Introduction to mlflow

WHAT IS MLFLOW?

MLflow is an open source platform for managing the end-to-end machine learning lifecycle. It tackles four primary functions:

- Tracking experiments to record and compare parameters and results (MLflow **Tracking**).
- Packaging ML code in a reusable, reproducible form in order to share with other data scientists or transfer to production (MLflow **Projects**).
- Managing and deploying models from a variety of ML libraries to a variety of model serving and inference platforms (MLflow **Models**).
- Providing a central model store to collaboratively manage the full lifecycle of an MLflow Model, including model versioning, stage transitions, and annotations (MLflow **Model Registry**). [1]

In gerneral, MLflow can be integrated at **any time** of the ML lifecycle.

WHAT IS MLFLOW?

- Developed by Databricks
- Alpha Launch in summer 2018
- 9.6k GitHub stars in June 22
- There is an **open-source** version available as well as a **paid version** ("Managed MLFlow by Databricks")
- In general, mlflow is library-agnostic since all functions are available via REST and CLI. But there exist APIs for Python, R and Java.
- The Python API provides modules with APIs for common ml libraries like scikit-learn, pytorch, keras and xgboost.