

We need to learn how to Teach Machine Learning

Advantages of Prior Mathematical Knowledge for Studying Machine Learning



By Oisín Hageman CSE3000 Research Project
Delft University of Technology

Abstract

With the growing need for machine learning knowledge for many fields, comes the challenge of teaching machine learning concepts to non-majors with diverse mathematical knowledge. This paper evaluated the correlation between the prior knowledge and topic specific machine learning topics via an experiment conducted on Computer Science and Physics students without prior machine learning education. We find no correlation between math knowledge and knowledge gain. However, a clear correlation of proficiency in probability and statistics, and algorithm heavy machine learning topics is observed, which students found particularly challenging, along with abstract concepts like perceptrons.

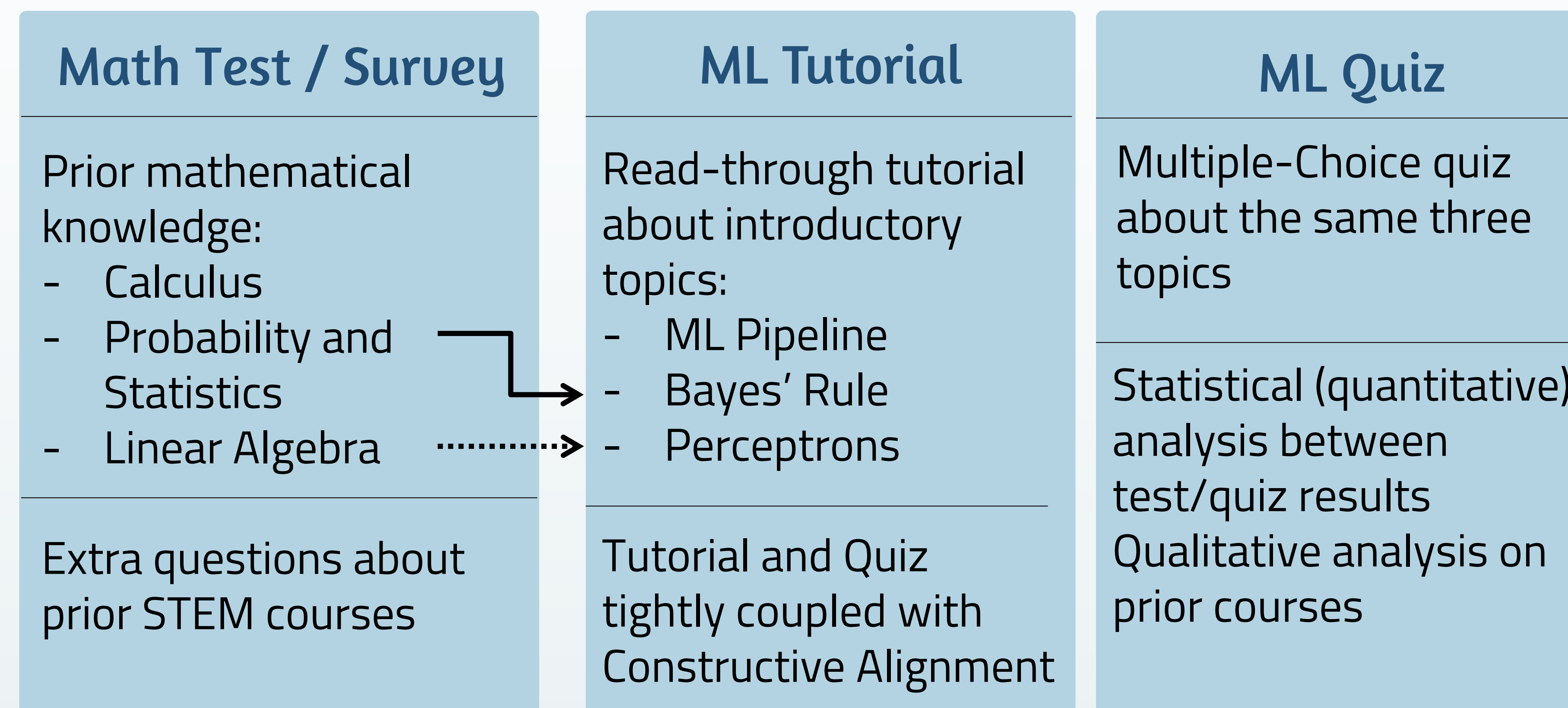
1. Introduction / Background

- Growing need for machine learning (ML) education for non-computer scientists [1]
- Non-majors have different background in mathematics -> more difficult to teach machine learning
- Correlation between prior mathematical proficiency and ML knowledge gain not quantified in literature

Research Question:

“How does prior mathematical knowledge influence students to learn specific machine learning topics between Computer Science and Physics majors?”

2. Methodology



Target Groups

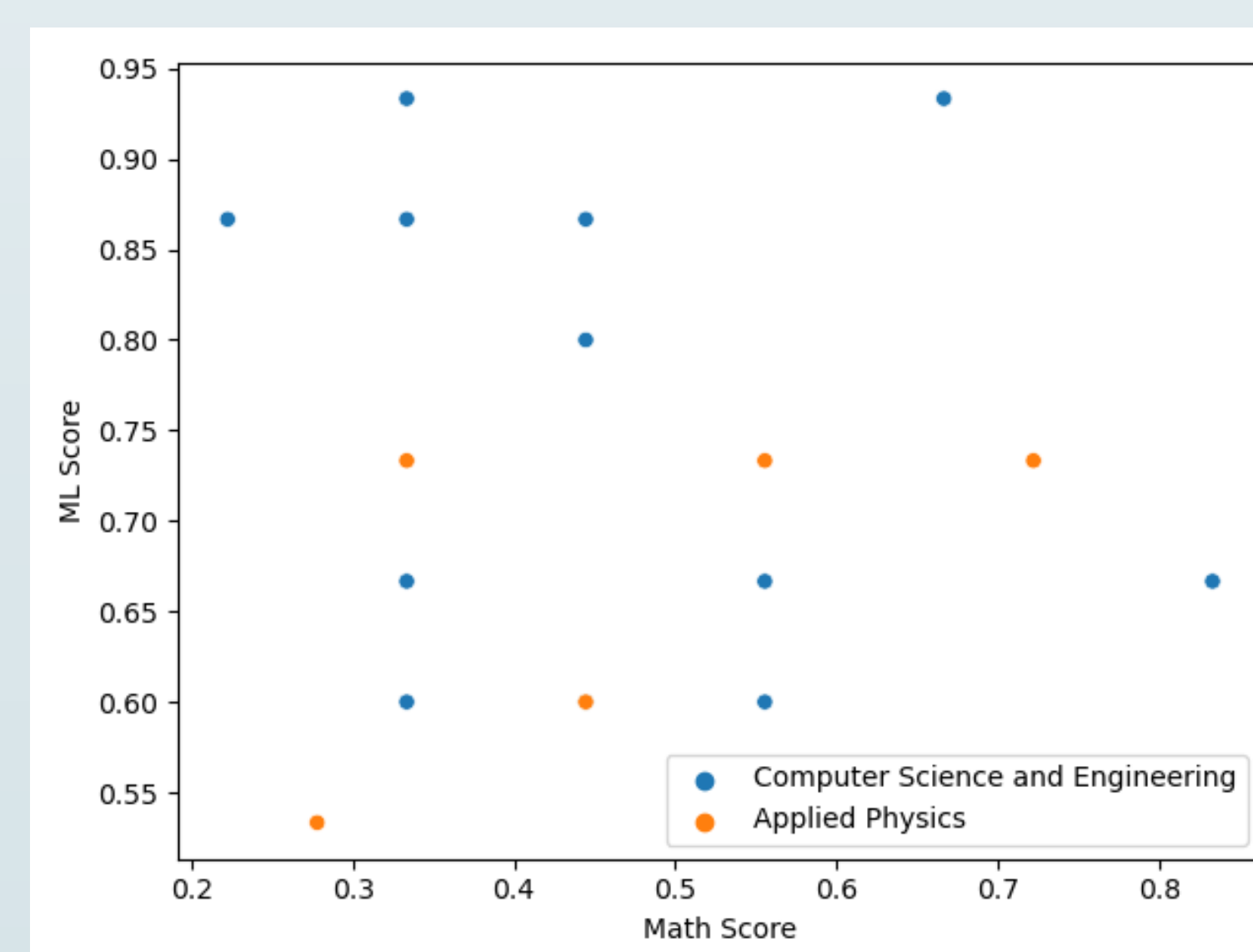
BSc Applied Physics students

N=5

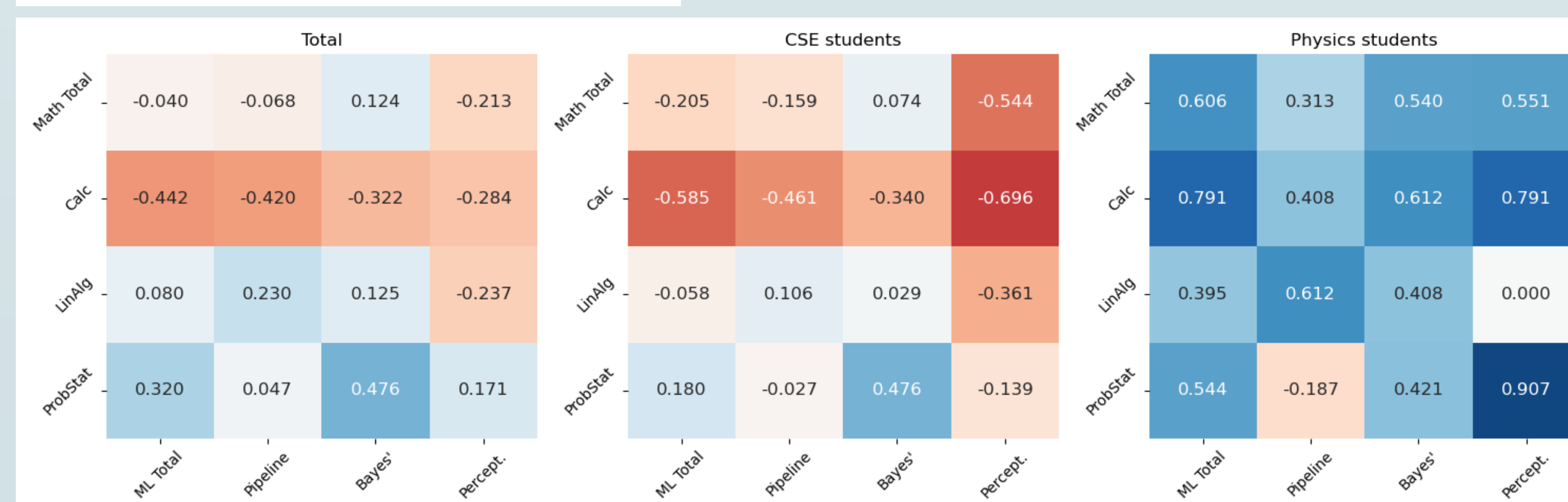
N=12

BSc Computer Science students

3. Results



	μ Math score	μ ML score	r	ρ
CSE	0.458	0.756	-0.205	-0.167
AP	0.467	0.667	0.606	0.671



4. Discussion

- AP students slightly better than CSE on math scores
 - More affinity with mathematics
- ML scores significantly higher for CSE
 - ML similar to rest of curriculum
- Correlation between Bayes' rule and ProbStat scores is moderately strong for both groups
- Negative correlations for Calculus scores to ML knowledge gain hard to justify
 - Need more students to get conclusive results
- Perceived difficulties for Bayes' line up with highest topic correlation found:
 - All physics students state highest difficulty with Bayes' rule, because of difficulty/dislike with probability theory
- Additional math education is required when teaching machine learning to students with weaker math backgrounds

	Commonly stated difficulties	Perceptrons	Bayes' Rule	Probability
CSE		67%	33%	8%
AP		0%	100%	100%

5. Future Work

- Similar experiment with changes:
 - Larger sample sizes
 - CSE students with more differences in completed math courses
 - More in-depth math test, with broader concept of math *proficiency*, and wider range of questions
- Combining dataset to compare different non-majors

References

- [1] R. B. Shapiro and R. Fiebrink, "Introduction to the special section: Launching an agenda for research on learning machine learning," *ACM Trans. Comput. Educ.*, vol. 19, no. 4, Oct. 2019.
- [2] E. Sulmont, E. Patissas, and J. R. Cooperstock, "What is hard about teaching machine learning to non-majors? insights from classifying instructors' learning goals," *ACM Trans. Comput. Educ.*, vol. 19, no. 4, Jul. 2019.