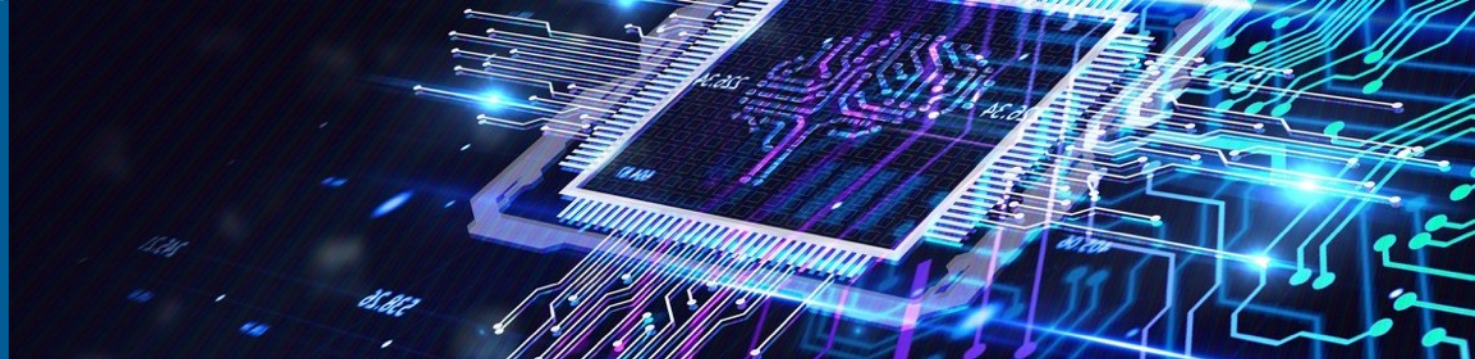




CSC

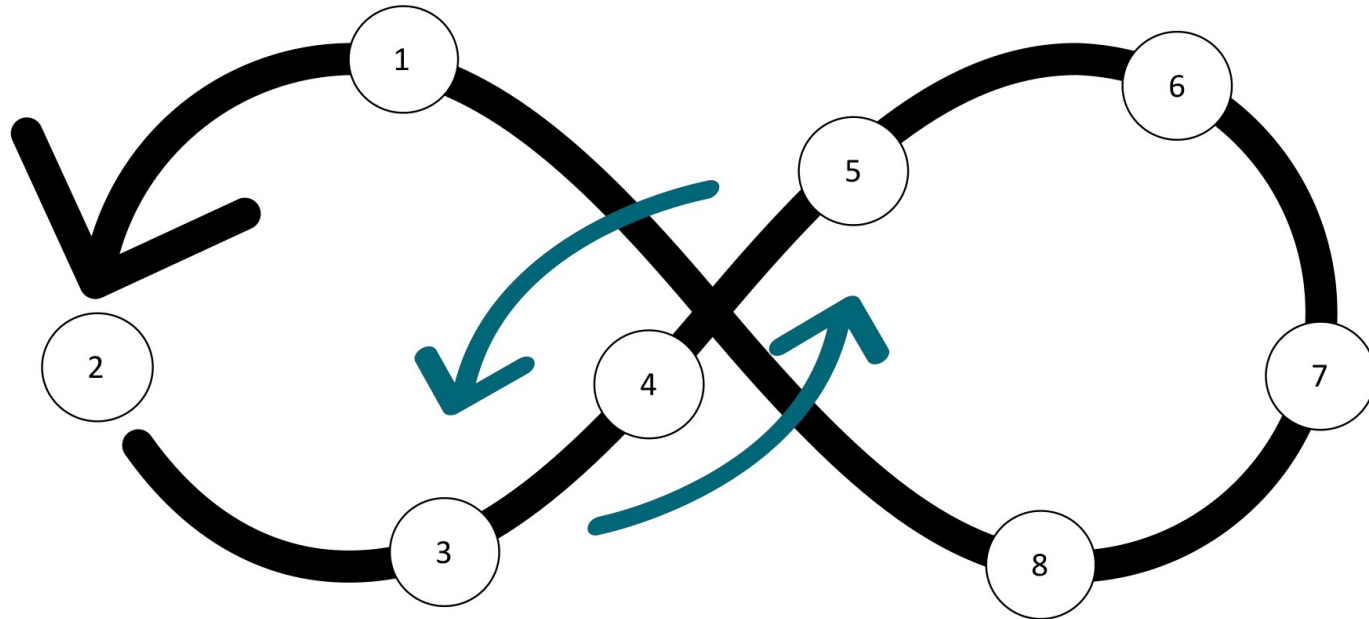
ICT Solutions for
Brilliant Minds



CSC Research Support Coffee: Streamlining ML workflows with MLflow

Erika Sternad 4.12.2024





1. Data collection
2. Data preparation
3. Model development
4. (Re-)Training

5. Evaluating
6. Deployment
7. Inference
8. Monitoring



Tracking Server

- Model tracking, saving metrics, used parameters etc.
- GUI

Models

- Saving models consistently
- Serving models locally

Model Registry

- Registering and versioning models
- Managing models in production with aliases

Projects

- Wrapping entire ML projects in git-style



Experiments

MLflow tutorial

Share

Search Experiments
Default
MLflow tutorial

Runs Evaluation Experimental Traces Experimental

metrics.rmse < 1 and params.model = "tree"

State: Active Datasets Sort: Created Columns

Group by

Run Name	Created	Dataset
3Layer_adam	2 months ago	dataset (5118c5...
3Layer_adagrad	2 months ago	dataset (869a0c...

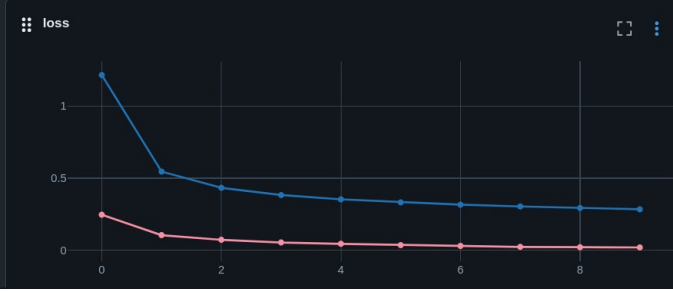
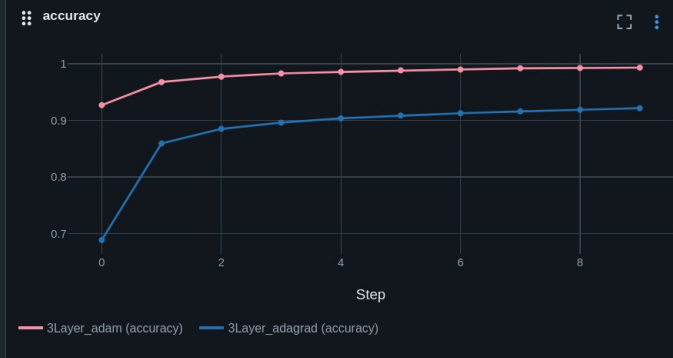
metrics.rmse < 1 and params.model = "tree"

Time created

State: Active Datasets Sort: Created Group by

Run Name

- 3Layer_adam
- 3Layer_adagrad



Registered Models

mnist_classifier

Created Time: 2024-11-03 17:34:01

Last Modified: 2024-11-03 17:34:35

Description Edit

Tags

Versions Compare

Version	Registered at	Created by	Tags	Aliases
Version 1	2024-11-03 17:34:01		Add	@ champion

erikaster / puhti_mlflow_tutorial (Public)

<> Code Issues Pull requests Actions Projects Security Insights

main 2 Branches 0 Tags

erikaster Update README.md 6h407a · 3 weeks

- pics added logo
- .gitignore Added '.gitignore'D
- README.md Update README.md
- puhti_mlflow_tutorial.ipynb cleaned up a bit

README

Instructions for Using the "Tutorial for ML Puhti"

By Erika Sternad / CSC 2024

Updates:

- 2024-9-2: Added first version
- 2024-9-13: Added README with instruction for testing
- 2024-9-25: Added second version with updates based on feedback
- 2024-10-25: Typo fixes and minor modifications

Getting Started

- Setup:**
 - You can either clone this repository directly to Puhti or download the zip file and n location, such as `/projapp1/project_id/username/` or `/scratch/project_id/` repository directly, first run `module load git` to load the Git module.

- Projects** is for packaging entire ML project code, enabling easy sharing and reproducibility.

By organizing your work into **experiments** and **runs**, MLflow ensures that you can systematically track progress, compare results, and refine your models effectively.

For more info on components visit [MLflow documentation](#).

```
In [ ]: # Prevent the creation of the .ipynb_checkpoints folder for current session
get_ipython().kernel.parent.contents_manager.checkpoints_class = None

In [ ]: # Let's make sure we have the latest version of MLflow (>2.15.1) in use. If not, run pip install below:
!mlflow --version
#!pip install --upgrade mlflow

In [ ]: # With sample code, we use Gradio Interface to demo deployment and inference:
!pip install gradio

In [ ]: !import pandas as pd
# Ensure that the pytorch/2.4 module is selected in the settings section, as it includes a sufficiently recent version of MLflow.
```

Additionally, make sure the file path points to the previously specified `mlruns` directory.

Settings

Module

pytorch/2.4

Location where MLflow files are stored

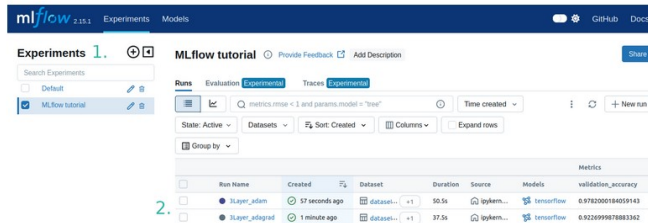
/scratch/project_2001234/path/to/mlruns

Directory or URI to where MLflow data is stored, typically set with the `mlflow.set_tracking_uri()` method. Examples:

- `/scratch/project_2001234/mlruns/`
- `sqlite:///scratch/project_2001234/mlruns.db`

Figure 1. Setup MLflow in Puhti

After launching successfully, we can view the previous runs:



mlflow 2.15.1 Experiments Models

Experiments 1. MLflow tutorial

Search Experiments: MLflow tutorial

Runs Evaluation Experimental Traces Experiment

metrics.mean < 1 and params.model = "tree"

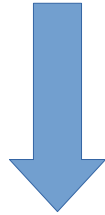
Time created

State: Active Datasets Sort Created Columns Expand rows

Group by

	Run Name	Created	Dataset	Duration	Source	Models	Metrics
	!layer_adam	57 seconds ago	dataset... +1	50.5s	ipynb...	tensorflow	0.9782000184059143
	!layer_adagrad	1 minute ago	dataset... +1	37.5s	ipynb...	tensorflow	0.922099978883262

Figure 2. Front page of MLflow UI



https://github.com/erikaster/puhti_mlflow_tutorial



Erika Sternad

Junior Machine Learning Engineer

+358 503373955

erika.sternad@csc.fi



facebook.com/CSCfi



twitter.com/CSCfi



linkedin.com/company/csc---it-center-for-science



github.com/CSCfi