

# \*\*\* Applied Machine Learning Fundamentals \*\*\*

## Machine Learning Introduction

Daniel Wehner

SAP SE

August 6, 2019



# Agenda August 6, 2019

## ① General Overview

## ② Problem Types in Machine Learning

- Type of Training Information
- Availability of Training Examples
- Type of Target Variable

## ③ Key Challenges in Machine Learning

## ④ Lecture Overview

## ⑤ Wrap-Up

- Summary
- Self-Test Questions
- Recommended Literature and further Reading

Section:  
**General Overview**



# Why Machine Learning?

- *'We are drowning in information and starving for knowledge.'*  
– *John Naisbitt*
- **Era of big data:**
  - In 2017 there are about **1.8 trillion** web-pages on the internet
  - **20 hours** of video are uploaded to YouTube every minute
  - Walmart handles more than **1 million** transactions per hour and has data bases containing more than **2.5 peta-bytes** ( $2.5 \times 10^{15}$ ) of information
- **No human being can deal with this data avalanche!**

## Why Machine Learning? (Ctd.)

*'I keep saying the sexy job in the next ten years will be **statisticians** and **machine learners**. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s? The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it - that's going to be a hugely important skill in the next decades.'*

*– **Hal Varian**, Chief Economist at Google, 2009*

# Definition of Machine Learning

- *'[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.'*

– **Arthur Samuel, 1959**

- *'A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .'*

– **Tom Mitchell, 1997**

## A more abstract Definition

- Our task is to learn a mapping from input to output:

$$h : \mathcal{I} \mapsto \mathcal{O}$$

- Put differently, we want to predict the output from the input:

$$y = h(\mathbf{x}; \boldsymbol{\theta}) \quad \text{also: } y = h_{\boldsymbol{\theta}}(\mathbf{x})$$

- $\mathbf{x} \in \mathcal{I}$  (Input)
- $y \in \mathcal{O}$  (Output)
- $\boldsymbol{\theta} \in \boldsymbol{\Theta}$  (Parameters: What needs to be 'learned')

Section:

## Problem Types in Machine Learning

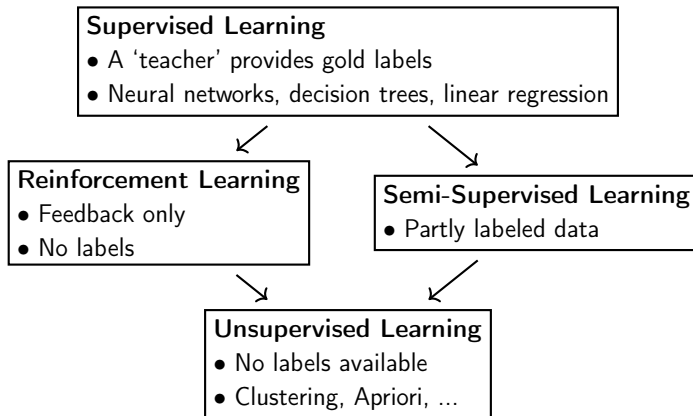




# Type of Training Information

- **Supervised learning**
  - 'Teacher' provides **gold labels**
  - E. g. neural networks, decision trees, linear regression
- **Unsupervised learning**
  - Labels are **not** known during training
  - E. g. clustering, density estimation, association rule mining
- **Reinforcement learning**
  - Environment provides rewards for actions but correct action is unknown
  - E. g. policy-iteration, Q-learning, SARSA
- **Semi-supervised learning** (Instances partly labeled)

# Type of Training Information (Ctd.)

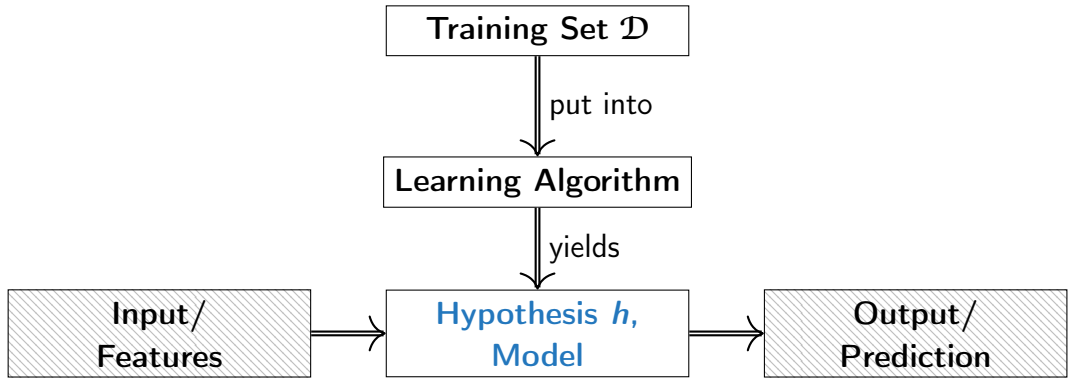


# Supervised Learning

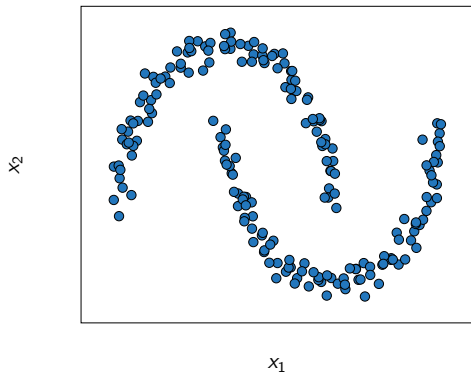
- A single row is called **example**
- An example without class label is called **instance**
- **Predictors:**
  - Outlook  $\in \{\textit{sunny}, \textit{overcast}, \textit{rainy}\}$
  - Temperature  $\in \{\textit{hot}, \textit{mild}, \textit{cool}\}$
  - Humidity  $\in \{\textit{high}, \textit{normal}\}$
  - Wind  $\in \{\textit{weak}, \textit{strong}\}$
- **Label:**
  - PlayGolf  $\in \{\textit{yes}, \textit{no}\}$
  - Given a new instance we want to predict the label
- **Label for the new instance???**

Outlook	Temperature	Humidity	Wind	PlayGolf
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rainy	mild	high	weak	yes
rainy	cool	normal	weak	yes
rainy	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rainy	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rainy	mild	high	strong	no
rainy	mild	normal	strong	???

# Supervised Learning: General Approach



# Unsupervised Learning



- There are **no** labels
- Try to find regularities in the data
- Examples for unsupervised learning:
  - Clustering
  - Density estimation
  - Dimensionality reduction

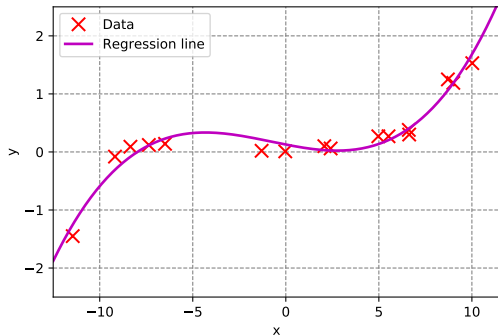
# Availability of Training Examples

- **Batch Learning**
  - The learner is provided with a fixed set of training examples
  - See weather data set
  - E. g. neural networks, decision trees
- **Incremental/Online Learning**
  - Constant stream of training examples
  - The model is updated as new training examples arrive
  - E. g. k-nearest-neighbors
- Active Learning (*not covered*)

# Type of Target Variable: Regression

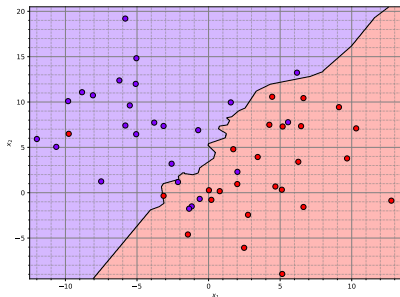
## Regression

- Learn a mapping into a continuous space
  - $\mathcal{O} = \mathbb{R}$
  - $\mathcal{O} = \mathbb{R}^3$
- E. g. curve fitting, financial analysis, housing prices, ...



# Type of Target Variable: Classification

## Classification



- Learn a mapping into a discrete space, e. g.
  - $\mathcal{O} = \{0, 1\}$  (binary classification)
  - $\mathcal{O} = \{0, 1, 2, 3, \dots\}$
  - $\mathcal{O} = \{verb, noun, adverb, \dots\}$
- Examples:
  - Spam / no spam
  - Digit recognition
  - Part of speech tagging



Section:

## Key Challenges in Machine Learning





Section:  
Lecture Overview



# Lecture Overview

## Unit I: Machine Learning Introduction

# Summary

# Self-Test Questions

# Recommended Literature and further Reading



## [1] Machine Learning

*Tom Mitchell. McGraw-Hill Science. 1997.*

→ See chapter 1 (Introduction)

# Thank you very much for the attention!

**Topic:** \*\*\* Applied Machine Learning Fundamentals \*\*\* Machine Learning Introduction

**Date:** August 6, 2019

**Contact:**

Daniel Wehner (D062271)

SAP SE

[daniel.wehner@sap.com](mailto:daniel.wehner@sap.com)

## Do you have any questions?