

AutoML - Project 2019

Improving CNN performance on K49

Philipp Jankov

Motivation

- ▶ New Task/Problem
 - ▶ Relying on known estimates for good values suboptimal
 - ▶ „Endless“ amount of configurations and algorithms
 - ▶ Large configspaces need more samples
- ▶ No Free Lunch Theorem^[1]
 - ▶ „... for any algorithm, any elevated performance over one class of problems is offset by performance over another class...”

[1] No free lunch theorems for optimization, David H. Wolpert and William G. Macready, 1997

Approach

- ▶ Use BOHB with the rightside configuration space
- ▶ Use refined gridsearch to determine initial LR
- ▶ Select pretrained weights by distance in feature space (warmstart)
- ▶ Let Baseline compete against un/-pretrained alternative

BOHB - ConfigSpace

- **Batchsize**
- **Model**
 - BaselineCNN
 - Resnet18
- **Warmstart** (Resnet18 with 64x64 input only)
- **Model arguments**
 - **Dropout**
 - **zero_init_residuals** (Resnet18 only)
- **Optimizers**
 - Adam
 - AdamW (with Nestervov)
- **Optimizer arguments**
 - **Amsgrad**
 - **Nesterov** (AdamW only)
 - **Weight decay**
- **Transforms**
 - None
 - **Predefined Pipeline**
 - Random Rotation
 - Random Perspective
 - Random Elastic Transformation
 - Resize

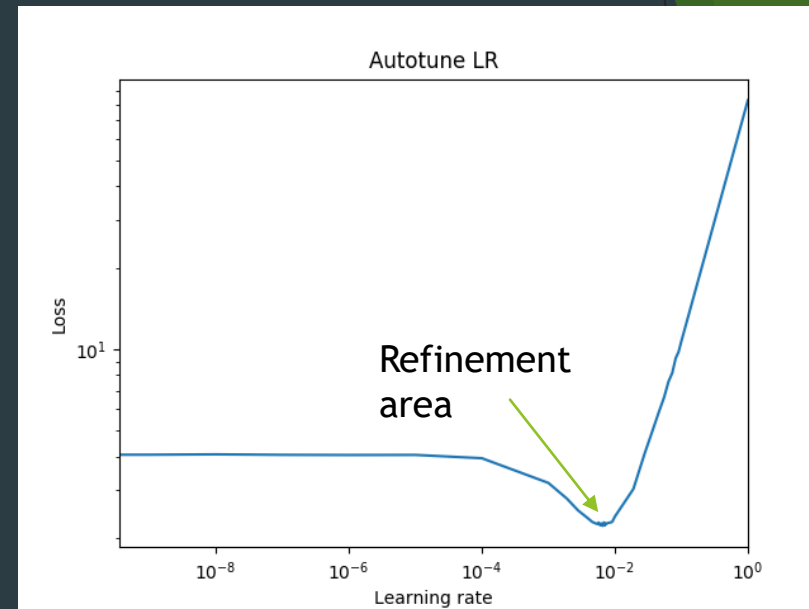
Categorical, **Bool**, **UniformFloat**, **UniformInt**, Choices

Autotune LR^[1]

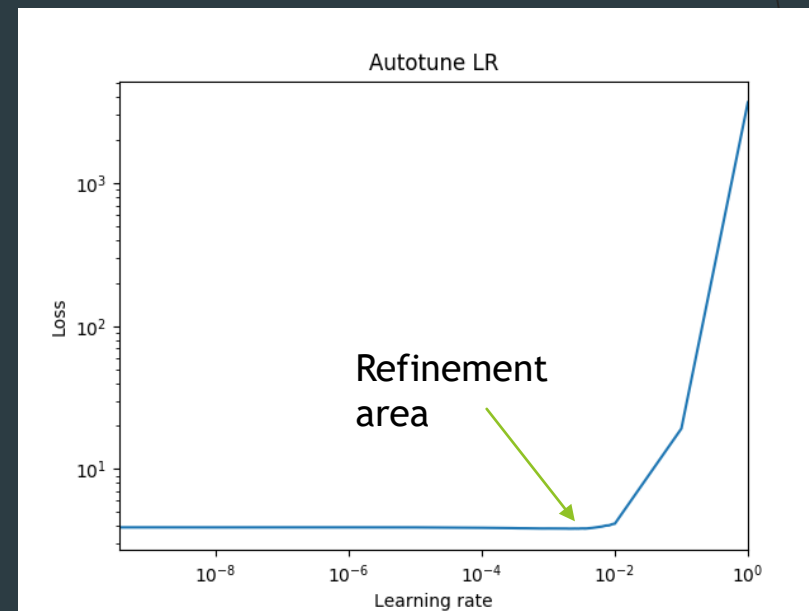
- ▶ Remove LR from configspace
- ▶ Refined Gridsearch to find initial LR
 - ▶ Determine magnitude
 - ▶ Refine around minimum
- ▶ Complements adaptive Optimizer
- ▶ Possible indicator for poor final performance

[1] Cyclical Learning Rates for Training Neural Networks, Leslie N. Smith et al, 2017

Philipp Janko - AutoML Project 2019

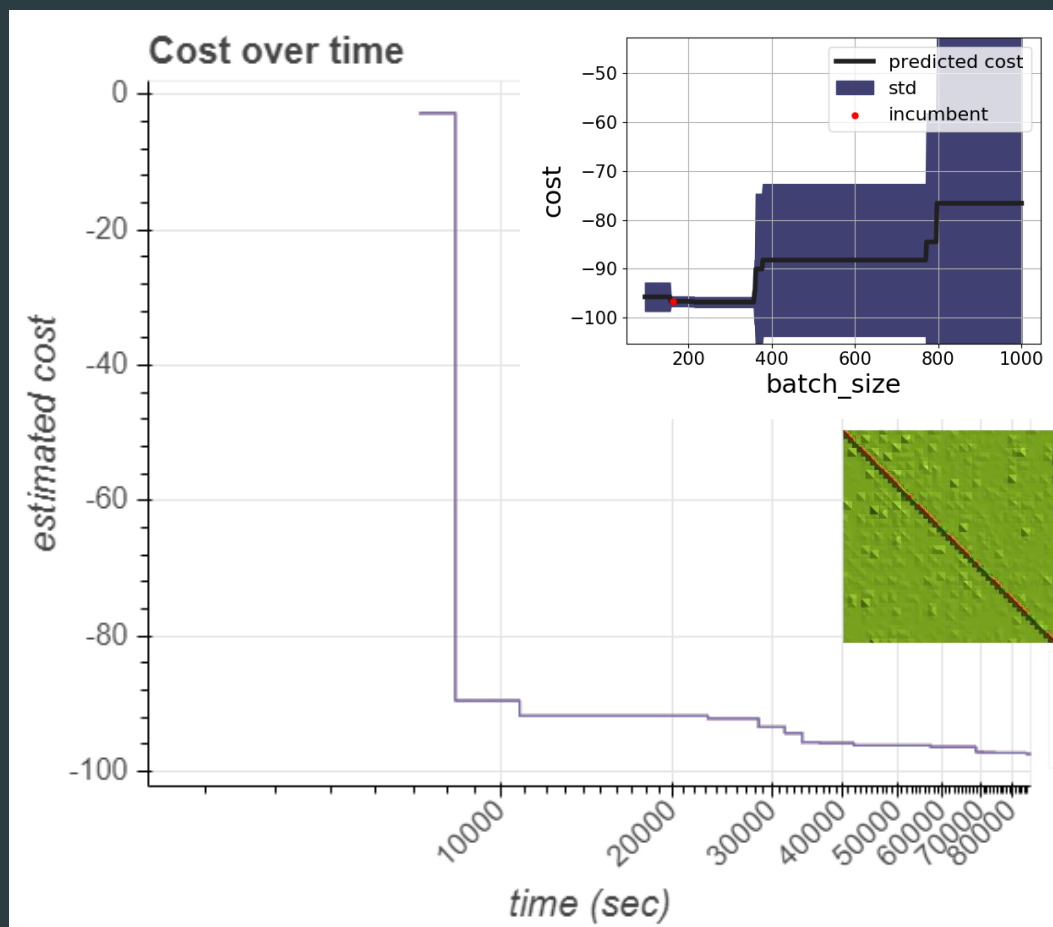


Batchsize 238
Resnet
Dropout 0.42
AdamW
WDecay 0.04
Validation acc
93.53



Batchsize 636
BaselineCNN
Dropout 0.53
AdamW
WDecay 0.14
Validation acc
52.81

Results



Total time spent evaluating configurations 85970.40 sec

Average time per configuration (mean / std) 2046.91 sec
(± 509.23)

evaluated configurations 42

Configuration origins Acquisition Function : 27
Random : 15

fANOVA

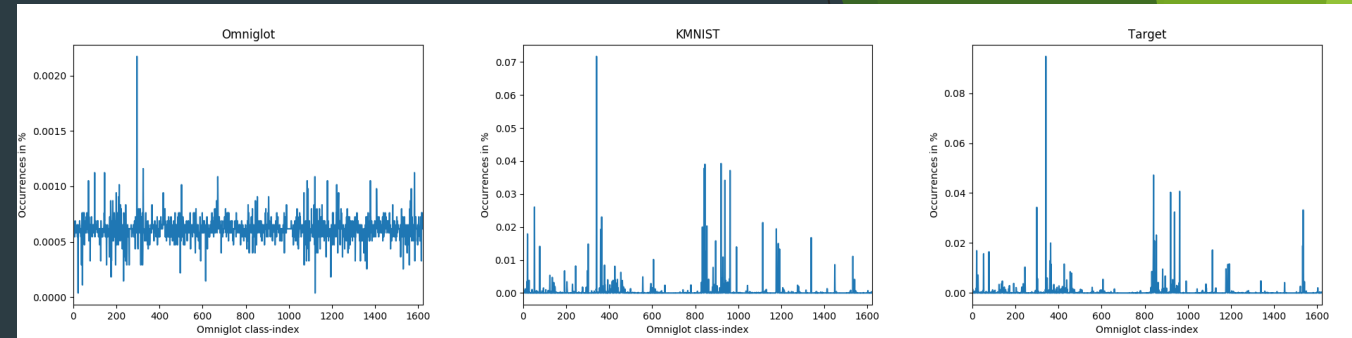
| | Single importance | Final choice |
|--------------------|---------------------|--------------|
| optimizer | 70.9337 +/- 36.37 | AdamW |
| batch_size | 15.2346 +/- 25.9232 | 163 |
| weight_decay | 1.1682 +/- 1.264 | 0.00122816 |
| amsgrad | 0.846 +/- 2.8297 | False |
| zero_init_residual | 0.7148 +/- 2.7682 | True |
| transforms | 0.549 +/- 1.4886 | NONE |
| dropout | 0.3154 +/- 0.6798 | 0.755828 |
| model | 0.1354 +/- 0.1334 | Resnet |

5-fold Cross Validation

| Dataset | BaselineCNN | Resnet |
|------------|--------------|---------------|
| Validation | 78.2373±0.99 | 97.3933±0.206 |
| Test | 66.1893 | 95.3434 |

Discussion Warmstarting

- ▶ Test initial weights from portfolio
- ▶ Use omniglot as feature space
 - ▶ Determine distance between supportsets and target
 - ▶ Select support set by smallest symmetric difference in feature space
 - ▶ Select closest set and use its pretrained weights
- ▶ Known issues
 - ▶ Made the use of augmentation and warmstart codependent
 - ▶ Omniglot's pretrained weights were never chosen



Warmstart effect on loss (final configuration)

