

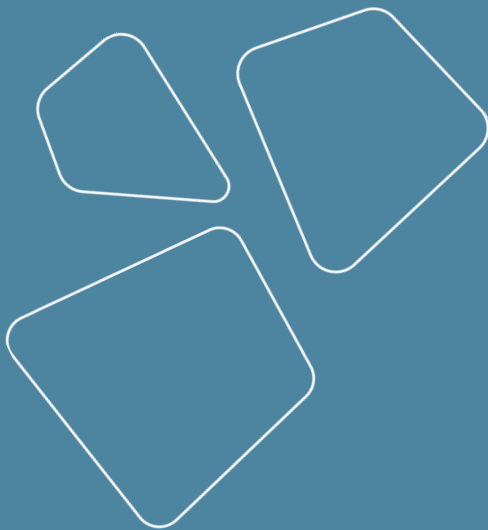
Логирование и управление экспериментами

Vladislav Goncharenko



HSE, fall 2023

Recap



- Источники данных
- Хранение данных
 - локально
 - удалённо
 - распределённо
- Data engineering
 - ETL and ELT
 - datahub
- Experiments parametrization
 - Hydra

Outline

- Логирование
- Визуализация
- Фреймворки для обучения



Логирование

girafe
ai

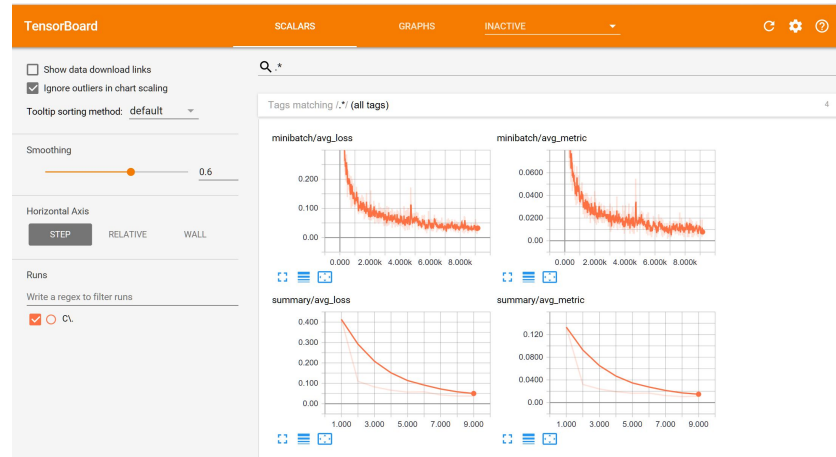
01

Логирование экспериментов

Тем важнее, чем больше экспериментов и крупнее каждый из них

Подходит для экспериментов, не для логирования прода

- отсутствие
- print
- local service
 - tensorboard
- remote service
 - ml-flow
 - clear-ml
 - kubeflow
 - w&b, neptune и прочая проприетарщина



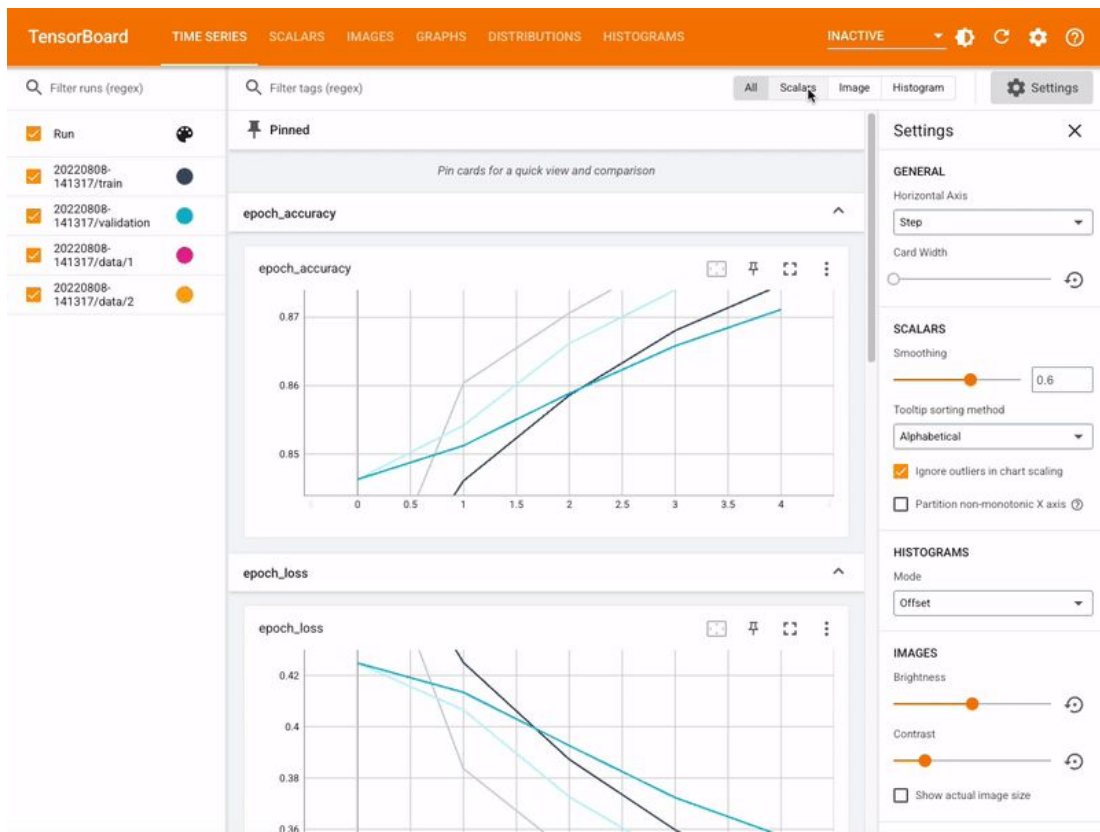
Ещё больше логирования...

- [Sacred](#)
- [Flyte](#)
- Comet
- Valohai
- Polyaxon
- Domino Data Lab
- Cortex
-

See overviews: [one](#), [two](#)

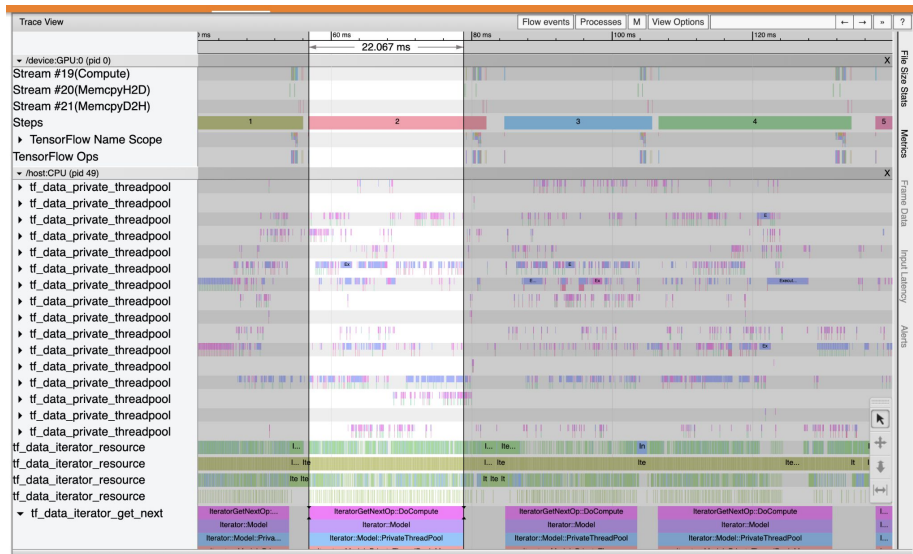
Tensorboard

- Является частью Tensorflow, но поставляется отдельным пакетом
- Лучше всего подходит для локального логирования
- Пытался развиваться в standalone сервер как TensorBoard.dev, но проект закрылся



Возможности Tensorboard

- Tracking and visualizing metrics such as loss and accuracy
- Visualizing the model graph (ops and layers)
- Viewing histograms of weights, biases, or other tensors as they change over time
- Projecting embeddings to a lower dimensional space
- Displaying images, text, and audio data
- Profiling TensorFlow programs



MLFlow



- Tracking
 - Record and query experiments: code, data, config, and results
- Projects
 - Package data science code in a format to reproduce runs on any platform
- Models
 - Deploy machine learning models in diverse serving environments
- Model Registry
 - Store, annotate, discover, and manage models in a central repository

ClearML



ClearML DataOps

ClearML Hyper-Datasets

ClearML Experiment

ClearML Train

ClearML Reports

ClearML Modelstore

ClearML Pipelines

ClearML Deploy

ClearML Orchestrate

ClearML Schedule

ClearML Compute

ClearML Autoscalers

ClearML Enterprise

Kubeflow

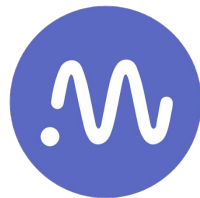
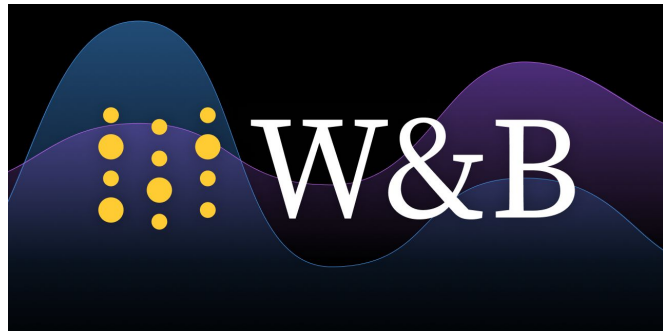
ML suite for Kubernetes (includes tracking)



Kubeflow

W&B, neptune и прочая проприетарщина

- neptune.ai
 - неплохие обзоры рынка и база знаний
- [Weights and biases](https://weightsandbiases.com)
 - предоставляют self-hosted версию

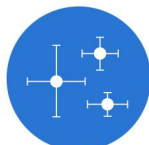


neptune.ai

Visualisations

Comprehensive list of visualisations:

<https://datavizcatalogue.com/>



Error Bars



Flow Chart



Flow Map



Gantt Chart



Heatmap



Histogram



Illustration Diagram



Kagi Chart



Line Graph



Marimekko Chart



Multi-set Bar Chart



Network Diagram



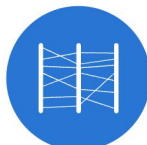
Nightingale Rose Chart



Non-ribbon Chord Diagram



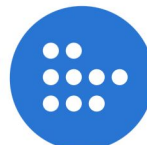
Open-high-low-close Chart



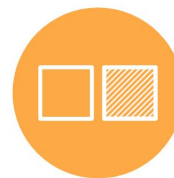
Parallel Coordinates Plot



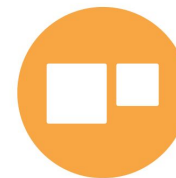
Parallel Sets



Pictogram Chart



Comparisons



Proportions



Relationships



Hierarchy



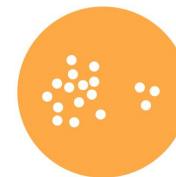
Concepts



Location



Part-to-a-whole



Distribution



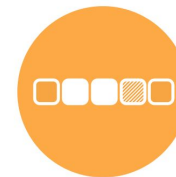
How things work



Processes & methods



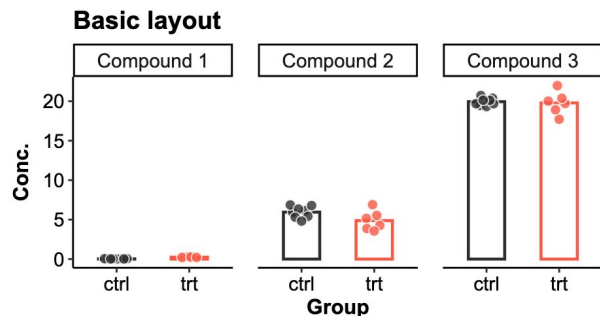
Movement or flow



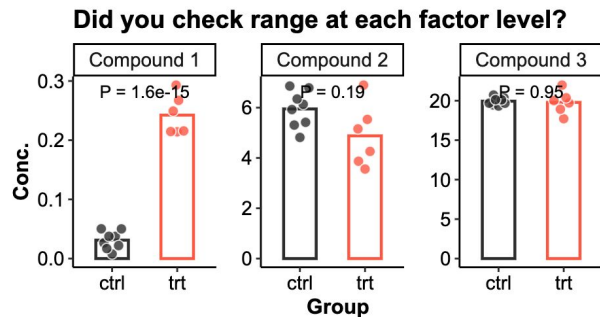
Patterns

Bad cases of visualization

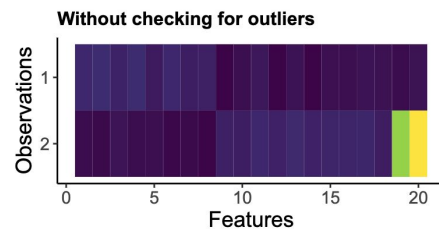
<https://github.com/cxli233/FriendsDontLetFriends>



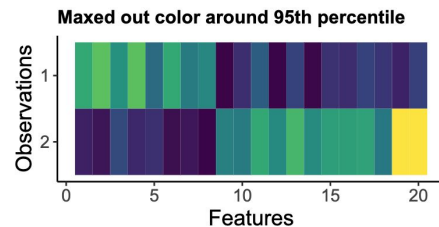
My treatment wasn't doing anything!



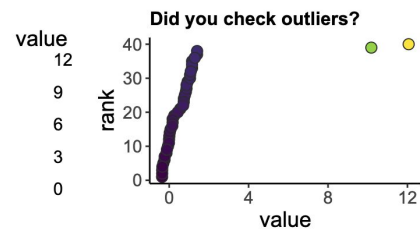
OH!!!



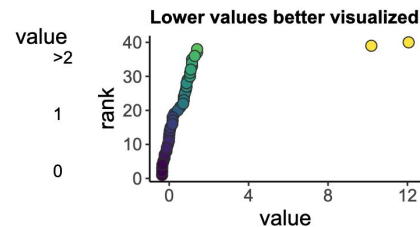
Only two values are different!



Two observations are very different!



Wait...



That's better...

Фреймворки для обучения

girafe
ai

02

Experiments Runners

- Pytorch
 - [Ignite](#)
 - [Catalyst](#)
 - [Lightning](#)
 - hugging face
- [Ray](#)
- DVC Experiments
- MLFlow Experiments
- [pachyderm](#)
- [fast.ai](#) (very poor coding level e.g. import *)
-

P.S. see <https://pytorch.org/ecosystem/>



PyTorch Lightning

Catalyst



PyTorch
Ignite

Revise

- Логирование
- Визуализация
- Фреймворки для обучения



Спасибо за внимание!

Жду вопросов и обсуждений

