University of Basel

SDSC

Advanced Python and Machine Learning

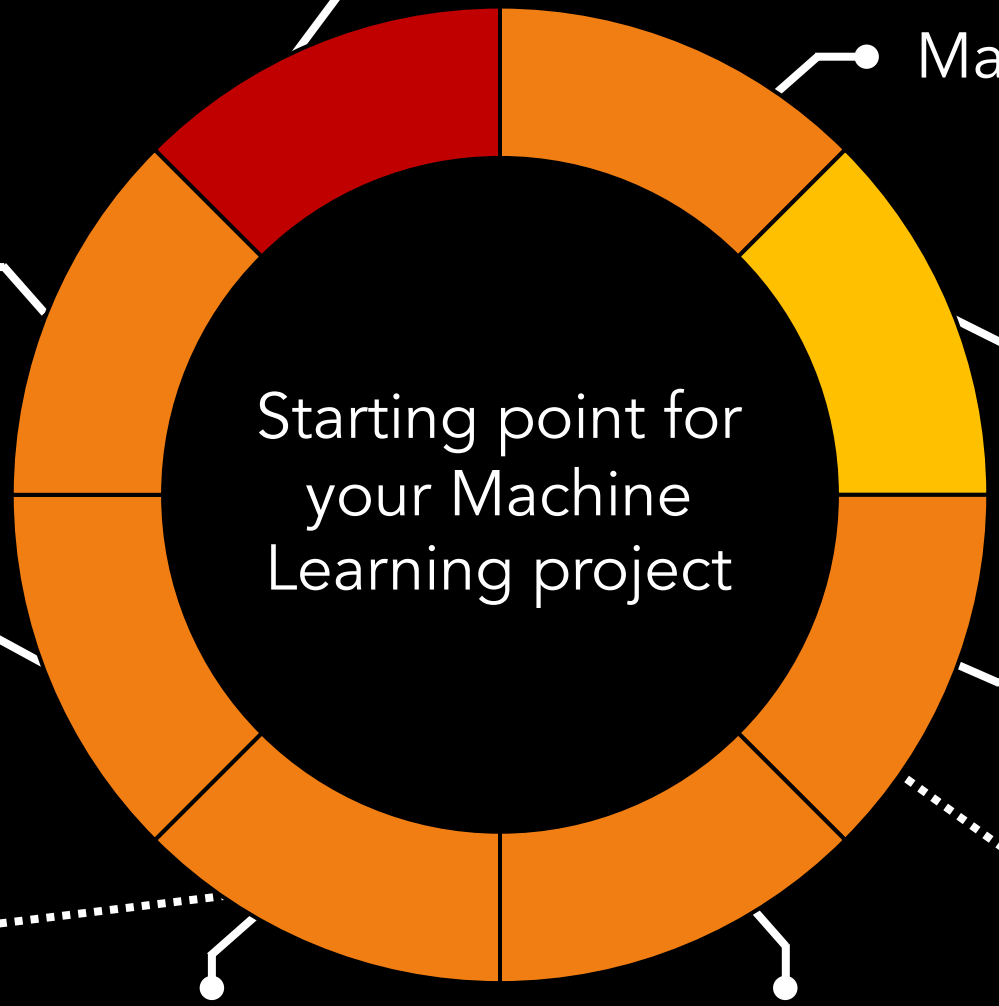
Dr. Maxim Samarin

Senior Data Scientist @ Swiss Data Science Center

29th / 30th April and 6th / 7th May 2024

Please:  for class;  for questions

Overview



Assignments (optional)

Machine Learning Introduction

Python Concepts

Regression & Classification

Today ↑

Dimensionality Reduction & Clustering

Kernel Methods

Tomorrow →

↑ Next Week

Neural Networks

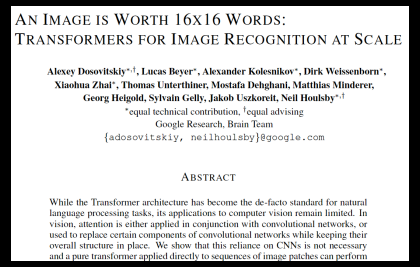
More on Machine Learning / Q&A

Practical Exercises

- End-of-notebook exercises → breakout rooms of 2-3 people
  - Goal: Understanding what is required to use different tools
  - Comparison of different methods for the same regression problem

- Optional assignments for more practice
  - 1 ECTS for completing all tasks fully and reasonably
  - Deadline: 31st of May

1. Parsing a PDF in Python + generators + decorators



AN IMAGE IS WORTH 16X16 WORDS:
TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

Alexey Dosovitskiy*,†, Lucas Beyer*, Alexander Kolesnikov*, Dirk Weissenborn*, Xiaohua Zhai*, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby*,†

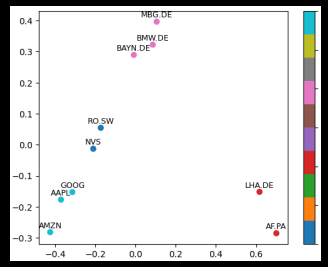
*equal technical contribution, †equal advising
Google Research, Brain Team
{adosovitskiy, neilhoulsby}@google.com

ABSTRACT

While the Transformer architecture has become the de-facto standard for natural language processing tasks, its applications to computer vision remain limited. In vision, attention is either applied in conjunction with convolutional networks, or used to replace certain components of convolutional networks while keeping their overall structure in place. We show that this reliance on CNNs is not necessary and a pure transformer applied directly to sequences of image patches can perform

2. Ear-shells data + ridge regression + decision trees + random forests

3. Finance data + PCA + k-means



4. Fashion images + neural networks

ETH zürich

ENS DE LYON

B.Sc. and M.Sc. in Physics

University of Basel

Ph.D. in Machine Learning

Research in Generative AI, Computer Vision, Deep Learning theory and various applications

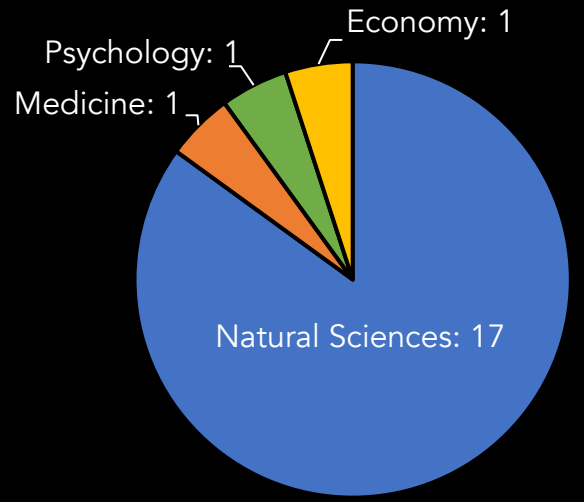
SDSC

Senior Data Scientist and ML / AI researcher

Projects on climate modelling and drug safety

Your Experience and Goal

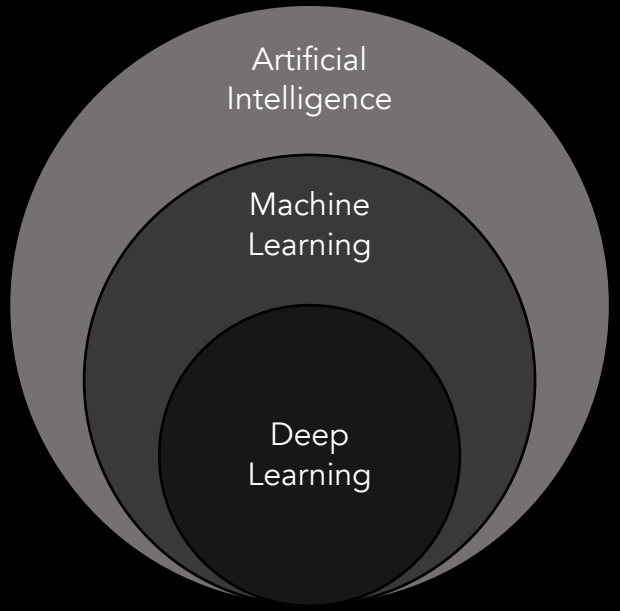
- What is your ML / programming experience so far and an intermediate goal for you?

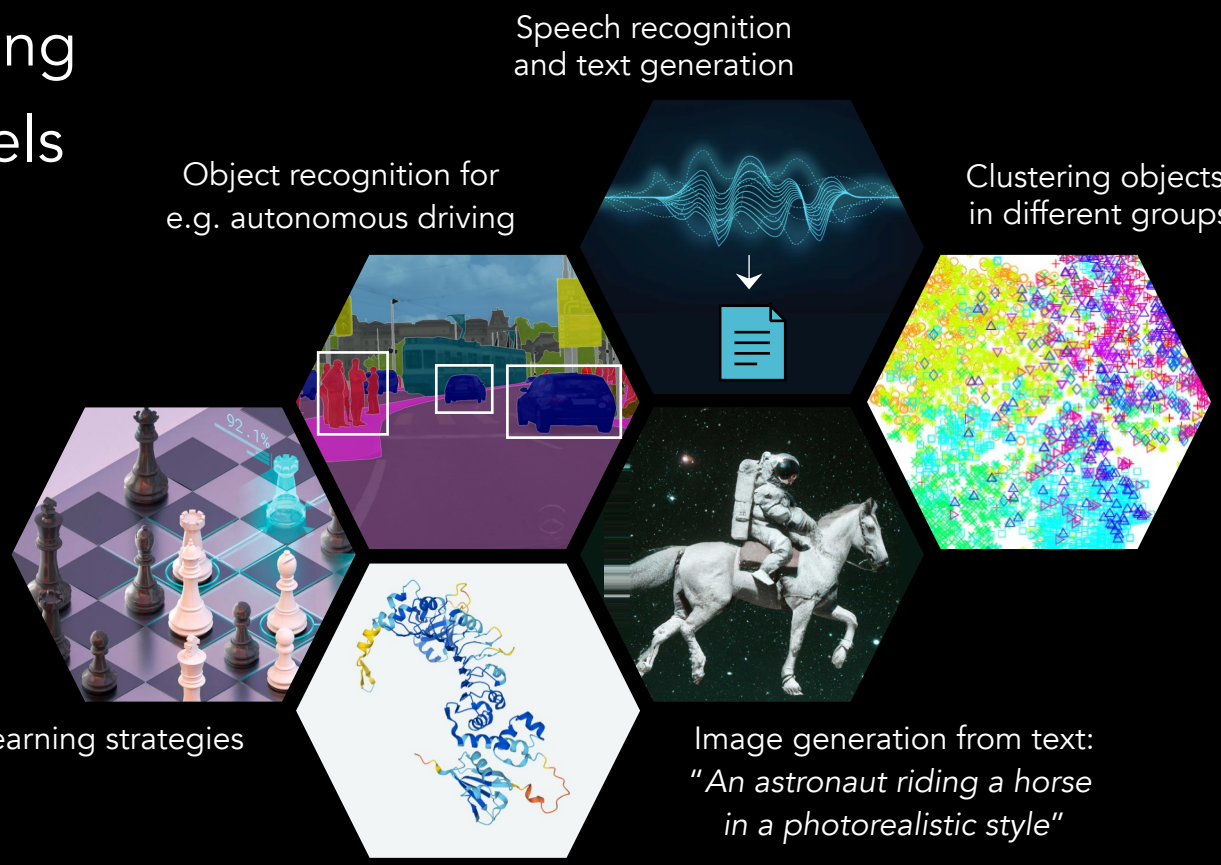


Samuel ⇢ Toni ⇢ Alexandra ⇢ Guido ⇢ Jakob ⇢ Jiami ⇢
⇢ Mattia ⇢ Maximilian ⇢ Lucca ⇢ Zarah ⇢ Raphael ⇢
⇢ Navish ⇢ Nila ⇢ Samuel ⇢ Fabrice ⇢ Paul ⇢
⇢ Danielle ⇢ Philipp ⇢ Bastian ⇢ Benedetta

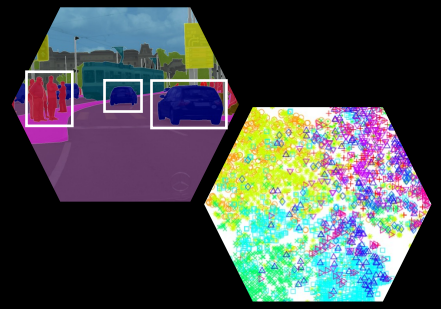
Machine Learning Landscape

- Supervised and unsupervised learning
- Discriminative and generative models
- Reinforcement learning

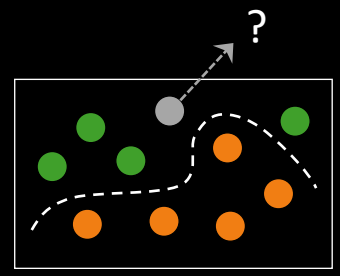




Supervised vs. Unsupervised

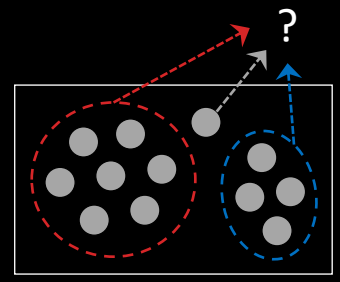


- Supervised learning: Labeled data guides learning
  - Classification: Class labels separate groups of data
  - Regression: Functional relationship between inputs (predictors) x and responses y $\rightarrow$ f(x) = y



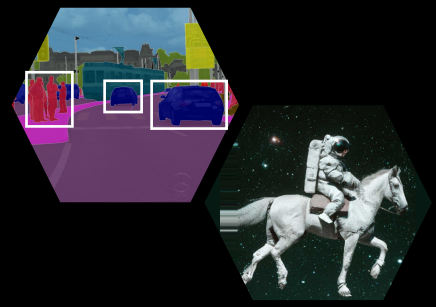
Supervised

- Unsupervised learning: No labels available
  - Clustering: Identifying clusters inherent to the data
  - Dimensionality reduction: Identifying relevant predictors / dimensions or factors of variation in data

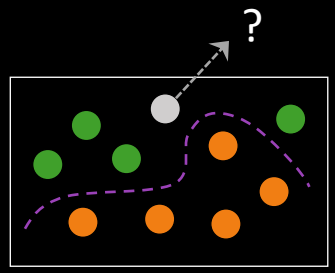


Unsupervised

Discriminative vs. Generative

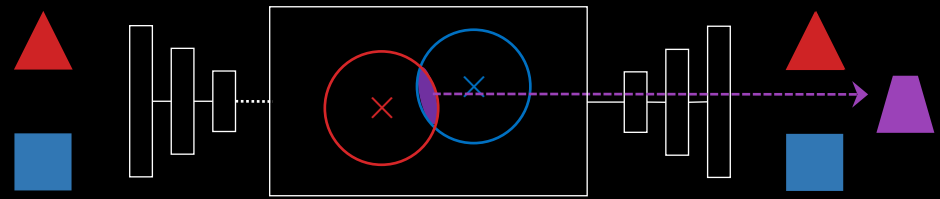


- Discriminative models: Learn decision boundary to differentiate data and provide accurate predictions to new, unseen data



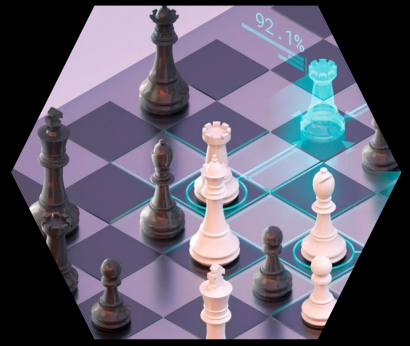
Discriminative

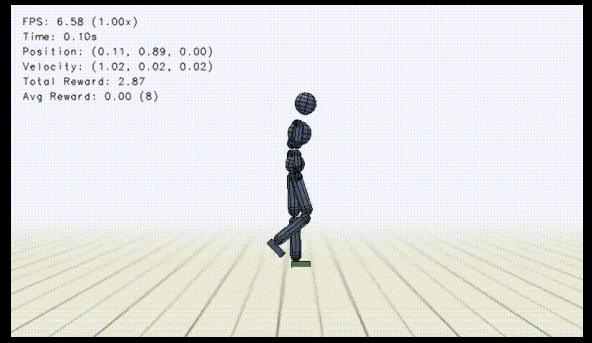
- Generative models: Approximate data distribution to provide accurate predictions but also allow generating artificial data similar to training data


Generative

Reinforcement Learning

- Agents explore environment, select actions, receive rewards, and form strategies / policies → exploration and exploitation
- Modelling (Markov) decision processes
- Challenging in (large) uncertain environments
- Application in games, robotic, and finance





Source: G. Berseth and C. Pal

Start your Working Environment

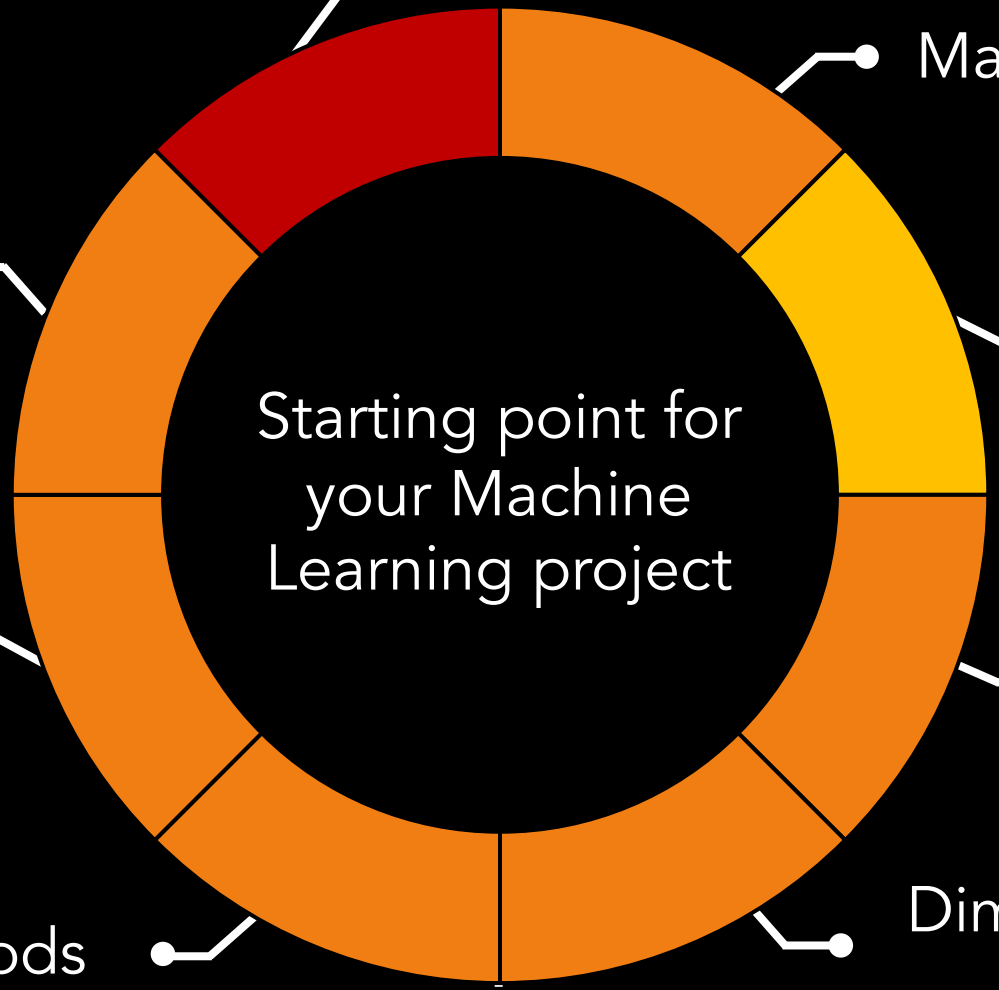
1. Access the course environment via Noto link provided in E-Mail and sign in with your SWITCH edu-ID ("*Use your Switch AAI login*")
2. Or: Download new material, start Jupyter Lab and open notebooks

- Suggestions on how to work in this course:

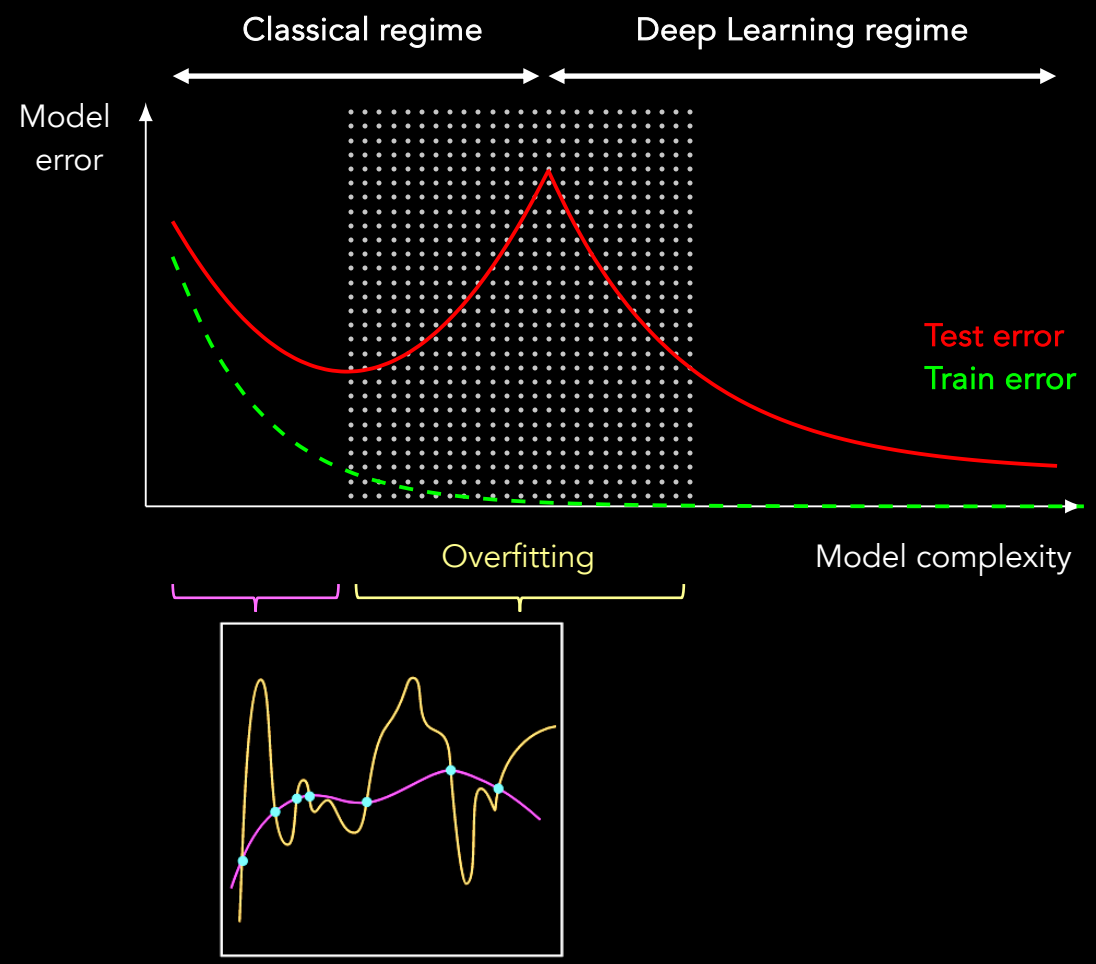
: Follow presentation, while executing scripts yourself, making adjustments and notes in your own notebook

: Follow presentation, switch to programming environment for exercises

Overview


Assignments (optional)
Machine Learning Introduction
Python Concepts
Regression & Classification
Dimensionality Reduction & Clustering
Kernel Methods
Neural Networks
More on Machine Learning / Q&A
Starting point for your Machine Learning project
← This week
Last week →

Bias-Variance Trade-off



- Flexible models are prone to overfitting
- Training / validation / test splits to ensure generalisation to new data
- Watch out for imbalance in training data

Machine Learning Topics

Regression & Classification	Dim. Reduction & Clustering	Kernel Methods	Neural Networks
Linear / Ridge / Logistic Regression	Principal Component Analysis	Kernel Ridge Regression	Feed-Forward / Convolutional Neural Networks
Decision Trees & Random Forests	k-Means	Gaussian Process	Variational Autoencoder
k-Nearest Neighbors	Gaussian Mixture Models	Kernel Density Estimation	Generative Adversarial Networks
Support Vector Machines	Factor Analysis	Kernel PCA	Transformer
Naïve Bayes	Independent Component Analysis	Kernel SVM	Diffusion Probabilistic Models
…	…	Kernel …	…

How to Continue with ML & Python

- Optional assignments: Deadline 31st of May, 1 ECTS for full completion

- Machine Learning scikit-learn tutorials
  https://scikit-learn.org/stable/tutorial/index.html
- TensorFlow tutorials
  https://www.tensorflow.org/resources/learn-ml
- Quick overview with interactive tutorials on some basic topics and more advanced concepts:
  https://www.learnpython.org/
- Great tutorials on specific (advanced) topics, easy-to-read books:
  https://www.realpython.com/

How to Continue Programming

- Use an integrated development environment (IDE) like
  - PyCharm:
    https://www.jetbrains.com/pycharm/
  - Visual Studio Code:
    https://code.visualstudio.com/

- Use high-performance computing (GPUs, TBs of RAM) cluster sciCORE
  - Getting started:
    https://scicore.unibas.ch/using-scicore/getting-started/
  - Courses:
    https://scicore.unibas.ch/events/teaching-training/

Suggestions for the Feedback

- Preparation information / YouTube videos adequate?
- What did / didn't you like about Noto?
- Content appropriate: anything missing for you (Python and Machine Learning)?
- Too fast or slow, shallow or deep?
- Insightful exercises and enough time?

Thank you and good luck on your Machine Learning journey! ☺