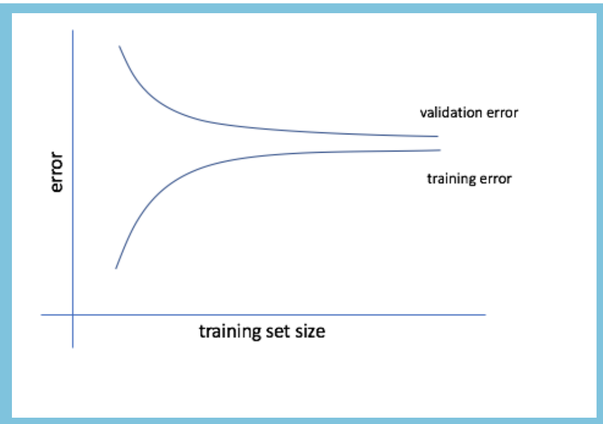
WHAT IS THE EFFECT OF GAUSSIAN FILTERING APPLIED BEFORE CURVE FITTING?

Author: Liviu Moanta (I.L.Moanta-1@student.tudelft.nl) Responsible Professor: Tom Viering (T.J.Viering@tudelft.nl)
Supervisors: Taylan Turan (O.T.Turan@tudelft.nl) and Cheng Yan (C.Yan-1@tudelft.nl)

TU Delft Delft University of Technology

1. INTRODUCTION

- Learning curves are graphical representations of the relationship between dataset size and error rate in machine learning
- Curve fitting is the process of estimating a learning curve using a mathematical formula. This research analyzes two ways of evaluating curve fitting: interpolation and extrapolation.
- Our study investigates the effects of the Gaussian filter on curve fitting and the potential to improve its performance.



2. RESEARCH QUESTION

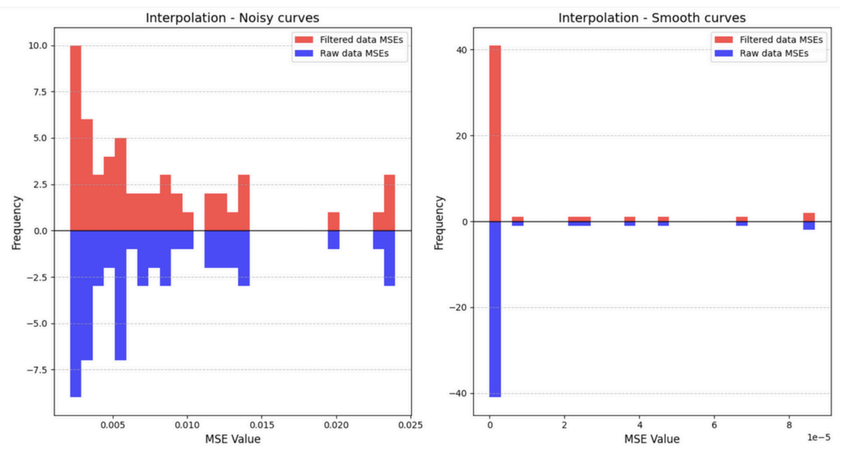
What is the impact of the Gaussian filter applied to the initial set of points before curve fitting for both interpolation and extrapolation?

3. METHODOLOGY

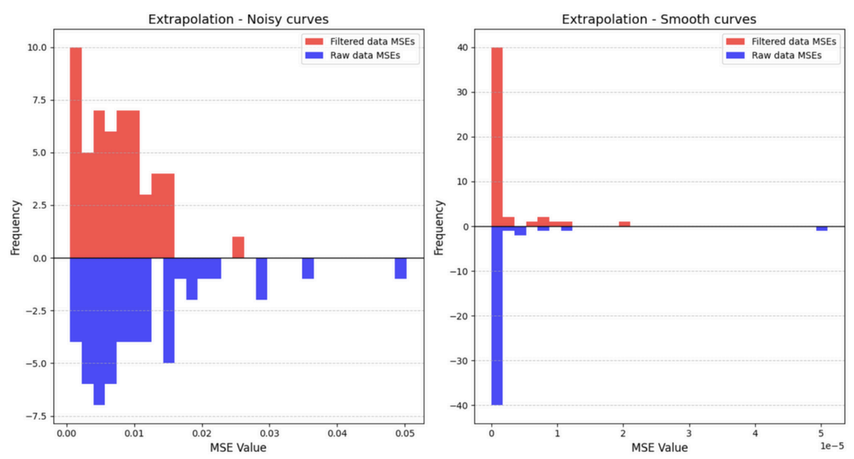
1. Create and design a framework that will allow me to import, analyze, and visualize learning curves
2. Apply the Gaussian filter on the raw points of the learning curve
3. Implement the curve fitting for both the initial learning curve and the filtered one
4. Evaluate the performance of the fitting using the raw curve vs the fitting using the filtered curve by calculating the Mean Squared Errors of interpolation and extrapolation
5. Conduct Mann-Whitney U tests for each hypotheses in order to determine if applying the Gaussian filter is statistically significant in each case

4. RESULTS

- Interpolation using noisy curves:
  Mann Whitney U test P-Value = 0.90
  The difference is not statistically significant.
- Interpolation using smooth curves:
  Mann Whitney U test P-Value = 0.97
  The difference is not statistically significant.

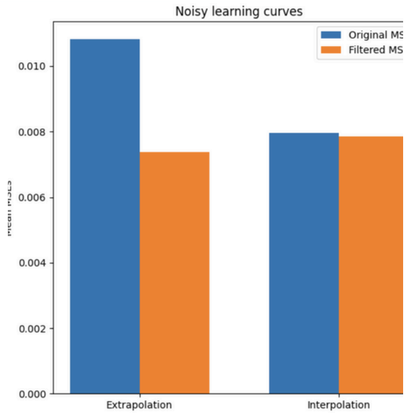


- Extrapolation using noisy curves:
  Mann Whitney U test P-Value = 0.02
  The difference is statistically significant.
- Extrapolation using smooth curves:
  Mann Whitney U test P-Value = 0.63
  The difference is not statistically significant.

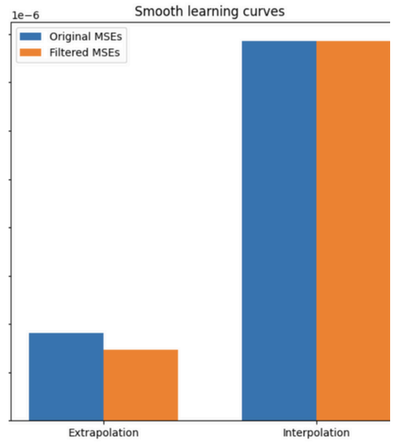


5. DISCUSSION

- The Gaussian filter improves the performance of fitting noisy curves, with a particularly significant improvement in extrapolation.



- The Gaussian filter provides a slight improvement in the performance of fitting smooth curves, with little to no impact on interpolation.



6. CONCLUSIONS

- Gaussian filter advantages
  - Reduces noise
  - Attenuates fluctuations
  - Improves generalization
- Gaussian filter improvement
  - Significant for extrapolating noisy curves
  - Insignificant for the other cases
- Best Sigma Values
  - For interpolation: Sigma = 2.0
  - For extrapolation: Sigma = 16.0

7. FUTURE WORK

- Research on multiple smoothing techniques(Low Pass Filter or Savitzky-Golay filter)
- Research on multiple Sigma Values, fitting percentages, models, ways of calculating Mean Squared Error
- Research on other types of learning curves