

1 Reproducibility 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] No unsubstantiated claims.
 - (b) Did you describe the limitations of your work? [Yes] Section 6 describes when method will perform poorly.
 - (c) Did you discuss any potential negative societal impacts of your work? [Yes] No foreseen negative societal impacts.
 - (d) Have you read the ethics 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? [N/A] The bound used comes from Chai (2009), which is directly referenced.
 - (b) Did you include complete proofs of all theoretical results? [N/A] We describe and evaluate theoretical results proved elsewhere.
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)? [No] The core experiments are on a real-world system. However, to enable reproducibility we did create a synthetic system that has similar properties to the real-world system being evaluated in the paper. A description of that is given in Online Appendix 2, and the code, data, and instructions to reproduce the results are all given. Instructions to reproduce results are in https://ax.dev/tutorials/multi_task.html, code is at <https://github.com/facebook/Ax>, and includes requirements.
 - (b) Did you include the raw results of running the given instructions on the given code and data? [Yes] The notebook at https://ax.dev/tutorials/multi_task.html has the raw outputs from the experiment.
 - (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? [No] Not for the experiments on the real-world system, but yes for the synthetic figures in Online Appendix 2.
 - (d) Did you ensure sufficient code quality such that your code can be safely executed and the code is properly documented? [Yes] Continuous integration testing ensures that the notebook continues to execute, and all code has unit test coverage.
 - (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] For the synthetic experiment, model fitting and search spaces are specified in the notebook, and all fixed hyperparameters are in the code.
 - (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] We used (and also have open-sourced) code for comparison

- methods (single-task modeling, and various multi-task approaches). All methods use the same approach for fitting and the same underlying kernel hyperpriors and hyperparameters. 43
44
- (g) Did you run ablation studies to assess the impact of different components of your approach? [Yes] See Online Appendix 2, and Fig. 10. 45
46
- (h) Did you use the same evaluation protocol for the methods being compared? [Yes] All methods evaluated in exactly the same manner. 47
48
- (i) Did you compare performance over time? [Yes] Fig. S4. 49
- (j) Did you perform multiple runs of your experiments and report random seeds? [Yes] We performed sufficiently many replications that the results are not sensitive to random seeds. 50
51
- (k) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Figs. 7, 8, 9, 10, S4. 52
53
- (l) Did you use tabular or surrogate benchmarks for in-depth evaluations? [No] We focused on the real-world system, and a synthetic problem constructed to have similar properties. 54
55
- (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] No GPUs or similar intensive computing units were used in the paper. 56
57
58
- (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)? [No] Hyperparameters were the standard BO ones, and we used off-the-shelf hyperpriors that are given in the code. 59
60
61
62
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets... 63
- (a) If your work uses existing assets, did you cite the creators? [N/A] 64
- (b) Did you mention the license of the assets? [N/A] 65
- (c) Did you include any new assets either in the supplemental material or as a URL? [N/A] 66
- (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] 67
68
- (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] 69
70
5. If you used crowdsourcing or conducted research with human subjects... 71
- (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] 72
73
- (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] 74
75
- (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] 76
77