“Tools”

CAS AML 2024

Dr. Mykhailo Vladymyrov

Data Science Lab
University of Bern

Build your workflow!

To get anywhere – you need multiple experiments.

- Different approaches
- Different data
- Different hyperparameters

You need to build your workflow such that it remains organied also after few dosens iterations.

MLFlow

MLflow is an open-source platform, purpose-built to assist machine learning practitioners and teams in handling the complexities of the machine learning process.

MLflow focuses on the full lifecycle for machine learning projects, ensuring that each phase is manageable, traceable, and reproducible.

MLFlow

1) Log experiment metrics, parameters, and models
2) Compare runs

Results can be stored in a file, or on a server.

In case of storing to file, I recommend local installation for running the visualization.

Optuna

An open source hyperparameter optimization framework to automate hyperparameter search.

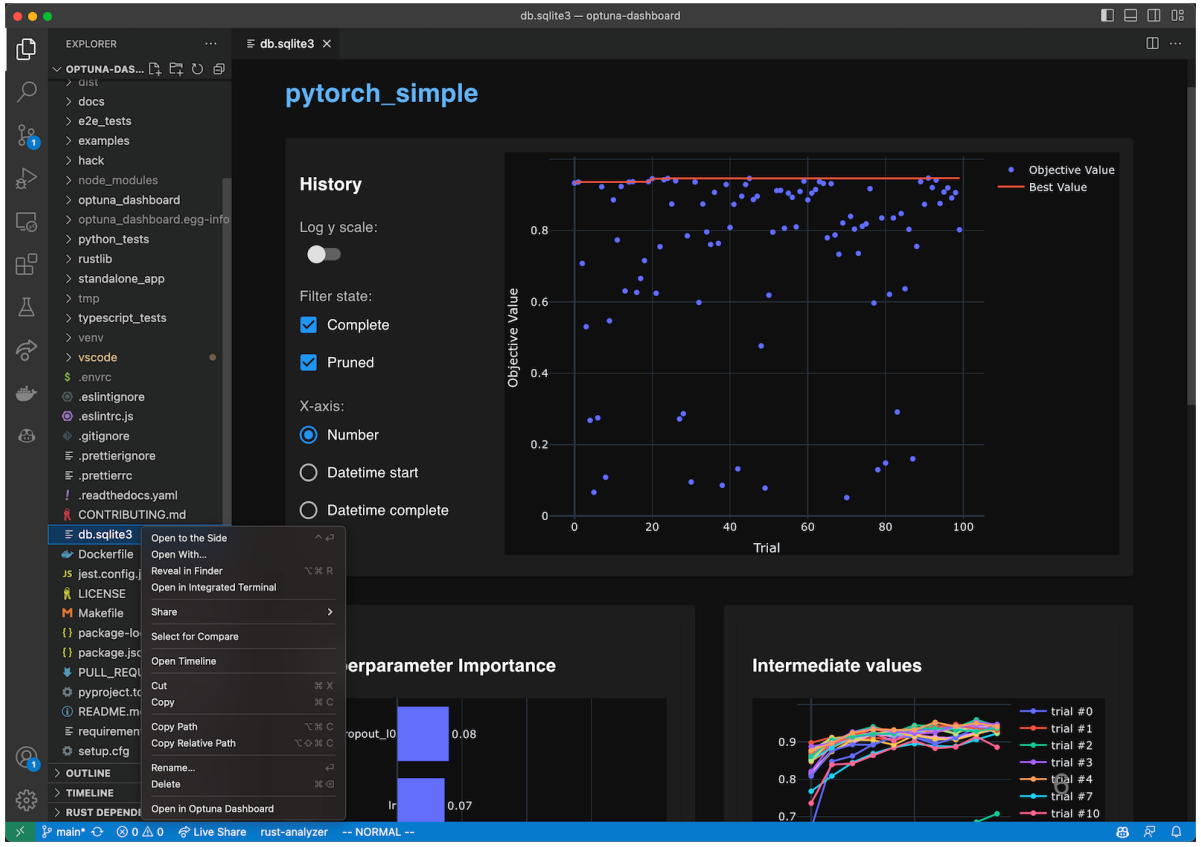
Helps tune hyperparameters in a systematic manner...

... does not replace your intuition and prior knowledge.

Optuna

Searches for optimal HP

Visualize effect of the parameters on performance



Optuna Applicability

Very good for medium-fast training models, few hyperparameters.

Limited applicability for models that train for weeks.

You need to know reasonable range of HPs.

Further readings

MLOps landscape:
https://neptune.ai/blog/mlops-tools-platforms-landscape

MLFlow docs:
https://mlflow.org/docs/latest/tracking.html
https://mlflow.org/docs/latest/llms/tracing/index.html

Optuna docs:
https://optuna.readthedocs.io/en/stable/tutorial/10_key_features/003_efficient_optimization_algorithms.html
https://optuna-dashboard.readthedocs.io/en/stable/getting-started.html

Workshop

45min