universität wien

Introduction to Machine Learning

Introduction

Nils M. Kriege

WS 2023

Data Mining and Machine Learning
Faculty of Computer Science
University of Vienna

Credit: Slides based on the IML Lectures by Sebastian Tschiatschek and Andreas Krause

About Myself

2023 Associate Professor

2020 Assistant Professor
University of Vienna

2019 Interim Professor, *TU Dortmund*

2015/16/18 Visiting Researcher
University of York

2015 PhD in Computer Science
TU Dortmund

About Myself

2023 Associate Professor

2020 Assistant Professor
University of Vienna

2019 Interim Professor, *TU Dortmund*

2015/16/18 Visiting Researcher
University of York

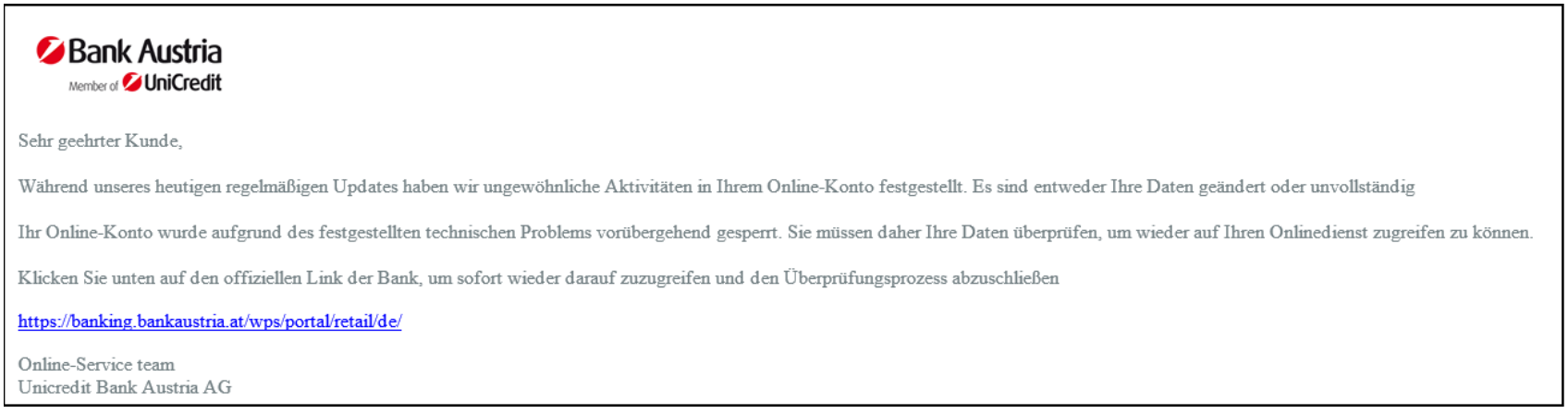
2015 PhD in Computer Science
TU Dortmund

Research Interests

- Data Mining, Machine Learning with Graphs
- Combinatorial Optimization, Graph Algorithms
- Applications in Cheminformatics

What is Machine Learning I: An Example

Classify email message as “spam” or “non spam”



Bank Austria
Member of UniCredit

Sehr geehrter Kunde,

Während unseres heutigen regelmäßigen Updates haben wir ungewöhnliche Aktivitäten in Ihrem Online-Konto festgestellt. Es sind entweder Ihre Daten geändert oder unvollständig

Ihr Online-Konto wurde aufgrund des festgestellten technischen Problems vorübergehend gesperrt. Sie müssen daher Ihre Daten überprüfen, um wieder auf Ihren Onlinedienst zugreifen zu können.

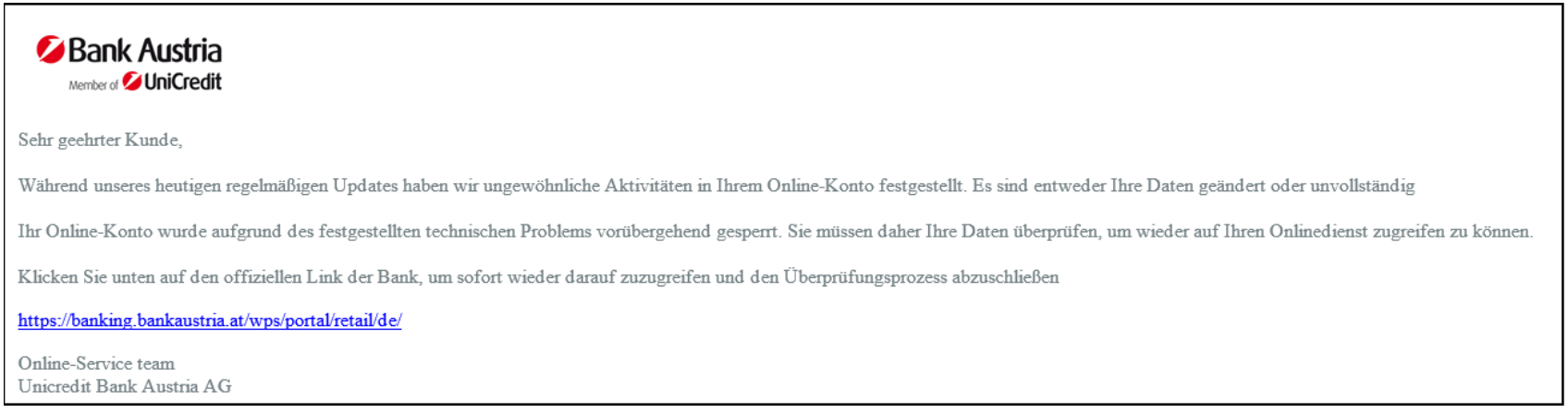
Klicken Sie unten auf den offiziellen Link der Bank, um sofort wieder darauf zuzugreifen und den Überprüfungsprozess abzuschließen

https://banking.bankaustria.at/wps/portal/retail/de/

Online-Service team
Unicredit Bank Austria AG

What is Machine Learning I: An Example

Classify email message as “spam” or “non spam”



Bank Austria
Member of UniCredit

Sehr geehrter Kunde,

Während unseres heutigen regelmäßigen Updates haben wir ungewöhnliche Aktivitäten in Ihrem Online-Konto festgestellt. Es sind entweder Ihre Daten geändert oder unvollständig

Ihr Online-Konto wurde aufgrund des festgestellten technischen Problems vorübergehend gesperrt. Sie müssen daher Ihre Daten überprüfen, um wieder auf Ihren Onlinedienst zugreifen zu können.

Klicken Sie unten auf den offiziellen Link der Bank, um sofort wieder darauf zuzugreifen und den Überprüfungsprozess abzuschließen

https://banking.bankaustria.at/wps/portal/retail/de/

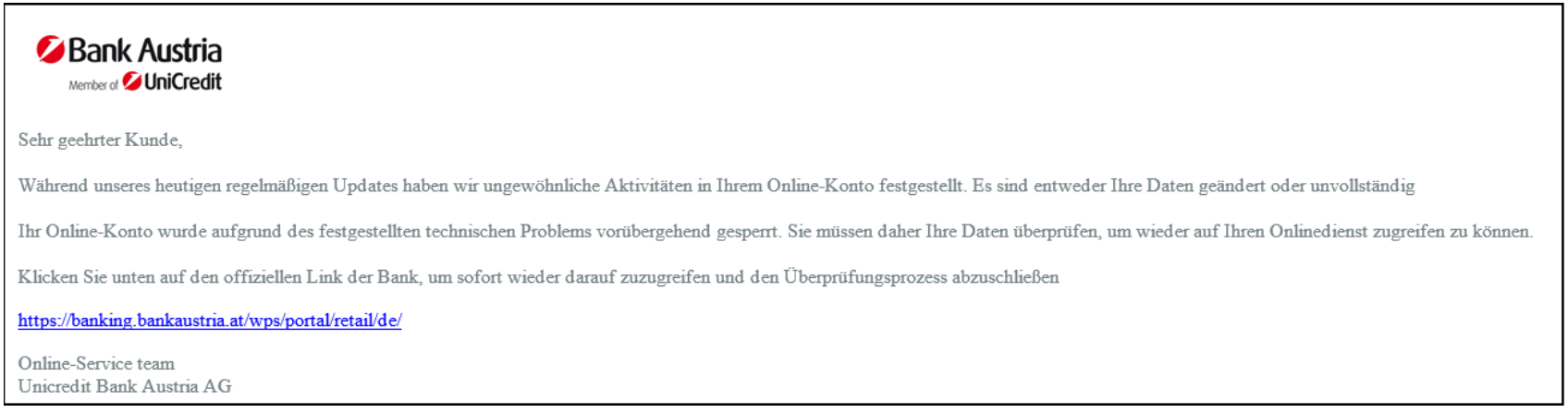
Online-Service team
Unicredit Bank Austria AG

Classical approach: manual rules

```
IF text body contains "Klicken Sie unten auf den
offiziellen Link"
THEN classify as "spam" ELSE "non spam"
```

What is Machine Learning I: An Example

Classify email message as “spam” or “non spam”



Bank Austria
Member of UniCredit

Sehr geehrter Kunde,

Während unseres heutigen regelmäßigen Updates haben wir ungewöhnliche Aktivitäten in Ihrem Online-Konto festgestellt. Es sind entweder Ihre Daten geändert oder unvollständig

Ihr Online-Konto wurde aufgrund des festgestellten technischen Problems vorübergehend gesperrt. Sie müssen daher Ihre Daten überprüfen, um wieder auf Ihren Onlinedienst zugreifen zu können.

Klicken Sie unten auf den offiziellen Link der Bank, um sofort wieder darauf zuzugreifen und den Überprüfungsprozess abzuschließen

https://banking.bankaustria.at/wps/portal/retail/de/

Online-Service team
Unicredit Bank Austria AG

Classical approach: manual rules

```
IF text body contains "Klicken Sie unten auf den
offiziellen Link"
THEN classify as "spam" ELSE "non spam"
```

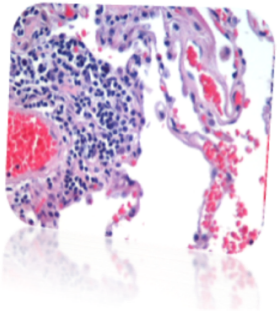
Machine Learning: Automatic discovery of rules from training data (examples)

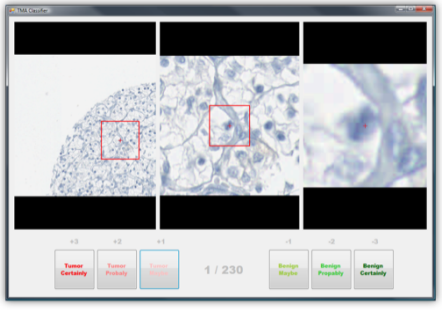
What is Machine Learning II: One Definition

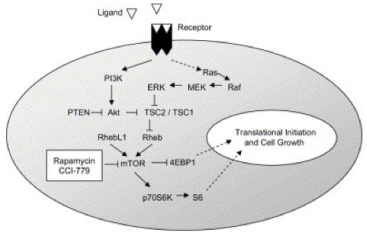
Tom Mitchell gave the following definition:

> "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 task *T*, as measured by *P*, improves with the experience *E*."

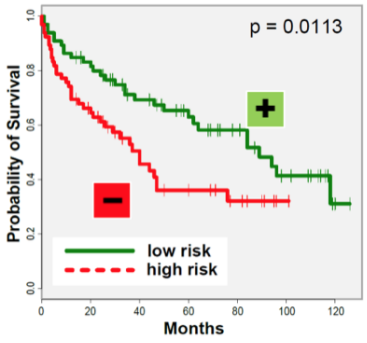
Our Digital Society and the Information Technology Value Chain







Activation of the mTOR Signaling Pathway in Renal Clear Cell Carcinoma. Robb et al., J Urology 177:346 (2007)



Data ⟶ Information ⟶ Knowledge ⟶ Value

Machine Learning plays a core role in this value chain!

Related Disciplinies

Information Theory

Statistics

Machine Learning

Philosophy
Epistemiology
Causality

Algorithms & Optimization

Neuro-informatics

Overview

- Introductory course
- Preparation for advanced ML courses and ML work
- Two main topics
  - Supervised learning
  - Unsupervised learning
- Algorithms, models & applications
- Lecture slides, pen & paper exercises, programming assignments etc. on course webpage (Moodle)
- Textbooks listed on course webpage (some available online)

Prerequisites

- Basic knowledge in linear algebra, calculus and probability
- If you need a refresher:
  - Part I of *Mathematics for Machine Learning* by Deisenroth, Faisal, Ong
  - Available online at `https://mml-book.com/`
- Basic programming (in Python)

Everyone familiar with Python?

Syllabus

- Linear regression
- Linear classification
- Kernels and the kernel trick
- Neural networks & deep learning
- Unsupervised learning
- The statistical perspective
- Statistical decision theory
- Discriminative vs. generative modeling
- Bayes' classifiers
- Bayesian approaches to unsupervised learning

After participating in this course you will

- Understand basic machine learning ideas & concepts
- Be able to apply basic machine learning algorithms
- Know how to validate the output of a learning method
- Have some experience using machine learning on real data
- Learn what role machine learning plays in decision making under uncertainty

Team

- Head

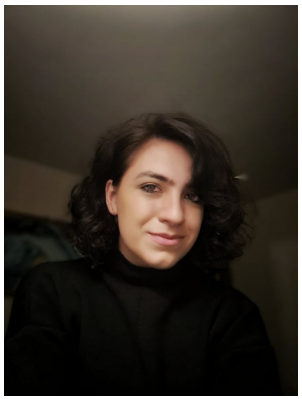
  Sebastian Tschiatschek
- Main instructor

  Nils Kriege (nils.kriege@univie.ac.at)
- Instructors/Tutors

Timo Klein
(timo.klein@univie.ac.at)

Simon Rittel
(simon.rittel@univie.ac.at)

Megi Teneqexhi
(megi.teneqexhi@univie.ac.at)



Lectures

- Lecture times:
  - Monday, 15:00-16:30
  - Friday, 13:15-14:45
- Attendance of the lecture is not mandatory, but recommended
- Previous year's lecture recordings by Prof. Tschiatschek will be made available via Moodle
- Attendance of exercise discussions is mandatory to get points

Exercises (Pen & Paper)

- The exercise discussions are on-site
- Exercises $\sim$ every 2nd Monday in the above mentioned time slots (check schedule on Moodle).
- There will be 7 pen & paper exercises
- Need to mark tasks you have solved on Moodle (Deadline 6 pm on the day before the exercise session)
- Present your solution in the exercise session (if selected)
- Sample solutions will be published on Moodle
- Attendance is mandatory to get points
- Limited number of absences allowed: You have to submit your solutions via Moodle **and** send an email stating a good reason for not attending

Programming Assignment

- There will be 6 programming assignments
- You will implement and apply basic machine learning methods to make predictions on real data
- Solve given tasks **on your own** (coding in Python using numpy, scikit-learn, . . .)
- Upload your (executable) code and write a short report
- First two assignments will be solely graded by us
- Peer-review for 3rd to 5th programming assignment
- Tentative schedule on Moodle

Questions

- Main resource: Forum available in Moodle
- During the lecture (“raise your hand”)
- Via email (only if Moodle does not work)
- If needed, we will do extra office hours

Feedback

- Feedback is very welcome!

Grading

- 30% Written exam [January 29th, 2024 (tentative)]
- 40% Programming assignments
- 30% Pen & paper exercises
- Let p be the weighted average percentage you achieved on the written exam, the pen-and-paper exercises, and the programming exercises
- If you achieved at least 50% on the written exam **AND** at least 50% on the programming assignments **AND** at least 50% on the pen & paper exercises, your grade is

$$\text{grade} = \begin{cases} \text{sehr gut} & 90\% \leq p \\ \text{gut} & 77\% \leq p < 90\% \\ \text{befriedigend} & 62\% \leq p < 77\% \\ \text{genügend} & 50\% \leq p < 62\% \\ \text{nicht genügend} & p < 50\% \end{cases}$$

ECTS Breakdown

- IML is a 6 ECTS course
- $\Rightarrow \approx$ 150 hours of work:

Lecture	31.5 h	(21 lectures with 90 minutes each)
Exam	15 h	(90 minutes exam + 13.5 hours preparation)
Pen & Paper	38.5 h	(10.5 hours exercise sessions + 4 hours of work per sheet)
Programming	65 h	(about 11 hours per assignment)
	150 hours	