To Tune or not to Tune: Influence of Hyperparameter Tuning on the Learning Curve

Author: Prajit Bhaskaran
Supervisor: Tom Viering
Responsible Professor: Marco Loog

1.Background

- Context: Learning curves display the measure of **accuracy** on **test data** against different amounts of **training data.**
- They can be modelled by parametric curve models

- Problem: No optimal parametric model[1].
- Little research on effects of hyperparameter tuning

2.Research Question

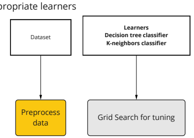
What is the influence of hyperparameter tuning on the learning curve?

- How does the general learning curve of an ML algorithm change when hyperparameter tuned?
- How does the curve fitting differ for hyperparameter tuned models against the default models?

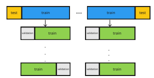
3.Method

General Learning curve

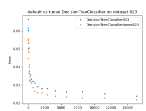
1.Preprocess data and setup appropriate learners



2.K stratified folds are created. Tuned learner undergoes k-cross validation. Default learner is trained on outer set



3.K learning curves are created which are then averaged to make a final general learning curve



Curve Fitting

model	Formula
POW2	$-an^{-b}$
POW3	$an^{-b}+c$
EXP2	$a\exp(-bn)$
EXPP3	$c-\exp((n-b)^{\alpha})$
log2	$c-\frac{a}{log(n)}$
logpow3	$a/((n\exp(-b))^{c}+1)$

Table 1: parametric learning curve models and their respective formulas where n is the varying training sizes

- Levenberg-Marquadt algorithm
- Random initial parameters
- Calculate MSE values (ranks)

4.Results

General Learning curve

Tuning displays smoothing behaviour

Error improvement follows parametric curve shape

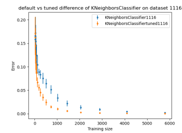


Figure 1: Default vs Tuned Kneighbors classifier on dataset 1116

Tuned learner performs better than default learner

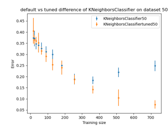


Figure 2: Default vs Tuned Kneighbors classifier on dataset 50

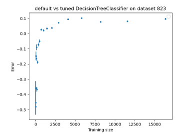
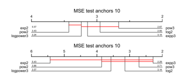


Figure 3: Error difference between default and tuned decision tree classifier

Curve Fitting

Ranks display how well a model performs (on average) on a dataset compared to other models



Critical diagram displays the significance of model performance

Decision tree tuned

curve	pow2	exp2	log2	pow3	expp3	logpow3
all	3.93	5.66	3.34	2.76	2.77	2.54
5%	3.33	5.83	3.00	3.33	2.67	2.83
10%	3.86	5.43	3.14	2.71	2.29	3.57
20%	3.60	5.93	3.20	3.80	1.93	2.53
40%	3.94	5.61	3.61	2.83	2.78	2.22
80%	4.42	5.32	3.37	2.68	2.95	2.26

Table 2:Ranks for extrapolating to all test anchors of tuned decision tree

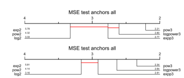
Figure 1: Critical diagram of decision tree default(top) and tuned(bottom) for MSE test anchor up to 10%. Models connected with a red line are statistically insignificant

Decision tree default

curve	pow2	exp2	log2	pow3	expp3	logpow3
all	4.09	5.59	3.11	2.54	2.91	2.77
5%	4.00	5.83	3.00	2.00	2.67	3.50
10%	4.17	5.83	3.25	2.58	2.17	3.00
20%	3.67	5.67	3.27	2.73	3.00	2.67
40%	4.42	5.58	3.26	2.58	2.53	2.63
80%	3.68	5.11	2.74	2.89	3.47	3.11

Table 3: Ranks for extrapolating to all test anchors of default decision tree

Default and tuned decision tree do not agree on optimal curve model



Default and tuned KNeighbors agree on pow3 being the optimal model

KNeighbors tuned

curve	pow2	exp2	log2	pow3	expp3	logpow3
all	4.14	5.81	3.16	2.34	2.91	2.64
5%	3.08	5.58	2.92	3.08	2.92	3.42
10%	3.21	5.64	2.86	2.64	3.57	3.07
20%	3.80	5.67	2.87	2.40	3.27	3.00
40%	4.21	5.42	3.21	2.42	3.16	2.58
80%	4.50	5.90	3.45	2.45	2.45	2.25

Table 4: Ranks for extrapolating to all test anchors of tuned KNeighbors

Figure 2: Critical diagram of Kneighbors default(top) and tuned(bottom) for MSE test anchor up to 10%. Models connected with a red line are statistically insignificant

KNeighbors default

curve	pow2	exp2	log2	pow3	expp3	logpow3
all	4.32	5.78	3.33	2.21	2.77	2.60
5%	3.89	6.00	3.56	2.00	2.89	2.67
10%	4.31	6.00	3.69	1.77	2.46	2.77
20%	4.47	6.00	3.53	1.93	2.33	2.73
40%	4.74	5.68	3.16	2.05	2.84	2.53
80%	4.60	5.55	3.30	2.60	2.20	2.75

Table 5: Ranks for extrapolating to all test anchors of default KNeighbors

5.Conclusion

-As training sizes increase, the tuned learner performs marginally better than default learner

-Tuned learner can make ill-behaved learning curves more well behaved

-Curve fitting for tuned learner is similar to that of the default learner

6.Dicussion and Future Work

-Chosen parametric models have been empirically analysed as strong models

-More accurate curve fitting tecnhiques

-Test on more datasets , learners, learning curve models

-Modellling the improvement in error between tuned learner and default learner

7.References

Viering, T., & Loog, M. (2021). The shape of learning curves: a review. arXiv preprint arXiv:2103.10948.