Non-Monotonicity in Empirical Learning Curves

Identifying non-monotonicity through slope approximations on discrete points

Author: Codrin Socol *(C.Socol@student.tudelft.nl)*

Supervisor: Dr. Tom Viering *(T.J.Viering@tudelft.nl)*
Responsible Professor: Dr. Jesse Krijthe *(J.H.Krijthe@tudelft.nl)*

TU Delft Delft University of Technology

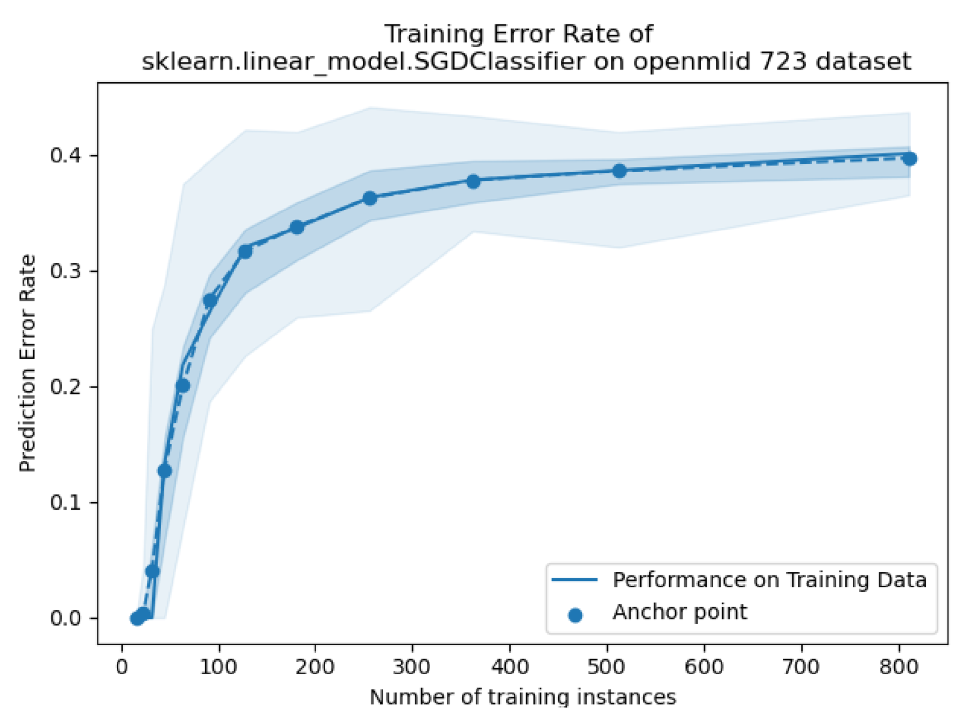
1. Introduction

- **Learning curves (LC)** outline the evolution of model performance with respect to increasing the training set size.
- They can be used to extrapolate training time & costs of machine learning models.

Terminology:

- **Curve monotonicity** – Adding more training samples will reduce the classifier error *(Descending Curve)*
- **Anchor** – point on the learning curve that describes the relationship between training size and model error. [1]

Intuition indicates that learning is monotone. However, that may not always be the case:

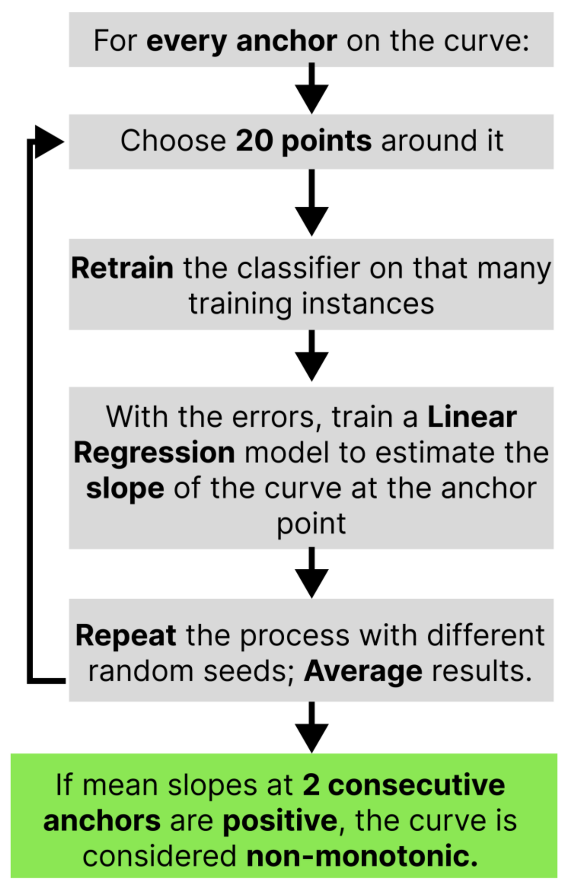


2. Research Question

How many learning curves are non-monotone and what influences this?

3. Methodology

1. Create an algorithm that can identify non-monotonicity in learning curves, by observing the **slopes** in the anchor points.
2. Evaluate the algorithm :
   - **Accuracy testing** using artificial LCs;
   - **Ablation study** to test whether 2 consecutive anchors should be considered to judge monotonicity;
   - **Peaking Test:** how well does the metric handle *sample-wise double descent [2];*
3. Using this metric, evaluate non-monotonicity on a subset of the **Learning Curve Database (LCDB)** [1].



5. Conclusions

1. It is useful to consider monotonicity analysis from a certain anchor point onwards;
2. Anchor slope approximation can be used to judge monotonicity of learning curves;
3. Some learners such as *LDA*, *QDA* experience more non-monotonic behavior compared to others, such as *Extra Trees* or *Random Forest.*

Future Work:

- Algorithm optimization through LR alternative, parallelization and/or GPU computing;
- Evaluating the entire LCDB database.

6. Limitations

1. The introduced metric cannot identify *peaking* if it only occurs around **one anchor** point;
2. Curves that are almost constant may be **misclassified** due to approximation errors from training Linear Regression;
3. Algorithm runs slowly, unfeasible for large scale analysis of many learning curves.

4. Results

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

	Actual Non-monotonic	Actual Monotonic
Predicted Non-monotonic	3140 (***98.77%***)	36 (*6.10%*)
Predicted Monotonic	39 (*1.23%*)	554 (***93.90%***)

- Algorithm correctly identified **most** artificial LCs generated;
- We concluded that the metric has **good potential** of identifying *real* non-monotonic curves.

Table 2: Ablation Study results. The brackets describe the increase/decrease compared to Experiment 1 Results from **Table 1**.

	Actual Non-monotonic	Actual Monotonic
Predicted Non-monotonic	3166 (***+0.82%***)	224
Predicted Monotonic	13	366 (***-31.87%***)

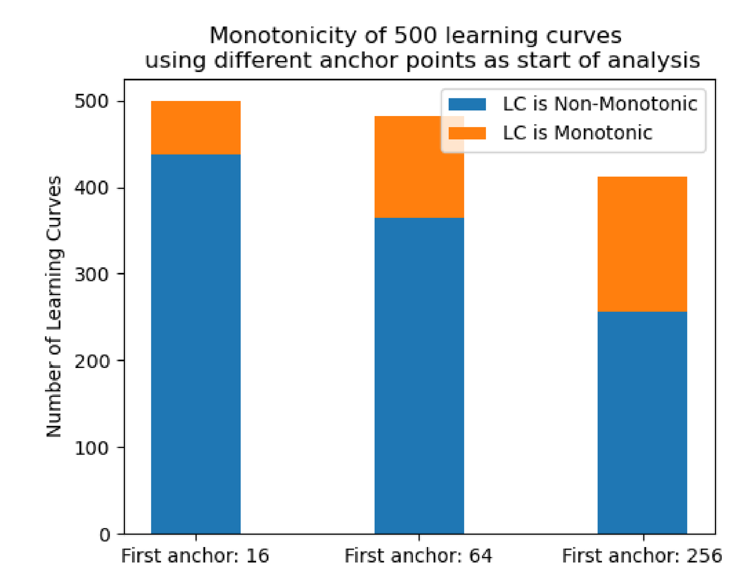
- Using 1 anchor slope instead of 2 results in a **worse performance** for identifying monotonicity.
- The ablated algorithm tends to classify curves as non-monotonic at the **slightest increase**.

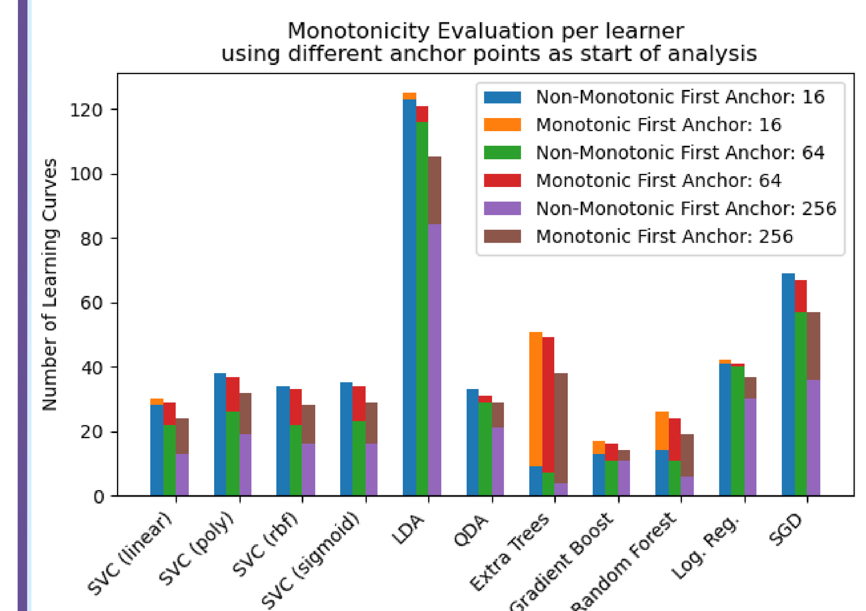
Table 3: Metric evaluation on 590 artificial monotonic LCs with peaking at anchor index 5.

	Correctly classified as non-monotonic	% of total curves (590)
Ablation Metric	267	**45.25%**
Initial Metric	19	**3.22%**

- The metric is **not suitable to identify peaking**, if it only occurs at one anchor point.

- Most classifiers tend to show **non-monotonic behavior** at the **very start** of the curve;
- For LCs with **smallest training times**, around **60%** showed non-monotonic behavior when starting to analyze from **anchor 256**.





For fastest trained datasets, some classifiers may be inherently non-monotonic:

- *Linear Discriminant Analysis* (LDA) and *Quadratic Discriminant Analysis* (QDA) showed **most non-monotonic behavior** at all 3 starting anchors.
- *Extra Trees* and *Random Forest* showed the **least non-monotonic behavior.**

[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.

[2] Tom Viering and Marco Loog. The shape of learning curves: a review, 2022.