Supplementary Material:
A simple transfer-learning extension of Hyperband

Lazar Valkov[*], **Rodolphe Jenatton**[*], **Fela Winkelmolen**[*], **Cédric Archambeau**[⋆]

Amazon AWS AI[*] & Amazon research[⋆]

{valkovl, jenatton, winkelmo, cedrica}@amazon.com

Abstract

We provide in this supplementary material details about about the experimental settings.

1 XGBoost binary classification tasks

In the second experiment of the paper, we focus on the tuning of XGBoost binary classifiers to optimize the validation AUC. We tune 8 HPs of XGBoost (reusing the terminology from its API):

- `eta` in $[0, 1]$
- `subsample` in $[0.5, 1]$
- `colsample_bytree` in $[0.3, 1]$
- `gamma` in $[2^{-20}, 64]$
- `min_child_weight` in $[2^{-8}, 64]$
- `alpha` in $[2^{-20}, 256]$
- `lambda` in $[2^{-10}, 256]$
- `max_depth` in $[2, 128]$

noting that all other HPs are left to their default values, in particular the booster is equal to `gbtree`. As discussed in the main paper, `num_round` acts as the resource parameter of Hyperband, in the range $[1, 81]$.

We consider the following subset of $T = 25$ datasets from the `libsvm` repository [1]: {`australian`, `fourclass`, `german.numer`, `gina_agnostic`, `madelon`, `splice`, `breast-cancer`, `higgs_small`, `a6a`, `a7a`, `a8a`, `ijcnn1`, `mushrooms`, `phishing`, `rcv1.binary`, `skin_nonskin`, `spambase`, `susy`, `svmguide1`, `w6a`, `w7a`, `w8a`, `cod-rna`, `a1a`, `w1a`}.

2 Using prior information in the definition of the search space

In the two experimental settings we have considered in the paper—tuning of SGD learning rates and XGBoost binary classifiers, we have not made assumptions on the search spaces. In particular, we have not used any prior information in the form of warping transformations, e.g., logarithmic transformations of some HP ranges spanning several order of magnitudes.

To assess the effect of injecting prior knowledge in the form of warping transformations, we run `random` for XGBoost in an appropriately transformed search space (with `log2_gamma`, `log2_min_child_weight`, `log2_lambda`, `log2_alpha` and a discretized list of `max_depth`={2, 3, 4, 6, 8, 11, 16, 23, 32, 45, 64, 91, 128} encoded as an integer HP corresponding to the indices in this list). We refer to this method as `random-log`.
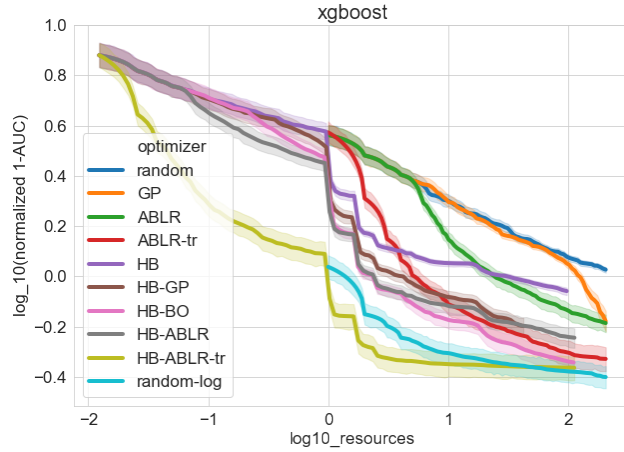
Figure 1: Comparison of different competing methods for the tuning of XGBoost binary classifiers.

In Figure 1, we can observe that appropriate transformations in the search space, if available, lead to a clear improvement. `HB_ABLR_transfer` is the only method operating in the "raw" search space that can first outperform, and then match, `random-log`.

As future work, we plan to further evaluate all the other methods in the transformed search space to better understand the impact of this prior information on the performance.

References

[1] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.