



MLflow Tracking

Erik Genthe

Hamburg, 01.12.2020



Background

A machine-learning project tends to evolve into a big mess during its lifetime

- In ML-Projects you often have hundreds of experiments
- You may remember the best ones, but very unlikely the 99 other things you unsuccessfully tried out.
- How can you document these experiments automatically.
- Version history with included residuals, results and Network-Weights.
- Git is not made for big files.
- How to compare different versions with each other?
 - You may want to sort experiments by specific metrics
 - Or compare curves of arbitrary experiments



MLflow Tracking

Overview

<https://mlflow.org>

MLflow Tracking

Record and query experiments: code, data, config, and results

[Read more](#)

MLflow Projects

Package data science code in a format to reproduce runs on any platform

[Read more](#)

MLflow Models

Deploy machine learning models in diverse serving environments

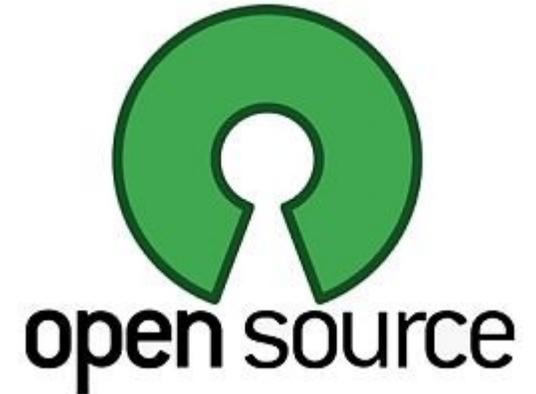
[Read more](#)

Model Registry

Store, annotate, discover, and manage models in a central repository

[Read more](#)

- MLflow a platform to manage the ML lifecycle, experimentation, maintain reproducibility and more
- Super simple API (Python, Java, R, REST), easy to include to existing code
- Works with any ML library, language.
- It provides an API and user interface for
 - logging parameters
 - code versions
 - logging metrics
 - logging output files (common filetypes can be directly visualized as well)



MLflow Tracking – Python API

All you need to know to include MLflow-functionality in your python-script

- `mlflow.start_run(run_id=None, experiment_id=None, run_name=None, nested=False)`
- `mlflow.end_run()`
- `mlflow.log_param(key, value)`
- `mlflow.log_metric(key, value, step=None)`
- `mlflow.log_artifact(path)`

For more details: <https://mlflow.org/docs/latest/tracking.html#logging-functions>

How to start the UI-Server

- Install locally
- No setup required
- In your project-path run the cli-command: `mlflow ui`
- Per default the server ist started on port 5000

MLflow Tracking - UI

Overview

localhost:5000/#/experiments/5/s?orderByKey=metrics.%60Validationloss%60

mlflow Experiments Models GitHub Docs

Experiments + <

Search Experiments

Default v_0.2 v_0.3 v_0.4 overfitting_test

v_0.4

Experiment ID: 5 Artifact Location: /home/egenthe/Desktop/mlflow/mlruns/5

Notes

None

Search Runs: metrics.rmse < 1 and params.model = "tree" and tags.mlflow.s State: Active Search Clear

Showing 13 matching runs Compare Delete Download CSV Columns

	Start Time	Run Name	User	Version	Parameters	Metrics
<input type="checkbox"/>					Batchsize Learningrate	Trainingloss ↑ Validationlo
<input type="checkbox"/>	2020-11-30 22:45:37	9_bb_loss...	egenthe	863f3d	16 0.03	0.07 0.09
<input type="checkbox"/>	2020-11-25 18:06:45	8	egenthe	93226c	16 0.03	0.01 0.095
<input type="checkbox"/>	2020-11-30 20:19:33	8_squared...	egenthe	20115a	16 0.03	0.022 0.102
<input type="checkbox"/>	2020-11-30 19:56:31	8_BS=16	egenthe	20115a	16 0.03	0.015 0.102
<input type="checkbox"/>	2020-11-30 20:19:33	8_BS=4	egenthe	20115a	4 0.03	0.02 0.362
<input type="checkbox"/>	2020-12-01 10:16:42	overfitting...	egenthe	863f3d	16 0.03	0.361 0.475
<input type="checkbox"/>	2020-12-01 13:25:15	overfitting...	egenthe	431512	4 0.1	0.446 0.476
<input type="checkbox"/>	2020-12-01 13:18:08	overfitting...	egenthe	863f3d	4 0.1	0.293 0.536
<input type="checkbox"/>	2020-12-01 10:38:00	overfitting...	egenthe	863f3d	16 0.03	1.09 1.079
<input type="checkbox"/>	2020-12-01 13:26:25	overfitting...	egenthe	431512	1 0.01	2.168 1.527
<input type="checkbox"/>	2020-12-01 13:25:36	overfitting...	egenthe	431512	4 0.01	1.638 2.29
<input type="checkbox"/>	2020-12-01 10:38:55	overfitting...	egenthe	863f3d	16 0.03	4.033 4.68

v_0.4 > E5

Detailed run information

Date: 2020-12-01 13:25:15

Source: ETPD.py

Git Commit: 431512c689e4aa490147fa6e142b6f

User: egenthe

Duration: 4.1min

Status: FINISHED

Notes

Testnote

Markup Test

Bold

Parameters

Name	Value
Batchsize	4
Filter divisor	1
Learningrate	0.1

Metrics

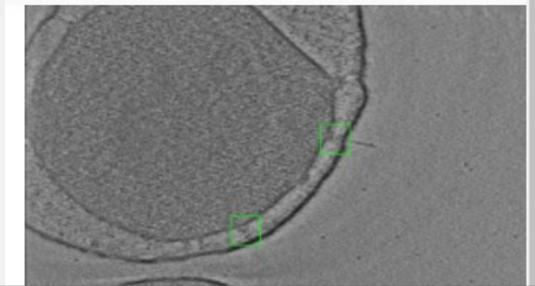
Name	Value
Epoch	100
Trainingloss	0.446
Validationloss	0.476

Tags

Artifacts

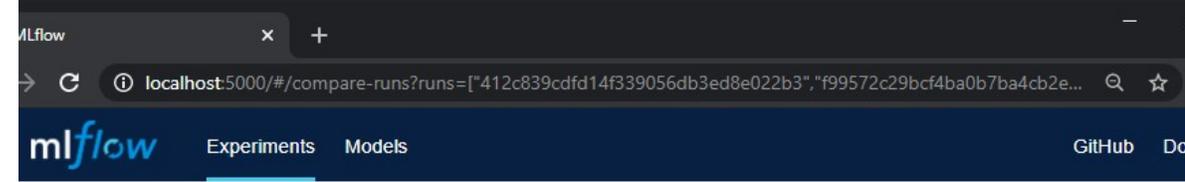
- console.txt
- gpu.txt
- model.pth
- train_4565.png
- train_64002.png
- train_80183.png
- validation_89707.png
- validation_92687.png
- validation_93728.png
- validation_96890.png

Full Path: /home/egenthe/Desktop/mlflow/mlr...
Size: 367.47KB



MLflow Tracking - UI

Comparison of runs



v_0.4 > Comparing 7 Runs

Run ID:	412c839cdf1...	f99572c29bcf...	d5dc0f260e9a...	2f270c4e6bbf...	a271e2f31f32...	eb13288fa13d...	c4796fcb6e0...
Start Time:	2020-12-01 13:26:25	2020-12-01 13:25:36	2020-12-01 13:25:15	2020-12-01 10:38:55	2020-12-01 10:38:00	2020-12-01 10:34:15	2020-12-01 10:16:42

Parameters

Batchsize	1	4	4	16	16	16	16
Filter divisor	1	1	1	1	1	1	1
Learningrate	0.01	0.01	0.1	0.03	0.03	0.03	0.03

Metrics

Epoch	100	100	100	40	40	2	40
Trainingloss	2.168	1.638	0.446	4.033	1.09	5.489	0.361
Validationloss	1.527	2.29	0.476	4.68	1.079	4.992	0.475

Scatter Plot

Contour Plot

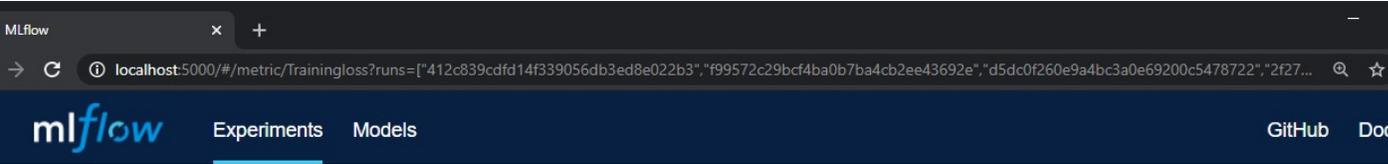
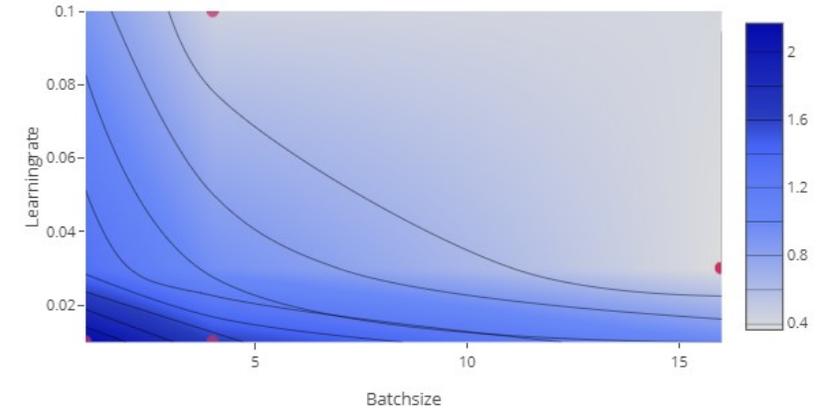
Parallel Coordinates Plot

X-axis:

Y-axis:

Z-axis:

Reverse color: Off



v_0.4 > Comparing 7 Runs > Trainingloss

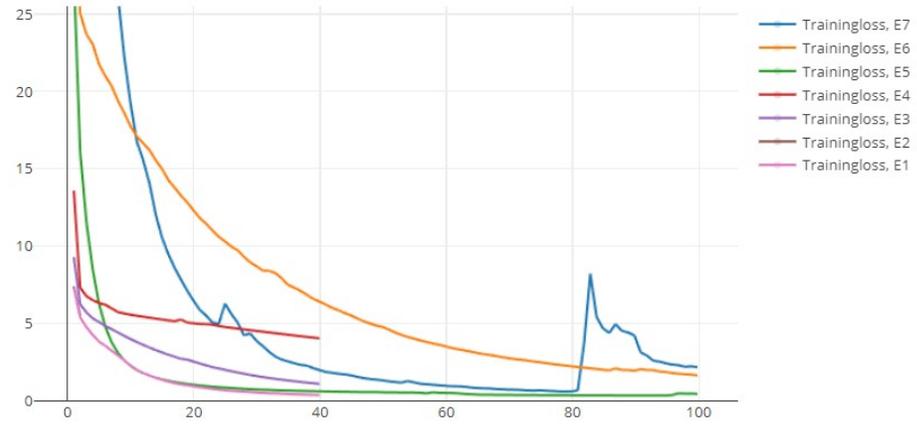
Points: Off

Line Smoothness

X-axis: Step Time (Wall) Time (Relative)

Y-axis:

Y-axis Log Scale: Off



Thank you very much

Kontakt

DESY. Deutsches
Elektronen-Synchrotron

Erik Genthe
mail@erikgenthe.de

www.desy.de