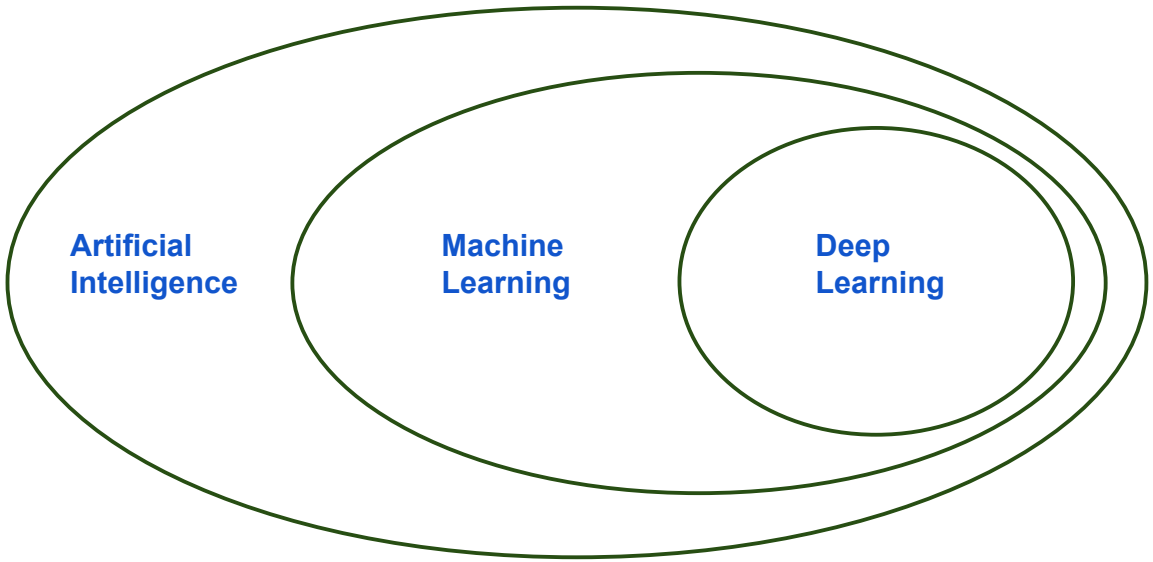
Introduction to Machine Learning

Introduction: What is Machine Learning?

compstat-lmu.github.io/lecture_i2ml

MACHINE LEARNING

Machine learning is a branch of statistics and computer science.



A computer program is said to **learn** from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Tom Mitchell, Carnegie Mellon University, 1998

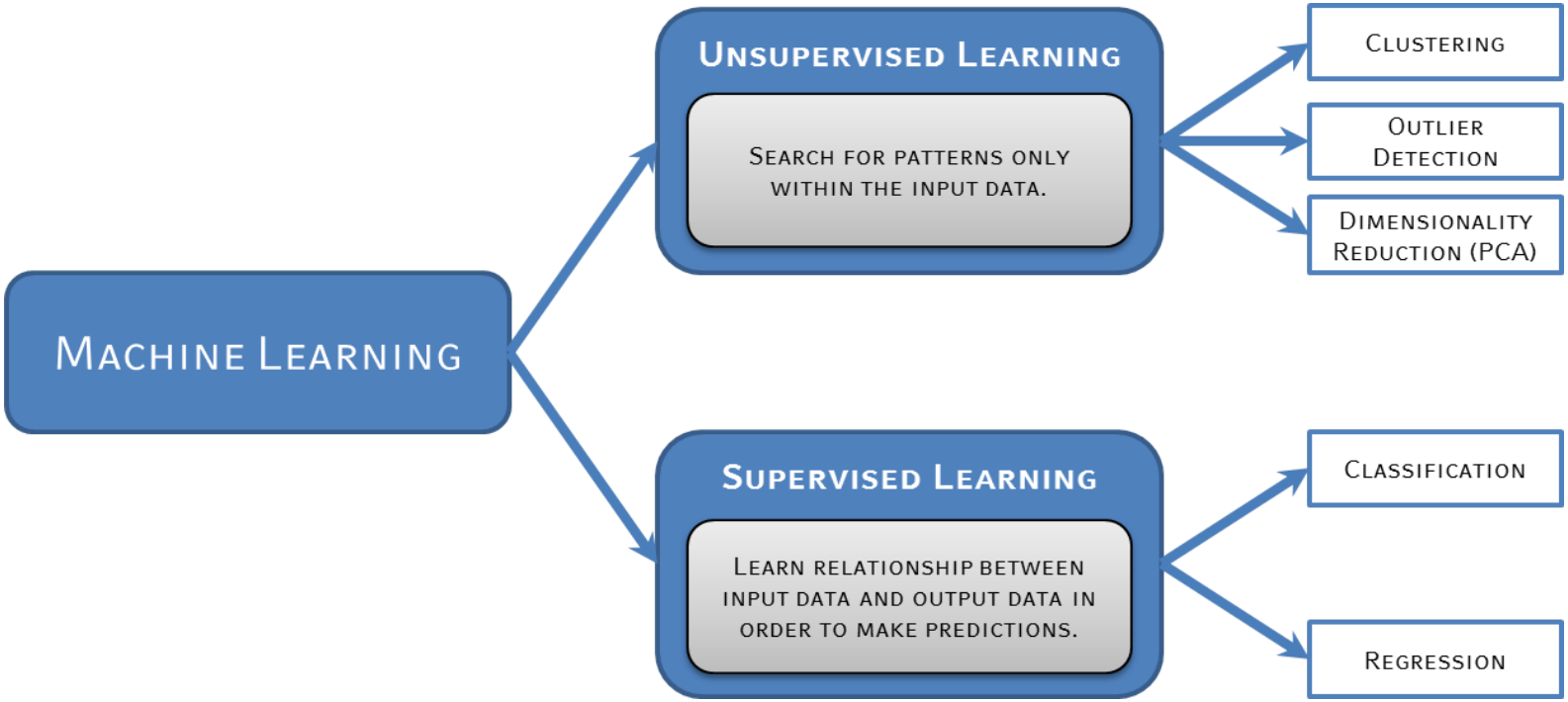
MACHINE LEARNING IS CHANGING OUR WORLD

- Search engines learn what you want
- Recommender systems learn your taste in books, music, movies,...
- Algorithms do automatic stock trading
- Google Translate learns how to translate text
- Siri learns to understand speech
- DeepMind beats humans at Go
- Cars drive themselves
- Smart-watches monitor your health
- Election campaigns use algorithmically targeted ads to influence voters
- Data-driven discoveries are made in Physics, Biology, Genetics, Astronomy, Chemistry, Neurology,...
- ...

COMING UP

- In this course, we focus on so-called supervised ML, in a nutshell: using ML to predict something.
- In the first chapters, we will go through the fundamental terminology and concepts in supervised ML which are relevant for everything that comes next:
  - What kind of "data" do we learn from?
  - How can we formalize the goal of learning?
  - What is a "prediction model"?
  - How can we quantify "predictive performance"?
  - What is a "learning algorithm" and how can we operationalize learning?
- We will also look at a couple of fairly simple ML models to obtain a basic understanding and look at some concrete examples.
- More complex stuff comes later.

MACHINE LEARNING TASKS



In this course, we will deal with **supervised learning** for regression and classification only: predicting labels y based on features x, using pattterns that we learned from labeled training data.

ADDITIONAL LEARNING TASKS

Unsupervised learning

- Data without labels y
- Search for patterns within the inputs x
- *Unsupervised* as there is no external criterion to optimize or "true" output
  - Dimensionality reduction (PCA, Autoencoders ...): compress information in $\mathcal{X}$
  - Clustering: group similar observations, separate dissimilar observations
  - Outlier detection, anomaly detection
  - Association rules

ADDITIONAL LEARNING TASKS

Semi-Supervised learning

- Large amount of labeled data necessary to train reliable model
- Creating labeled datasets often very expensive
- Learn from labeled (expensive) **and** unlabeled (cheap) data
- Unlabeled data in conjunction with a small amount of labeled data improves learning accuracy
- FIGURE R course repository –> Tobi

ADDITIONAL LEARNING TASKS

Reinforcement learning

- Select actions in subsequent states within a certain environment to maximize lagged future reward
- Example: train neural net to play mario kart (environment)
  - Accelerate/ steer/ break (actions) at each time point (states) during playing
  - Reward: ranking after finish, should be maximized
- FIGURE R course repository –> Tobi