Machine Learning Course Workbook

– Before the Course –

## Introduction

### ML is everywhere!

#### Where (else) do you use ML in your everyday life incl. work?

### ML history: Why now?

#### What accelerated the rise of ML in the last few years?



#### What is the difference between ANI and AGI?

– Day 1 (Part 1) –

## The Basics

### Data is the new oil!?

#### What does Goodhart’s Law warn us about?

#### With what KPI could your department’s performance be quantified? What would be the target state, alert threshold, and what could be possible corrective actions?

### What is ML?

#### What is the difference between Machine Learning, Artificial Intelligence, and Deep Learning?

#### What are the benefits of ML compared to traditional software?

#### When should you not use ML?

#### Which kind of ML problems have a high chance of success and when is the outcome uncertain?

### How do machines “learn”?

#### Describe the different learning strategies and what their requirements (in terms of data) are:

* Unsupervised Learning:
* Supervised Learning:
* Reinforcement Learning:

#### What are “features” and what are “labels”?

* Features:
* Labels:

### ML use cases

#### What does structured and unstructured data look like?

* Structured Data:
* Unstructured Data:

#### What is the drawback of unsupervised learning methods?

#### What are the benefits of breaking down a complex input-output problem into simpler subproblems?

#### What is the downside of a system composed of multiple ML models?

### Solving problems with ML

#### Which tasks take up most of a Data Scientist’s time?

### 

#### What are the two deployment options for an ML model and when should you use which?

– Day 1 (Part 2) –

#### Take another look at the [ML algorithm cheat sheet](https://franziskahorn.de/mlws_resources/algorithm_cheatsheet.pdf) & try to find examples where you could be using each of these algorithms to improve your organization’s products or processes.

* Anomaly Detection:
* Clustering:
* Regression:
* Classification:
* Deep Learning:
* Recommender Systems/Information Retrieval:

## Data Analysis & Preprocessing

### Data Analysis

#### You want to pick a restaurant for dinner. Your data source is Google Maps. What information do you consider when making a decision and what makes you choose one restaurant over another?

### Garbage in, garbage out!

#### What do you think are the most common ways in which datasets in your organization are messy?

#### Which concrete next steps could your organization take to improve their data quality?

## – Day 2 –

## Avoiding Common Pitfalls

#### With which stupid baseline should you compare regression and classification models respectively?

#### When is it a really bad idea to evaluate a classification model with the accuracy metric?

#### What does it mean for a model to over- or underfit?

#### Why can a model still be wrong, even though it generates correct predictions for data points from the testset?

#### What are “Adversarial Attacks”?

#### In what ways can a biased model negatively affect users?

#### How can you check whether a model discriminates?

#### What is the difference between data and concept drift?

#### What could be reasons for data or concept drift in your domain / next project?

## Conclusion

#### According to Andrew Ng, what are the 5 steps for a successful AI transformation of a company and where do you think your organization stands in this process?