Fast and Informative Model Selection using Learning Curve Cross-Validation

Anonymous[1]

[1]Anonymous Institution

1 Broader Impact Statement

In this work, we present an approach to check options for early discarding in model selection based on an analysis of learning curves. We think that this work has potentially a broad impact in environmental and societal terms since it aims at reducing evaluation runtimes for machine learning models. The potential reduction of computation times implies not only a reduction of the CO_2 footprint but also potentially enables researchers or practitioners with less computational resources to employ AutoML tools that employ the presented technique rather efficiently. After careful reflection, the authors have determined that this work presents no notable negative impacts to society or the environment.

2 Submission Checklist

1. For all authors...

    (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]

    (b) Did you describe the limitations of your work? [Yes]

    (c) Did you discuss any potential negative societal impacts of your work? [Yes] We state above that we do not see any.

    (d) Have you read the ethics author's and review guidelines and ensured that your paper conforms to them? https://automl.cc/ethics-accessibility/ [Yes]

2. If you are including theoretical results...

    (a) Did you state the full set of assumptions of all theoretical results? [Yes] reviewers have checked our proof based on the assumptions we stated.

    (b) Did you include complete proofs of all theoretical results? [Yes]

3. If you ran experiments...

    (a) Did you include the code, data, and instructions needed to reproduce the main experimental results, including all requirements (e.g., `requirements.txt` with explicit version), an instructive `README` with installation, and execution commands (either in the supplemental material or as a URL)? [Yes] The code is available at https://github.com/fmohr/lccv, and the stuff is also pip installable.

    (b) Did you include the raw results of running the given instructions on the given code and data? [Yes] Can be found in the repository

    (c) Did you include scripts and commands that can be used to generate the figures and tables in your paper based on the raw results of the code, data, and instructions given? [Yes] There is a notebook in the repository to do that

(d) Did you ensure sufficient code quality such that your code can be safely executed and the code is properly documented? [Yes] But of course this depends on the commitment of the user

(e) Did you specify all the training details (e.g., data splits, pre-processing, search spaces, fixed hyperparameter settings, and how they were chosen)? [Yes]

(f) Did you ensure that you compared different methods (including your own) exactly on the same benchmarks, including the same datasets, search space, code for training and hyperparameters for that code? [Yes]

(g) Did you run ablation studies to assess the impact of different components of your approach? [Yes] these can be found in the appendix

(h) Did you use the same evaluation protocol for the methods being compared? [Yes]

(i) Did you compare performance over time? [Yes]

(j) Did you perform multiple runs of your experiments and report random seeds? [Yes]

(k) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] in the appendix

(l) Did you use tabular or surrogate benchmarks for in-depth evaluations? [No] all computations were computed from scratch. There is no such benchmark for what we did.

(m) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] we did not track this datum

(n) Did you report how you tuned hyperparameters, and what time and resources this required (if they were not automatically tuned by your AutoML method, e.g. in a NAS approach; and also hyperparameters of your own method)? [Yes] yes we reported this

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets…

(a) If your work uses existing assets, did you cite the creators? [Yes] as far as we are aware of citability

(b) Did you mention the license of the assets? [No]

(c) Did you include any new assets either in the supplemental material or as a URL? [Yes]

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] we only use data from OpenML, where consent is granted by definition

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] Does not apply