Royal International Global Journal of Advance and Applied Research
Peer Reviewed International, Open Access Journal.
ISSN: 2998-4459 | Website: https://rlgjaar.com Volume-2, Issue-3 | March - 2025

Original Article

Mathematical Foundations and Statistical Applications of Artificial Intelligence and Machine Learning

Ajay Hanamant Tarange
Assistant Professor, Department of Education Shivaji University, Kolhapur

Manuscript ID:
RIGJAAR-2025-0203036

ISSN: 2998-4459

Volume 2

Issue 3

Pp. 173-176

March 2025

Submitted: 18 Feb 2025

Revised: 27 Feb 2025

Accepted: 15 Mar 2025

Published: 31 Mar 2025

Correspondence Address:
Ajay Hanamant Tarange
Assistant Professor, Department of Education
Shivaji University, Kolhapur
Email: ajaytarange01@gmail.com

Quick Response Code:

Web. https://rlgjaar.com

DOI: 10.5281/zenodo.22953625

DOI Link:
https://doi.org/10.5281/zenodo.22953625

Creative Commons

Abstract

Artificial Intelligence (AI) and Machine Learning (ML) are strongly dependent on mathematical structures that support representation, learning, optimization, uncertainty quantification and generalization. This paper rewrites and extends the supplied review on mathematical foundations of AI and ML by adding a quantitative, secondary-data perspective. Linear algebra, probability and statistics, optimization, information theory and statistical learning theory are examined as interconnected components of modern AI. The statistical component uses documented benchmark datasets and published benchmark results rather than invented experimental observations. The MNIST database contains 60,000 training examples and 10,000 test examples, while Fashion-MNIST contains 60,000 training and 10,000 test examples distributed across 10 balanced classes. For MNIST, the documented test error rates of a linear classifier are 12.0% without preprocessing and 8.4% with deskewing; on the 10,000-image test set these correspond to approximately 1,200 and 840 misclassified images, respectively. The analysis demonstrates how mathematical and statistical measures convert abstract AI concepts into measurable evidence. The paper concludes that mathematical foundations are not merely theoretical prerequisites but practical tools for model selection, performance evaluation, uncertainty analysis and responsible AI.

Keywords: *Mathematics; Artificial Intelligence; Machine Learning; Linear Algebra; Probability; Statistics; Optimization; Information Theory; Statistical Learning Theory; MNIST; Fashion-MNIST; Deep Learning.*

Introduction

Artificial Intelligence aims to develop systems capable of perception, reasoning, decision making, learning and action. The supplied source emphasizes that mathematical rigor helps formulate AI problems precisely, identify logical relationships and analyse models. AI and ML therefore depend on mathematical descriptions of data, functions, objectives and uncertainty. Representation, functional approximation and clustering provide a mathematical language for understanding how learning systems transform inputs into outputs. Machine Learning can be viewed as the construction of mathematical functions from data. In supervised learning, a model estimates a relationship between inputs and known targets; in unsupervised learning, it identifies structure without target labels; and in reinforcement learning, an agent learns from observations, actions and rewards. These paradigms require mathematical tools for optimization, probability, statistics and numerical computation.

The purpose of this rewritten paper is to preserve the mathematical emphasis of the supplied document while strengthening its research-paper structure through a transparent statistical analysis of established benchmark data. The quantitative section is deliberately based on documented secondary sources; no fabricated survey or laboratory observations are presented.

Objectives of the Study

- To explain the principal mathematical foundations underlying AI and ML.
- To examine the role of linear algebra, probability, statistics, optimization and information theory.
- To connect statistical learning theory with model complexity and generalization.

Creative Commons (CC BY-NC-SA 4.0)
This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License, which allows others to remix, tweak, and build upon the work noncommercially, as long as appropriate credit is given and the new creations ae licensed under the idential terms.

How to cite this article:
Tarange, A. H. (2025). Mathematical Foundations and Statistical Applications of Artificial Intelligence and Machine Learning. Royal International Global Journal of Advance and Applied Research, 2(3), 173–176. https://doi.org/10.5281/zenodo.22953625

Royal International Global Journal of Advance and Applied Research
Peer Reviewed International, Open Access Journal.
ISSN: 2998-4459 | Website: https://rlgjaar.com Volume-2, Issue-3 | March - 2025

- To demonstrate the use of descriptive statistical measures on established AI benchmark datasets.
- To interpret documented classification error rates using absolute error counts and percentage measures.
- To identify implications of mathematical and statistical analysis for reliable and responsible AI.

Research Methodology

This study adopts a secondary-data and analytical review methodology. The conceptual material is based primarily on the supplied paper, which discusses mathematical prerequisites, machine learning paradigms, statistical learning theory, optimization, probabilistic reasoning, information theory, applications and responsible AI. Quantitative observations are taken from publicly documented benchmark dataset descriptions and published benchmark results.

The statistical analysis uses frequency, percentage, class distribution, train-test proportions, absolute error count and relative improvement. For a reported error rate e and test size N, the estimated number of classification errors is calculated as $E = N \times e/100$. Percentage improvement in error rate is calculated as $((e1 - e2)/e1) \times 100$. These calculations are descriptive and derived directly from the documented benchmark values.

Mathematical Foundations of AI and ML

1. Linear Algebra

Linear algebra provides the language for vectors, matrices, transformations and high-dimensional representations. In machine learning, an input observation may be represented by a vector x, model parameters by a vector or matrix, and a linear transformation by $Wx + b$. Eigenvalues, eigenvectors, matrix factorization and vector norms are useful in dimensionality reduction, optimization, classification and neural-network computation.

2. Probability and Statistics

Probability supplies a formal framework for uncertainty, while statistics provides methods for learning from finite samples. Mean, variance, covariance, correlation, probability distributions, likelihood and hypothesis testing are central to data analysis. In supervised learning, empirical risk is estimated from observed samples; in probabilistic AI, posterior distributions quantify uncertainty after evidence is observed.

3. Optimization

Learning commonly involves minimizing a loss function or maximizing an objective function. Gradient descent and stochastic gradient descent use derivatives to update model parameters. Convex optimization provides conditions under which local and global optima coincide, while stochastic optimization is particularly important when datasets and models are large.

4. Information Theory

Entropy measures uncertainty and mutual information measures statistical dependence between variables. These concepts are important in feature selection, representation learning, decision trees, probabilistic modelling and information-bottleneck approaches.

Machine Learning and Statistical Learning Theory

The supplied paper describes statistical learning theory as a mathematical framework for analysing the relationship between model complexity and generalization from finite training samples. Important concepts include empirical and expected risk, approximation error, estimation error, VC dimension, sample complexity, the bias-variance trade-off and Probably Approximately Correct (PAC) learning.

A simplified decomposition of prediction error is useful for interpretation: model performance reflects the interaction of approximation limitations, estimation from finite data and implementation or optimization effects. Increasing model complexity can reduce bias but may increase variance, which is why validation and regularization are important.

Statistical Data and Quantitative Analysis

The quantitative analysis uses two established image-classification benchmark datasets. MNIST contains 60,000 training examples and 10,000 test examples. Fashion-MNIST contains 60,000 training examples and 10,000 test examples, with 10 classes and 28×28 grayscale images. These datasets are useful for illustrating the relationship between dataset structure, statistical evaluation and machine-learning performance.

Dataset	Training	Test	Total	Classes	Image Format
MNIST	60,000	10,000	70,000	10	28×28 grayscale
Fashion-MNIST	60,000	10,000	70,000	10	28×28 grayscale

For both datasets, the training proportion is 60,000/70,000 = 85.71%, while the test proportion is 10,000/70,000 = 14.29%. In Fashion-MNIST, the 70,000 observations are evenly distributed over 10 classes, giving 7,000 observations per class; the predefined split corresponds to 6,000 training and 1,000 test images per class.

Measure	MNIST	Fashion-MNIST
Total observations	70,000	70,000
Training observations	60,000 (85.71%)	60,000 (85.71%)
Test observations	10,000 (14.29%)	10,000 (14.29%)
Number of classes	10	10
Observations per class	Dataset is balanced across labels in the standard split	7,000
Training observations per class	—	6,000
Test observations per class	—	1,000

Royal International Global Journal of Advance and Applied Research
Peer Reviewed International, Open Access Journal.
ISSN: 2998-4459 | Website: https://rlgjaar.com Volume-2, Issue-3 | March - 2025

Statistical Interpretation of MNIST Benchmark Results

The official MNIST documentation reports a 12.0% test error rate for a linear classifier without preprocessing and an 8.4% test error rate when deskewing is applied. Because the official test set contains 10,000 examples, these percentages can be translated into absolute counts.

Method	Test Size	Error Rate	Estimated Errors	Estimated Correct	Accuracy
Linear classifier, no preprocessing	10,000	12.0%	1,200	8,800	88.0%
Linear classifier, deskewing	10,000	8.4%	840	9,160	91.6%

The absolute reduction is 360 classification errors per 10,000 test images. The error rate falls from 12.0% to 8.4%, an absolute reduction of 3.6 percentage points and a relative reduction of 30.0% in the reported error rate. This result illustrates an important mathematical principle: a comparatively simple preprocessing transformation can alter the statistical distribution of the input and consequently change classification performance.

Descriptive Statistical Measures for AI Evaluation

Metric	Formula	Purpose in AI/ML
Mean	$\bar{x} = \Sigma x_i / n$	Central tendency of numerical variables
Variance	$\sigma^2 = \Sigma (x_i - \bar{x})^2 / n$	Measures dispersion
Standard deviation	$\sigma = \sqrt{\sigma^2}$	Quantifies spread around the mean
Accuracy	(TP + TN) / Total	Overall classification correctness
Error rate	Errors / Total × 100	Percentage of incorrect predictions
Precision	TP / (TP + FP)	Reliability of positive predictions
Recall	TP / (TP + FN)	Coverage of actual positives
F1-score	2PR / (P + R)	Balance between precision and recall

These measures allow researchers to move beyond qualitative descriptions of AI performance. Accuracy is useful for balanced classification problems, whereas precision, recall and F1-score become especially important when classes are imbalanced or when the costs of false positives and false negatives differ.

Optimization and Statistical Performance

Optimization links mathematical theory with empirical model performance. A learning algorithm seeks parameters θ that minimize an objective such as $L(\theta)$. In stochastic gradient descent, an update may be represented as $\theta(t+1) = \theta(t) - \eta\nabla L(\theta(t))$, where η is the learning rate. Statistical evaluation is required because minimizing training loss alone does not guarantee strong performance on unseen data. Generalization, validation error and test error therefore provide complementary evidence.

Probabilistic Reasoning and Uncertainty

Bayesian reasoning represents uncertainty through prior, likelihood and posterior quantities. For a hypothesis H and data D, Bayes' rule is $P(H|D) = P(D|H)P(H)/P(D)$. Probabilistic graphical models extend this framework by representing dependencies between variables. When exact inference is computationally difficult, approximate inference methods such as variational approaches and Monte Carlo sampling can be used.

Information-Theoretic Analysis

Entropy is commonly expressed as $H(X) = -\Sigma\, p(x) \log p(x)$. Mutual information can be expressed as $I(X;Y) = \Sigma\, p(x,y) \log[p(x,y)/(p(x)p(y))]$. These measures provide quantitative tools for assessing uncertainty and dependency. In representation learning, they help frame the trade-off between compressing information and retaining information relevant to a target variable.

Applications of Mathematical AI

Application Area	Mathematical/Statistical Tools	Typical Analytical Task
Computer Vision	Linear algebra, probability, optimization	Classification and feature extraction
Natural Language Processing	Probability, vectors, information theory	Language modelling and representation
Robotics and Control	Differential equations, optimization	Planning and control
Healthcare and Bioinformatics	Statistics, probability, optimization	Prediction and risk estimation
Finance	Statistics, stochastic models, optimization	Forecasting and decision support
Environmental Analysis	Statistics, spatial mathematics, optimization	Pattern detection and prediction

Ethics, Fairness and Responsible Statistical Modelling

The supplied paper notes that AI systems can propagate or amplify undesirable behaviours, including bias, discrimination and privacy violations. Quantitative evaluation can contribute to responsible AI by comparing performance across relevant subgroups, monitoring error distributions and reporting uncertainty. Mathematical regularization, calibration, threshold analysis and statistical validation can support more transparent model assessment, although technical measures alone cannot resolve every ethical question.

Discussion

The statistical evidence illustrates how mathematical concepts become operational in AI research. Dataset size determines the amount of empirical evidence available for training and evaluation. Class balance affects the interpretation of accuracy. Error rates convert model behaviour into measurable quantities, while optimization

determines how model parameters are learned. Statistical learning theory explains why performance on training data must be distinguished from generalization to unseen data.

The MNIST example also demonstrates the importance of preprocessing. The documented change from 12.0% to 8.4% test error corresponds to 360 fewer errors on a 10,000-image test set. This is not evidence that preprocessing will produce the same improvement on every dataset; rather, it provides a concrete illustration of how a mathematical transformation of data can affect empirical performance.

Limitations

- The quantitative component is based on secondary benchmark documentation rather than a newly trained experimental model.
- Reported benchmark results may depend on implementation, preprocessing, training protocol and historical experimental conditions.
- The paper does not claim that the documented MNIST results generalize directly to all AI applications.
- A future empirical study should train multiple models under controlled conditions and report confidence intervals, cross-validation results and computational cost.

Conclusion

Mathematics provides the conceptual and computational foundation of Artificial Intelligence and Machine Learning. Linear algebra supports representation, probability and statistics quantify uncertainty and empirical performance, optimization drives parameter learning, and information theory provides measures of uncertainty and dependence. Statistical learning theory connects these components to generalization and model complexity.

The secondary-data analysis shows the value of presenting AI research with explicit statistical evidence. MNIST provides 60,000 training and 10,000 test examples, while Fashion-MNIST provides the same split across 10 balanced classes. The documented MNIST linear-classifier error rates of 12.0% and 8.4% translate to approximately 1,200 and 840 errors on the 10,000-example test set, respectively. Such conversions make model performance easier to interpret and demonstrate the practical relationship between mathematical formulation and statistical evaluation.

Future research can extend this framework through controlled experiments comparing classical statistical models, optimization algorithms and neural architectures using reproducible datasets, confidence intervals, hypothesis tests and computational-cost measures. In this way, mathematics and statistics can continue to provide a rigorous basis for transparent, reliable and responsible AI.

Acknowledgement

I would like to express my sincere gratitude to all those who supported and encouraged me in the completion of this research paper, "Mathematical Foundations and Statistical Applications of Artificial Intelligence and Machine Learning."

Financial support and sponsorship

Nil.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of thesis paper.

References

1. Chodrow, P. (2017). Divergence, Entropy, Information: An Opinionated Introduction to Information Theory.
2. Glauner, P., Meira, J. A., & State, R. (2018). Machine Learning for Data-Driven Smart Grid Applications.
3. Hafez-Kolahi, H., & Kasaei, S. (2019). Information Bottleneck and its Applications in Deep Learning.
4. Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine Learning and Deep Learning.
5. Kutyniok, G. (2022). The Mathematics of Artificial Intelligence.
6. LeCun, Y., Cortes, C., & Burges, C. J. C. MNIST Handwritten Digit Database. Official dataset documentation.
7. Morse, L., Teodorescu, H. M., Awwad, Y., & Kane, G. (2020). A Framework for Fairer Machine Learning in Organizations.
8. Sen, J., et al. (2022). Machine Learning: Algorithms, Models, and Applications.
9. Sun, S., Cao, Z., Zhu, H., & Zhao, J. (2019). A Survey of Optimization Methods from a Machine Learning Perspective.
10. von Luxburg, U., & Schölkopf, B. (2008). Statistical Learning Theory: Models, Concepts, and Results.
11. Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms.