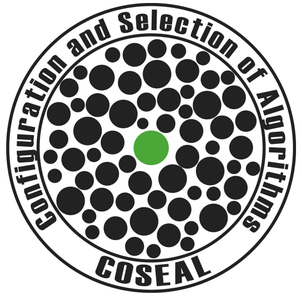


Validation and Explaining Meta-Algorithmics

COSEAL Workshop 2021 - Panel Discussion

Thomas Bartz-Beielstein

Technology
Arts Sciences
TH Köln

1 Validation

“Research strands into ML performance evaluation remain arguably disorganised”

(applies to Auto*, AC, Meta-algorithmic approaches, etc.)

David Jacob Kedziora, Katarzyna Musial, and Bogdan Gabrys. AutonoML: Towards an Integrated Framework for Autonomous Machine Learning, 2020.

Result A:

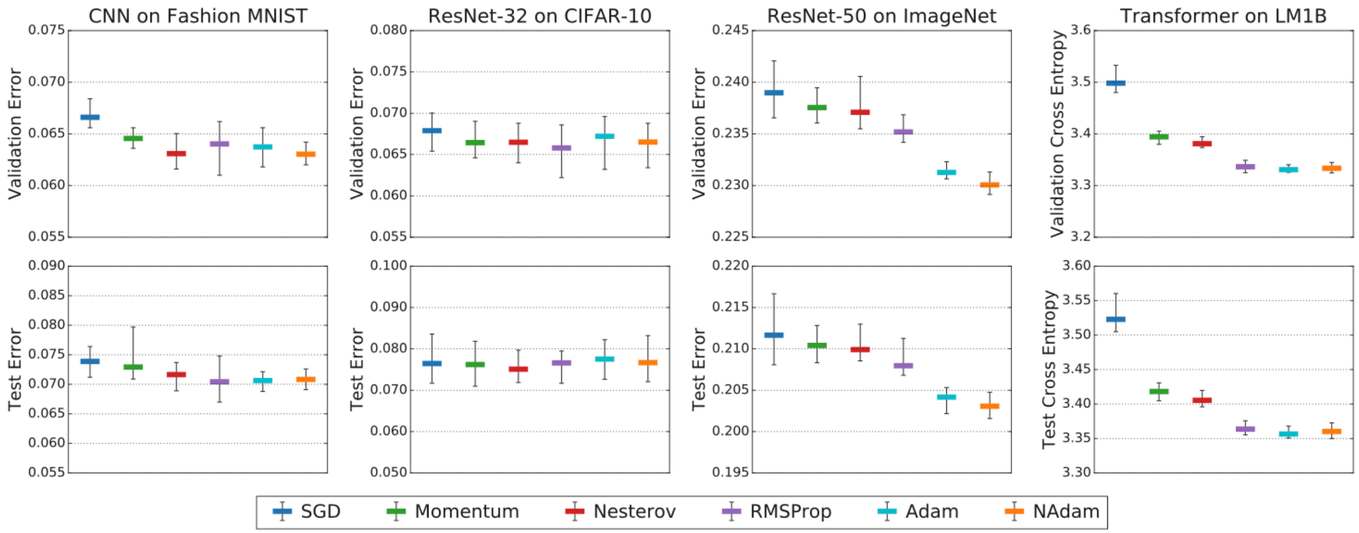
$$\psi_i^{(\text{train})} < \psi_j^{(\text{train})} \not\Rightarrow \psi_i^{(\text{test})} < \psi_j^{(\text{test})}, \tag{2.5}$$

where $\psi_i^{(\cdot)}$ denotes the value of the hyperparameter response surface for the i-th hyperparameter configuration λ_i.

Test error is (not) monotone in validation error

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. J. Mach. Learn. Res., 13:281–305, 2012.

Result B:



"For a relative comparison between models during the tuning procedure, in-sample error is convenient and often leads to effective model selection. The reason is that the relative (rather than absolute performance) error is required for the comparisons."

Dami Choi et al. On Empirical Comparisons of Optimizers for Deep Learning, 2019.

Validation of HPO Results ?

- During Tuning/Optimization: how can progress (improvement of hyperparameters) be validated?
- Several measures (loss and accuracy) and data sets (training, validation, test)
- Model selection versus model assessment
- Role of cross-validation?
  - Variance reduction versus computational time
- HPO w/o validation data, w/o labels?
- Replicability (forget Python - but: even problems with R)

2 Security

Privacy

- Infeasible or undesirable to transmit data to servers
- Devices generate data: bringing code to the data. Federated Learning?
- Still limited to trial-and-error, ad-hoc approaches
- No rigorous theoretical research on the topics such as
    - transferability of data
    - evaluation on surrogates

3 Explaining

XAI

- Fully automated HPO or domain experts?
  - Experts introduce bias
  - Who explains the results (if no experts were involved in the setup)?
- The Mythos of Model Interpretability [Lipt16a]
  - Interpretability Measures: do we seek understandable features, parameters, models, or algorithms?
    - Simulatability, Decompostability, Algorithmic Transpancy, Post-hoc Interpretability (Text Explanation, Visualization, Explaining by Example)!
- European Commission: new rules for Artificial Intelligence

Finally

- Multitude of issues reported when attempting to execute automatic ML frameworks (wrong data splits, numerical instabilities) [Balaji and Allen, 2018]

- Auto-*: human experts are shifted to a higher level, but are still in the loop [Liu, 2018]

- [It] “requires only a single line of Python to train highly accurate machine learning models on an unprocessed tabular dataset such as a CSV file” [Erickson et al., 2020]

Feedback Welcome

- Bartz-Beielstein et al.: Surrogate Model Based Hyperparameter Tuning for Deep Learning with SPOT
- https://arxiv.org/abs/2105.14625
- Working paper: version 4 will be uploaded soon