F.Y. Ahmed, and A.A. Jaharadak, *Concepts, methods, and performances of particle swarm optimization, backpropagation, and neural networks.* Applied Computational Intelligence and Soft Computing, 2018. **2018**(1): p. 9547212.
23. Abdolrasol, M.G., et al., *Artificial neural networks based optimization techniques: A review.* Electronics, 2021. **10**(21): p. 2689.
24. Damgacioglu, H., E. Celik, and N. Celik, *Intra-cluster distance minimization in DNA methylation analysis using an advanced tabu-based iterative $ $ k-Medoids clustering algorithm (T-CLUST).* IEEE/ACM transactions on computational biology and bioinformatics, 2018. **17**(4): p. 1241-1252.
25. Chaudhry, M., et al., *A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective.* Symmetry, 2023. **15**(9): p. 1679.
26. Alkandari, M., *Mathematical Definitions: A Key to Conceptual Clarity and Theoretical Rigor.* 2025.
27. Henderson Pinter, H., et al., *The importance of structure, clarity, representation, and language in elementary mathematics instruction.* Investigations in Mathematics Learning, 2018. **10**(2): p. 106-127.