

# Machine Learning Course Workbook

## 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?*

### **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:*

- Unsupervised Learning:
- Supervised Learning:
- Reinforcement Learning:

## **When should you use ML?**

When should you not use ML?

*For which kinds of problems does ML have a high chance of success and when is the outcome uncertain?*

*What distinguishes an ML project from a data science project (in terms of deliverables)?*

*In what ways can you create value with ML?*

## **Solving problems with ML: Workflow**

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

- 1.
- 2.
- 3.

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

- 
- 
- 
- 

*What are the two main deployment possibilities 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:
- Classification:
- Regression:

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

## Supervised Learning Basics

*What is the difference between a regression and a classification problem?*

*What is the stupid baseline you should always compare a regression model against?*

*What is the stupid baseline you should always compare a classification model against?*

## Avoiding Common Pitfalls

*When should you absolutely evaluate a model with the balanced accuracy metric?*

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

*Why might a model still be wrong, even if it generates correct predictions for new data points?*

*What are “Adversarial Attacks”?*

*Why can it happen that a model discriminates?*

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

## Conclusion

*According to Andrew Ng, what are the 5 steps for a successful AI transformation of a company?*

- 1.
- 2.
- 3.
- 4.
- 5.

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