In Search of Best Learning Curve Model

Dean Nguyen

Responsible Professor: Marco Loog
Supervisor: Tom Viering

1. INTRODUCTION

Research Question

Which parametric learning curve model provides the best fit when applied to empirical learning curves?

Why study learning curves? [1]

- Better model selection
- Extrapolation to reduce data collection costs
- Speeding up training and tuning

Current Literature

- Power law with some divergent results
- Few datasets and learners
- No study on regression tasks
- Faulty metrics (using R^2 for non-linear models)
- Some only looking at interpolation

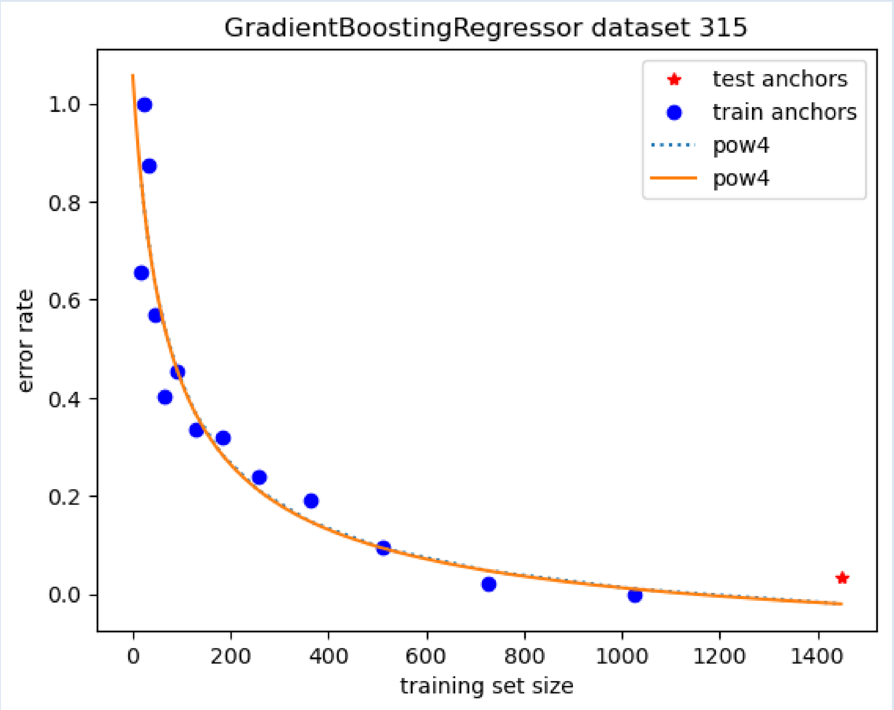

Figure 1: Example of a regression learning curve

2. METHODOLOGY & EXPERIMENTAL SETUP

Generating Empirical Learning Curves

- K-fold (train and test set)
- Anchor points given by $a_i := \lceil 2^{\frac{7+i}{2}} \rceil$
- 16, 23, 32,...
- Training data are reused in the testing data.

Curve Fitting Procedure

- Data is split into different partitions for fitting: 5%, 10%, 20%, 40% and 80%
- Levenberg-Marquardt
  - Random start for regression tasks
  - Specialized methods for classification tasks
- 16 parametric models fitted in total

Model	Formula	Model	Formula
last1	a	vap3	$\exp(a + \frac{b}{x} + c\log(x))$
pow2	$-ax^{-b}$	expp3	$c - \exp((-b + x)^a)$
log2	$-a\log(x) + b$	expd3	$c - (-a + c)\exp(-bx)$
exp2	$a\exp(-bx)$	logpow3	$a/((x\exp(-b))^c + 1)$
lin2	$ax + b$	pow4	$a - b(d + x)^{-c}$
ilog2	$-a/\log(x) + b$	mmf4	$(ab + cx^d)/(b + x^d)$
pow3	$a - bx^{-c}$	wbl4	$-b\exp(-ax^d) + c$
exp3	$a\exp(-bx) + c$	exp4	$c - \exp(-ax^d + b)$

Table 1: Parametric curve models

Setup

- Classification: 20 learners, 246 datasets (from the LCDB [2])
- Regression: 5 learners, 10 datasets
- Normalise the curves

3. RESULTS

- Friedman test ranks models per fitting experiment and takes the average to calculate the test statistics
- Wilcoxon signed-rank test with Holm's alpha correction to detect differences
- Visualized using Critical Diagrams

Average Rank

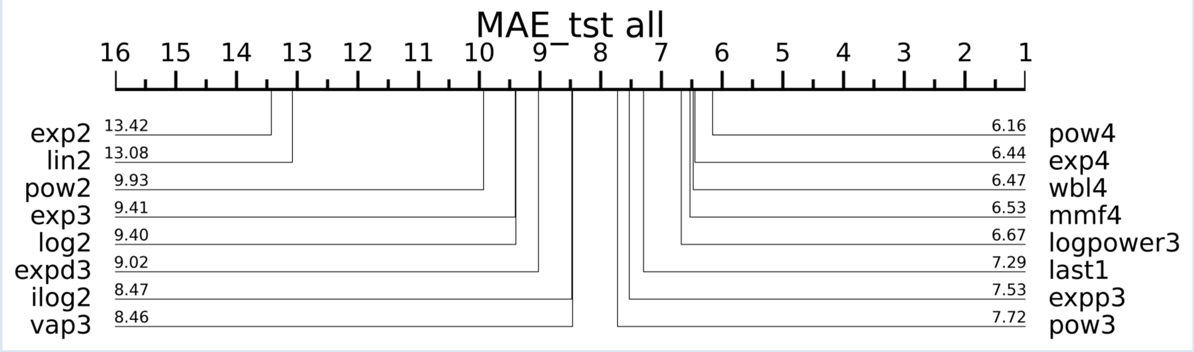

Figure 2: Critical diagram for classification tasks

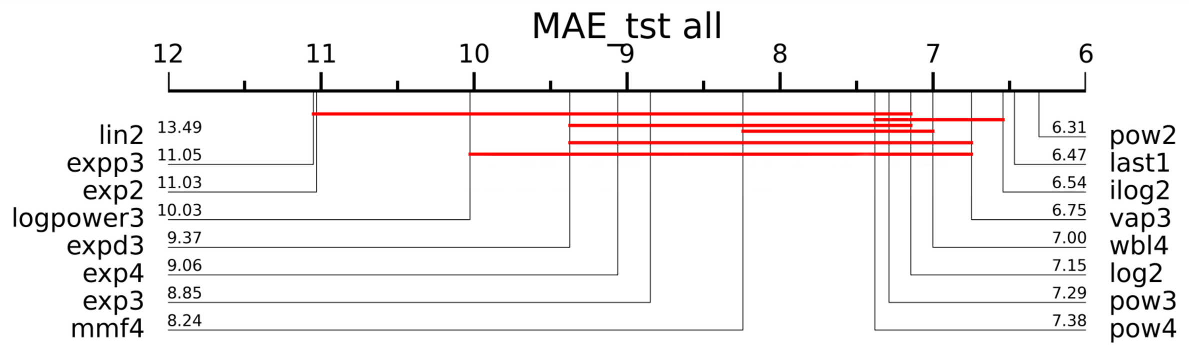

Figure 3: Critical diagram for regression tasks

Learner	Best Curve Model
LinearDiscriminantAnalysis	last1
QuadraticDiscriminantAnalysis	last1
BernoulliNB	last1
SVC_sigmoid	last1, wbl4

Table 2: Learners with divergent best curve models in average rank

Average MAE

Curve Model	Average MAE
last1	0.204 ± 0.0007
ilog2	0.244 ± 0.0008
mmf4	0.244 ± 0.0019
exp4	0.250 ± 0.0018
logpower3	0.256 ± 0.0024
wbl4	0.258 ± 0.0025
pow4	0.259 ± 0.0020
expp3	0.276 ± 0.0019
expd3	0.285 ± 0.0019
log2	0.292 ± 0.0012
pow3	0.305 ± 0.0021
pow2	0.311 ± 0.0012
exp3	0.312 ± 0.0020
vap3	0.323 ± 0.0020
lin2	0.572 ± 0.0022
exp2	0.606 ± 0.0026

Table 3: Average MAE of different parametric models for classification tasks

Curve Model	Mean MAE
last1	0.210 ± 0.0066
pow2	0.253 ± 0.0109
wbl4	0.270 ± 0.0119
ilog2	0.284 ± 0.0120
exp3	0.306 ± 0.0154
mmf4	0.313 ± 0.0148
expd3	0.315 ± 0.0144
logpower3	0.333 ± 0.0134
pow3	0.333 ± 0.0177
pow4	0.333 ± 0.0199
exp4	0.334 ± 0.0123
vap3	0.347 ± 0.0176
log2	0.375 ± 0.0134
expp3	0.474 ± 0.0220
exp2	0.484 ± 0.0200
lin2	1.062 ± 0.0376

Table 4: Average MAE of different parametric models for regression tasks

4. CONLUSIONS

- Power law but no universal model
  - pow4 is best for classification tasks
  - pow2 is best for regression tasks
  - Deviations for certain learners
- Different analyses show different curve models to be the best (outliers)

5. FUTURE RECOMMENDATIONS

- Investigate how preprocessing impacts the shape of learning curves
- Combine extensive hyperparameter tuning with preprocessing
- Further work into fitting the learning curves
- Perform an extensive study combining all recommendations using optimal settings

References

[1] Tom J. Viering and Marco Loog. The shape of learning curves: a review. CoRR, abs/2103.10948, 2021.
[2] F. Mohr, T. J. Viering, M. Loog, and J. N. van Rijn, "Lcdb 1.0: An extensive learning curves database for classification tasks," unpublished