PREVALENCE OF NON-MONOTONICITY IN LEARNING CURVES

Author: Dinu Gafton(D.Gafton@student.tudelft.nl) Supervisor: Taylan Turan (O.T.Turan@tudelft.nl)
Responsible Professor: Tom Viering (T.J.Viering@tudelft.nl)

TU Delft Delft University of Technology

1. INTRODUCTION

Learning curves (LC) depict how model performance evolves as the training set size increases.
They serve as a tool for estimating the training time and costs associated with machine learning models.

Terminology:

- Anchor: A point on the learning curve that defines the correlation between training size and model error.
- *Y-distance*: vertical separation or difference in the Y-coordinate values between two anchor points on a graph, along the Y-axis

2. RESEARCH QUESTION

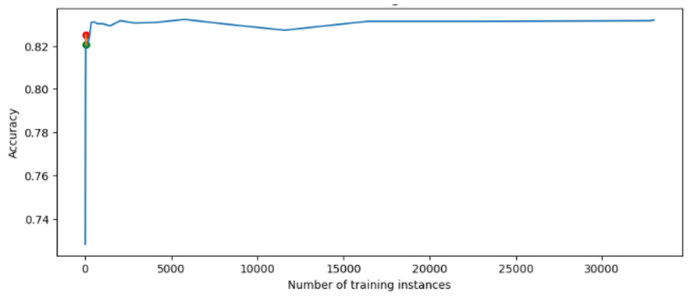
How many learning curves are nonmonotone and what influences this?

3. METHODOLOGY

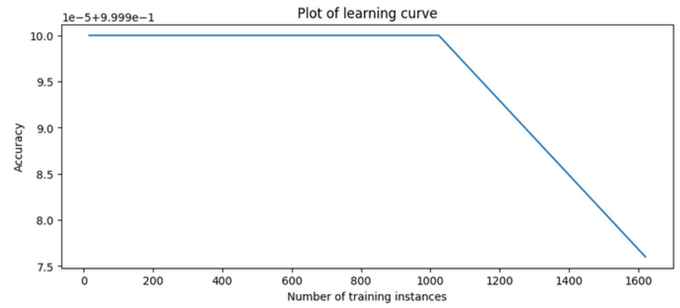
1. Create an algorithm that can identify non-monotonicity in learning curves, by observing the degree of monotonicity violation in the anchor points.
2. **Evaluate the algorithm :**
   - Accuracy testing using artificial LCs;
   - Significance Test: how significant is the non-monotonicity in each learning curve;
3. Compare performance to another algorithm
4. Using this metric, evaluate non-monotonicity on a subset of the Learning Curve Database (LCDB) [1]

5. LIMITATIONS

- Sensitive against noise data, the threshold is still set too low



- Fails to identify peaking on almost constant learning curves with sudden drops in performance



6. CONCLUSIONS

- Y-distances can be used to judge monotonicity of learning curves, the proposed algorithm providing high accuracy
- O(N) complexity (N - nr. of anchor points), fast and reliable enough to be able to conduct large scale analysis of many learning curves
- Results indicate that there might be little to no correlation between occurrences and significances

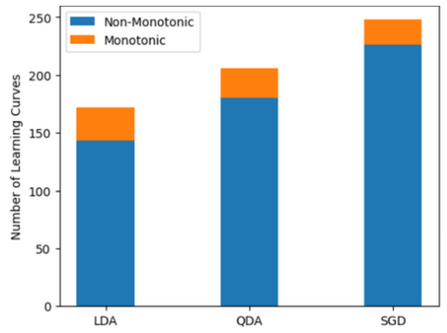
Future work:

- Run the proposed algorithm over the entire LCDB database
- Optimize the existing threshold or try using standard deviation as threshold

[1] Felix Mohr et al. "LCDB 1.0: An extensive learning curves database for classification tasks". In: Machine Learning and Knowledge Discovery in Databases (2023), pp. 3–19. doi: 10.1007/978-3-031-26419-1_1.

4. RESULTS

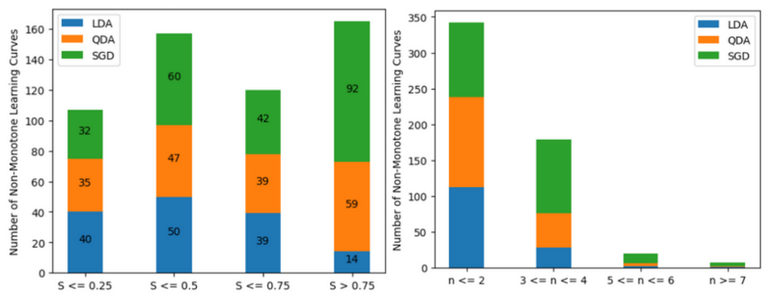
- Algorithm correctly identified most non-monotone learning curves
- Stochastic Gradient Descent (SGD) learner has the biggest ratio of non-monotone learning curves
- **83** non-monotone learning curves (**15.63%** out of the total) have a drop in performance right at the very end.



	Actual non-monotonic	Actual monotonic
Classified non-monotonic	526 (99.05%)	23 (24.21%)
Classified monotonic	5 (0.95%)	72 (75.79%)

Table 1: Accuracy Test Results. The brackets describe the percentage of correctly classified curves from the total number of monotonic or non-monotonic learning curves, respectively.

- Majority of non-monotone learning curves encounter **at most** 2 non-monotone intervals on the curve.
- The Stochastic Gradient Descent (SGD) learner has the most **volatile** learning curves (displays the behavior of peaking the most)



- Neural network models and SGD display the most non-monotone behavior, also being prone to display peaking behavior
- Tree-like learners show significantly less non-monotonic behaviour than the rest