Cheat Sheet
Machine Learning

Frameworks

PyTorch
Research oriented framework with many pretrained models.

Tensorflow
Large sets of integrations & libraries supported with a focus on production-ready use cases.

JAX
Google-supported framework designed for high-performance numerical computing.

Pytorch Lightning
Provides a high level interface for Pytorch by focusing on reproducibility, easy SOTA integration and best practices for training DL-models.

scitkit-learn
Machine Learning framework offering many algorithms out of the box & data analysis tools.

Keras
High-level abstraction layer for TensorFlow. Allows for quickly defining and training models with the drawback of reduced flexibility.

Technical Set-up

Datasets

CIFAR10
Image dataset with objects belonging to 10 different classes (famous benchmark).

Titanic
Tabular dataset describing different passengers on the titanic (famous toy dataset).

SQuAD
Question and Answer dataset for training and evaluating language models.

Imagenet
Used for evaluating SOTA and pre-training vision models. Consists of over 14 million images split into 22k different classes.

MS COCO
Large dataset for object detection and image segmentation dataset used for pre-training and benchmarking.

Kaggle
Platform for open datasets where the users can compete in competitions and share code.

Papers with Code
Provides benchmark leader boards, links to SOTA papers, and descriptions about methods.

Tip
«How to develop a ML model on the MNIST handwritten digit recognition task in Python using Keras.»
By J. Brownlee

Tip
«A visual introduction to machine learning»
By R2D3

Tip
SL requires large amounts of labelled examples for training.

Supervised Learning (SL)

Classification
Prediction of discrete classes, e.g. image classification.

Anomaly Detection
Find anomalies in the data by finding outliers (data that does not fit the rest). Also works for UL.

Regression
Prediction of continuous values, e.g. house price prediction.

Methods & Applications

Tip
UL focuses on extracting patterns in unlabelled data.

Unsupervised Learning (UL)

Generative Modelling
Generate new data from a set of known samples, useful for simulations of expensive physical operations.

Latent Variable models
Extract the underlying variables that influence the data for compression and further analysis.

RL trains an agent which observes its environment and selects an appropriate action. The agent then receives feedback in the form of a reward which it uses for learning. Can be used for the control of robots

Tip

Reinforced Learning (RL)

Use cases

Predictive Maintenance
Predict failures of machines from their logs and sensor data.
SL + RL

User Input
Learn from user input to create new interfaces (e.g. Human-Brain interfaces).
SL

Fraud Detection
Detects fraudulent or unexpected behavior. (e.g. credit card identity fraud).
SL + UL

Medical Analysis
Crawl patients' medical histories to predict likelihoods of certain diseases.
SL + UL

Autonomous Control & Driving
Autonomously drive a car.
SL + RL

Object Localization
Locates objects in images.

Optimal Control
Optimally tune a controller which runs a physical system.
UL + RL

Production Fault Check
Automatically check manufactured objects for problems or defects.
SL + UL

Data Correction
Automatically detect and optionally correct wrong user inputs.
SL + UL

Timeseries Prediction
Predict the value of a variable based on previous knowledge.
SL

Tip
«10 real-world examples of ML and AI»
By M. Feldmann

MOTIUS
WE R&D.