Machine Learning Course Workbook

– Before the Course –

## Introduction

### ML is everywhere!

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

– Part 1 –

### ML history: Why now?

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



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

### What is ML?

#### Define ML:

#### What do you need to create an ML-powered product (i.e., value)?



#### AI and ML Researchers, Statisticians, and Data Scientists all use a certain set of tools. What is the difference between...

* ML vs. AI?
* ML vs. Deep Learning?
* ML vs. Statistics?
* ML vs. Data Science?

### How do machines “learn”?

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

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

### When should you use ML?

#### In what ways can ML create value?

#### When should you not use ML?

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

### Solving problems with ML: Workflow

#### What are the 3 main steps to create value with ML?



#### What should you check before starting an ML project?



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



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

## Data & Preprocessing

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

* Features:
* Labels:

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

* Structured Data:
* Unstructured Data:

### 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?

## ML Solutions: Overview

#### What does the output of the different algorithm categories look like for one data point?

* Dimensionality Reduction:
* Anomaly Detection:
* Clustering:
* Regression:
* Classification:
* Recommender Systems/Information Retrieval:

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

## 

## – Part 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?



#### 

#### Where do you think your organization stands in this AI transformation process?