Knowledge Retention and Mathematical Foundations in Machine Learning Education

Exploring the Role of Prior Mathematical Knowledge in Retaining Core Machine Learning Concepts

Author: **Carina-Silvia Oprean**

Email: C.S.Oprean@student.tudelft.nl

Responsible Professor: Gosia Migut; Supervisor: Ilinca Rențea

CSE3000 Research Project

1. Introduction

Machine Learning (ML) drives innovation in fields like healthcare, AI, and engineering. As an interdisciplinary subject blending computer science, mathematics, and statistics, ML education is essential for preparing students to meet the demands of a rapidly evolving workforce [1].

- **Mathematics: The Foundation of ML**

Mathematical concepts - such as Linear Algebra, Calculus and Probability - are fundamental to understanding ML algorithms and methods [2, 3].

- **Bridging the Gaps**

Students struggle to retain ML concepts long-term, often due to gaps in foundational math knowledge [4]. This study aims to address these challenges by investigating:

  - How prior knowledge of specific math topics influences retention of core ML concepts
  - How perceived difficulty and confidence shape learning outcomes

2. Research Question

To what extent do students who have completed the Machine Learning course (CSE2510) retain core concepts within two years, and how does prior knowledge of specific mathematical topics influence this retention?

3. Methodology

Literature Review →

- Inspired item creation by reviewing established educational measurement tools [5 - 7]

Survey Design →

- **Purpose:** assess retention of ML concepts and influence of prior math knowledge, perceived difficulty, and confidence
- **Structure:**
  - Opening statement & Demographics
  - One section per ML topic: 5 MCQs, 5 Likert items, 1 open-ended question
- **Balanced design:** randomized topic order

Survey Distribution →

- **Target group:** students who passed the ML course within the past two years
- **Final sample size**: N = 28
- **Ethical compliance:** anonymity and voluntary participation ensured
- **Instructions:** complete independently, without external resources

4. Results

Core ML Topics: Principal Component Analysis (PCA), Gradient Descent (GD), Bayes' Theorem (BT), Hierarchical Clustering (HC)

4.1 Correlation Between Prior Math Knowledge and ML Retention

- Examined how grades in foundational math courses relate to retention of core math-intensive ML topics
- Stronger math foundations, particularly in Calculus are linked to better retention of related ML concepts like GD (Figure 1)

>>> Highlighting math's role in ML can improve retention outcomes

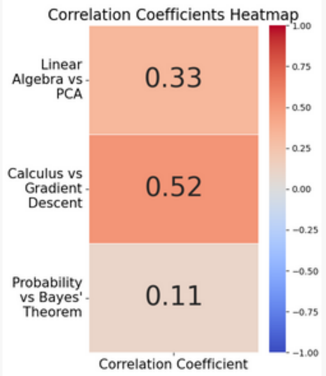

Figure 1: Targeted Correlations Heatmap between Grades in Math Courses and Retention Scores Across ML Topics

4.2 Retention, Perceived Difficulty, and Confidence

- Difficulty Trends: topics perceived as difficult showed mixed effects on retention. For GD, higher difficulty hindered retention, while for PCA and BT, difficulty appeared to motivate engagement
- Confidence Boosting Retention: higher confidence consistently correlated with better retention across all topics

>>> Building confidence and addressing difficulty through tailored strategies can enhance retention

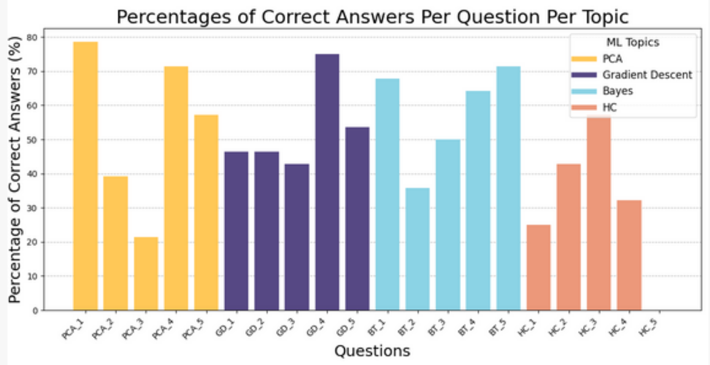

Figure 2: Percentages of Correct Answers for Each Question Across ML Topics

4.3 Comparative Retention Across ML Topics

- HC had the lowest retention, despite minimal math prerequisites (Figure 2)
- PCA, GD, and BT had higher retention, but average scores remain unsatisfactory

>>> Frequent application and reinforcement of concepts improve retention, but must be coupled with stronger foundational understanding for truly satisfactory outcomes

4.4 Performance on Math-Linked Questions

- Question 3 had the highest "Don't Remember" response rates for the math-intensive ML topics, compared to HC where this trend was reversed

>>> Gaps in applying mathematical principles can hinder retention, reinforcing the need for targeted math refreshers (Figure 3)

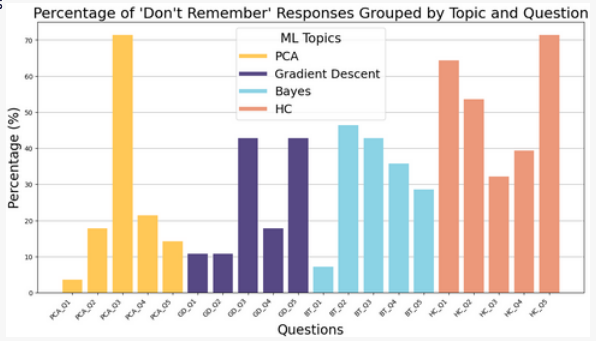

Figure 3: "Don't Remember" Response Percentages Across ML Topics

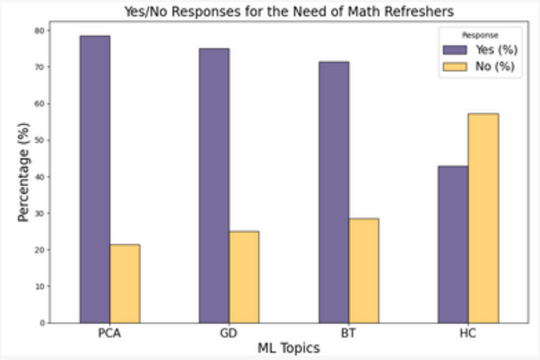

Figure 4: Proportion of Yes/No Responses for the Need of a Math Refresher Across ML Topics

4.5 Thematic Analysis from Open-Ended Responses

- Challenges included recalling eigenvectors and eigenvalues (PCA), derivatives (GD), and probability terms (BT)
- Many noted that gaps between math courses and ML topics accelerated forgetting

>>> Students emphasized the need for targeted refreshers and practical approaches to strengthen understanding, suggesting a curriculum improvement opportunity (Figure 4)

5. Discussion

Students struggled with applying mathematical principles to ML concepts, with high "Don't Remember" rates for questions requiring advanced math, while confidence positively correlated with retention [8], emphasizing the need for strategies to boost student confidence [9].

Proposed Solution: Targeted Math Refreshers

- Concise, topic-specific refreshers reinforcing key math concepts before related ML lectures
- Format: Jupyter Notebooks, recap slides, or short video tutorials including:
  - Concept explanations (e.g., eigenvalues for PCA, derivatives for GD)
  - Worked examples and practice problems with feedback
- **Aim:** strengthen math foundations to improve understanding and retention of ML concepts

Limitations:

- Sample Size: relatively small, N = 28, limiting generalizability
- Survey Design: limited question depth and only four ML concepts assessed
- Math Proxies: grades used to measure prior knowledge may lack granularity
- Timeline: no longitudinal data to assess and compare retention decay over time

Future Work:

- Develop and test interactive modules for revisiting math concepts
- Expand analysis to additional ML topics and larger sample sizes
- Conduct longitudinal studies to track retention at multiple intervals
- Explore alternative refresher formats, such as quizzes or collaborative workshops

6. Conclusions

- Retention of ML concepts is shaped by mathematical foundations, with Calculus having the strongest influence on Gradient Descent retention
- Confidence correlates positively with retention, while perceived difficulty shows mixed effects across topics
- **Math refreshers matter:** concise, targeted refreshers linked to ML topics can bridge foundational gaps and improve long-term retention
- **Curricular insights:** reinforcing math-intensive topics and fostering confidence are crucial for sustained learning in ML education

References

[1] Grabarnik, G., Kim-Tyan, L., & Yaskolko, S. (2020). Addressing prerequisites for STEM classes using an example of linear algebra for a course in machine learning. Mobile, Hybrid, and On-line Learning (eLmL 2020), 29.

[2] Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for machine learning. Cambridge University Press.

[3] Alnaimi, I. A., Alenezi, et al. The Role of Academic Self-Efficacy in Predicting Mathematical Confidence among University Students in Light of Some Demographic and Academic Variables.

[4] Ralston, A., & Shaw, M. (1980). Curriculum'78—is computer science really that unmathematical?. Communications of the ACM, 23(2), 67-70.

[5] Nuutila, K., Tapola, A., Tuominen, H., Molnár, G., & Niemivirta, M. (2021). Mutual relationships between the levels of and changes in interest, self-efficacy, and perceived difficulty during task engagement. Learning and Individual Differences, 92, 102090.

[6] Ribeiro, N. F., & Yarnal, C. M. (2010). The Perceived Difficulty Assessment Questionnaire (PDAQ): Methodology and applications for leisure educators and practitioners. Schole: A Journal of Leisure Studies and Recreation Education, 25(1), 111-115.

[7] Loorbach, N., Peters, O., Karreman, J., & Steehouder, M. (2015). Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. British journal of educational technology, 46(1), 204-218

[8] Bandura, A. (1977). Self-Efficacy: Toward a Unifying Theory of Behavioral Change. Psychological Review, 84.

[9] McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger III, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. Journal of Experimental Psychology: Applied, 20(1), 3.