Does Representation Matter? Comparing Algebraic and Geometric Approaches to Teaching L1/L2 Regularization.

Effects on conceptual understanding, problem-solving, and knowledge transfer

TU Delft
Author: Ivan Nikolov
Email: I.I.Nikolov@student.tudelft.nl
Supervisors: Gosia Migut, Ilinca Rențea

1 Background

- Machine learning is now a core skill, so students need better ways to learn difficult ML concepts.
- **Regularization** is important because it prevents **overfitting** by penalizing large model weights.
- *L1* and *L2* behave differently: *L1 creates sparse models*, while *L2 shrinks weights smoothly*.
- Regularization is usually taught **algebraically** or **geometrically**, but it is unclear which approach supports learning better.
- Prior work shows students often know ML terms but struggle to explain the mechanisms behind them.
- This study addresses that gap by comparing how algebraic and geometric explanations affect **conceptual understanding, problem-solving, and transfer**.

2 Research Question

Does teaching L1 and L2 regularization algebraically or geometrically lead to differences in student performance across conceptual understanding, problem-solving, and knowledge transfer?

3 Materials

Algebraic representation notebook

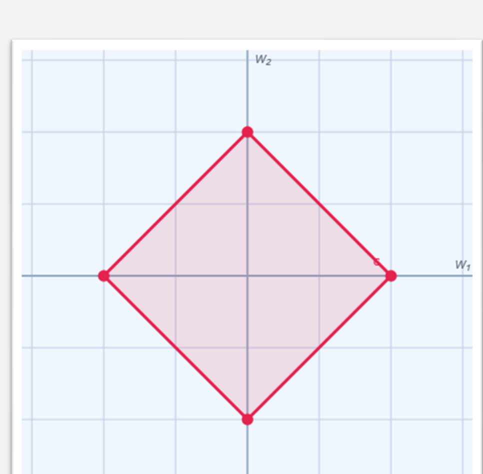
- Frames regularization as changing **what the model minimises** – the loss function plus a penalty term - $L(w) + \lambda \cdot R(w)$.
- L1 vs L2 is the **form of the penalty**: $\Sigma|w_j|$ vs Σw_j^2.

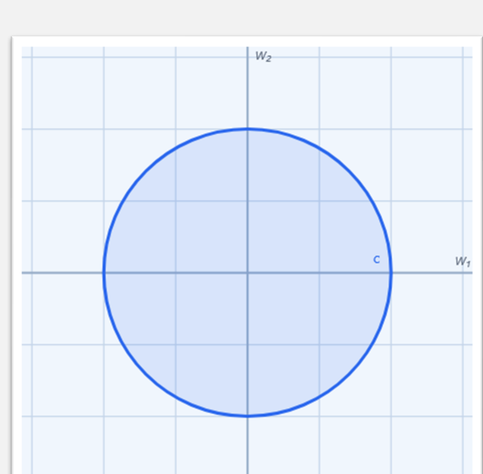
$$\frac{1}{n}\sum_{i=1}^{n}(y_i - b - w_1x_{1i} - \cdots - w_5x_{5i})^2 + \lambda\sum_{j=1}^{5}|w_j|$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - b - w_1x_{1i} - \cdots - w_5x_{5i})^2 + \lambda\sum_{j=1}^{5}w_j^2$$

Geometric representation notebook

- Frames the identical problem as restricting **where the weights can live** - a constraint region in weight space.
- L1 vs L2 is the **shape of the region**: a diamond with corners on the axes vs a smooth circle.





4 Methodology

Participants & Assignment

- 20 Bachelor students from Computer Science and Aerospace Engineering participated.
- All had completed an introductory Machine Learning course.
- Students were split into two matched groups: Algebraic notebook (n = 10) and Geometric notebook (n = 10), balanced by field of study and pre-test score.

Study Procedure

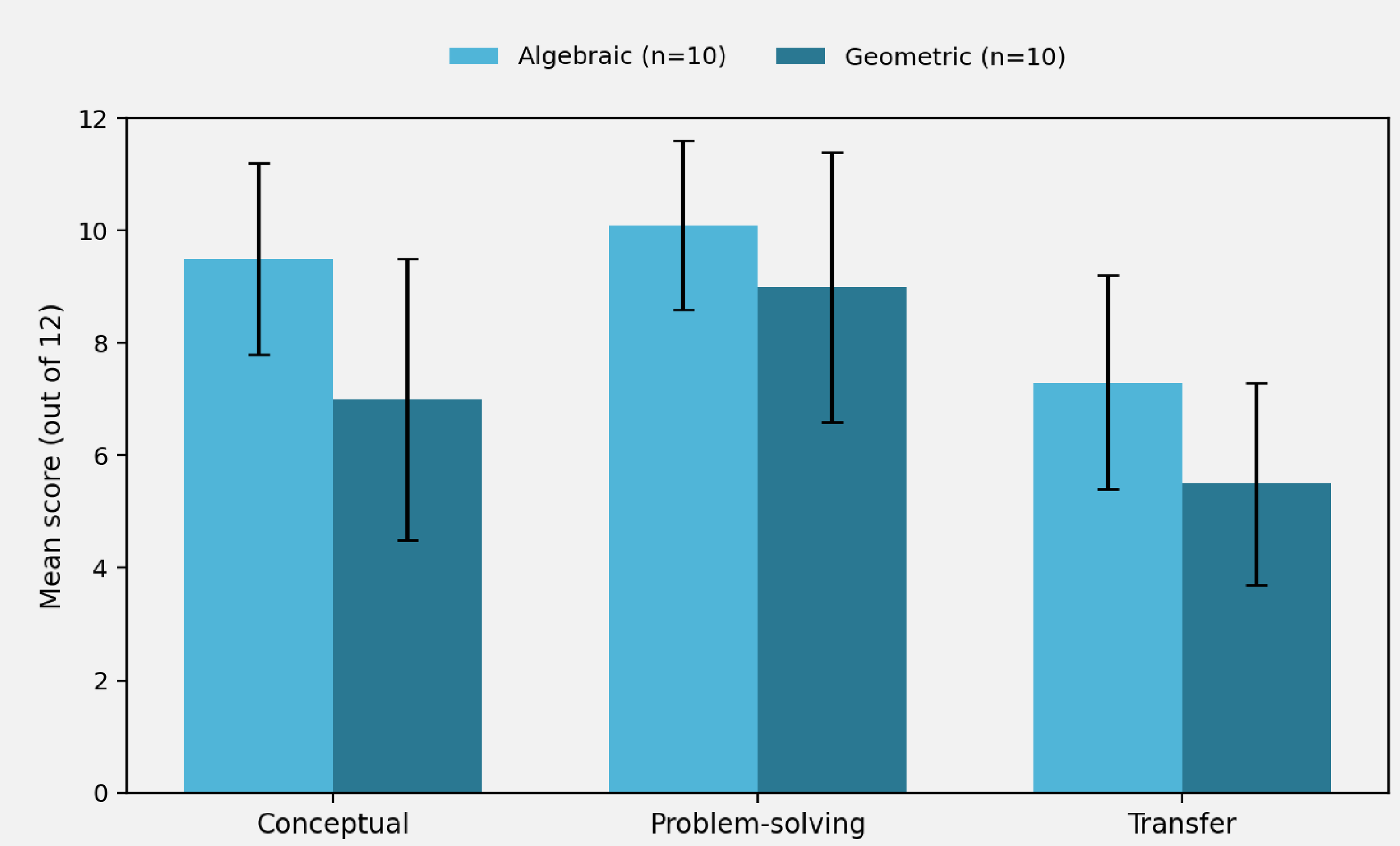
- Procedure: pre-test → assigned interactive notebook → post-test.
- 8 participants also completed a structured interview about their reasoning and learning experience.

Measurements & Analysis

- The post-test measured conceptual understanding, problem-solving, and knowledge transfer.
- A mixed ANOVA compared performance across group and question category.
- Bonferroni-corrected follow-up tests identified group differences; thematic analysis examined written explanations.

5 Quantitative Results

- Both groups started equal: 4.70/5 on the pre-test.
- Algebraic group scored higher overall: 26.9 vs 21.5 out of 36.
- Overall difference showed a large effect size: $d \approx 1.08$.
- Conceptual understanding showed the clearest algebraic advantage.
- Conceptual difference was the only category result that survived correction: $d \approx 1.18$.
- Problem-solving showed no reliable difference between groups.
- Transfer was higher for the algebraic group on average, but inconclusive after correction.



6 Qualitative Analysis

- Students used different explanatory language: algebraic students framed regularization as a penalty on the loss function, while geometric students framed it as a constraint region in weight space.
- Both groups reached **similar practical judgments** about overfitting, underfitting, λ, and validation-based tuning.
- *Algebraic students* more often justified their answers by checking validation error or testing several λ values.
- *Geometric students* sometimes gave correct outcomes but explained less of the mechanism behind them.

7 Implications

- Representation shaped **how students explained** regularization more than whether they understood it.
- Algebraic explanations support formal ML vocabulary: penalties, weights, λ, training error, and test error.
- Geometric explanations support intuition for why L1 creates sparse weights and L2 shrinks weights smoothly.
- Teaching should combine both representations instead of choosing only one.

8 Conclusion

- Both formats helped students reason about the **core ideas** of regularization.
- The **algebraic group scored higher overall**, especially on conceptual understanding.
- Problem-solving performance was **similar** across both groups.
- The main difference was not practical reasoning, but **how students expressed their understanding**.

9 Limitations and future work

- **Small sample size**: 20 participants.
- Participants already had introductory ML background, so results may not generalize to beginners.
- Only **immediate post-test performance** was measured, not long-term retention.
- Future work should use **larger samples, delayed testing**, and directly test whether combining both formats improves learning.