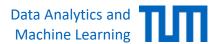
Machine Learning for Graphs and Sequential Data

Introduction

Lecturer: Prof. Dr. Stephan Günnemann

cs.cit.tum.de/daml

Summer Term 2024



Roadmap

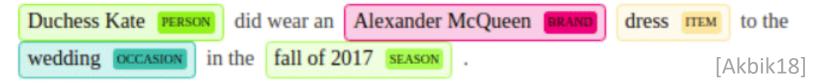
- Chapter: Introduction
 - 1. What will you learn in this lecture?
 - 2. Organizational aspects + project tasks

What is this course about?

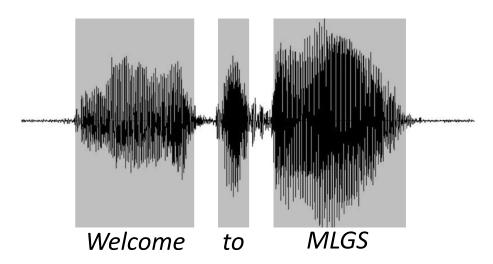
- In short: A continuation of our intro ML lecture (IN2064) now focusing on advanced learning principles and covering more complex data domains
- Focus on algorithms and general principles, not limited to a single domain
- Project tasks will give you hands-on experience
- At the end you should also be able to extend existing techniques and adapt them to applied problems

How do you learn from sequences?

How can we understand written language?



How do you detect patterns in sound waves?

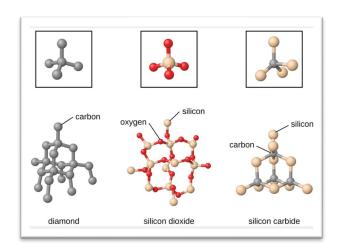


How can we predict the future?

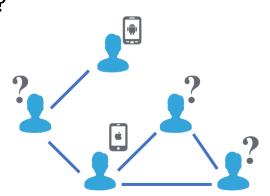


How to predict...?

- ...the properties of molecules?
 - instances are not vectors but graphs!



- ...the preference of a user in a social network?
 - instances are connected by a graph

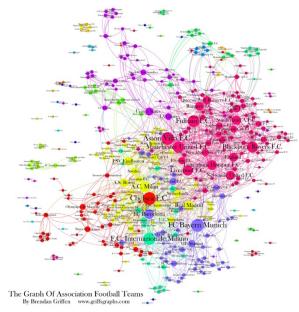


How to use/incorporate the graph structure for classification?

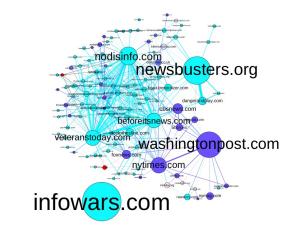
How to exploit graph/network data...?

... to find communities of people?





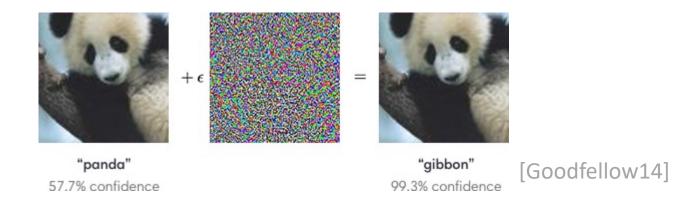
... to detect relevant websites?



← or the most favorite football clubs ©

Is Deep Learning fragile?

Neural networks can be tricked by small perturbations



- How can we find such a perturbation?
- How do you measure the susceptibility of a model to these attacks?
- Can we ensure the robustness of our models against such adversarial examples, e.g. through specialized training procedures?

The Challenge of Non-I.I.D. data

- Classical Machine Learning approaches assume data instances to be i.i.d.
 - Independent, identically distributed
- However, most real-world datasets show dependencies
 - Temporal data/sequences
 - Text: sequence of words (discrete values)
 - Sensor measurements: time series (continuous values)
 - Current value depends on previous values
 - Graphs/networks
 - Already seen before: molecules, social networks, knowledge graphs, connected devices (IoT)
 - Arbitrarily complex dependency structure possible

Lorem ipsum dolos
sit amet, consecte
adipiscing elit.
posuere tortor vitae
elit. Sed vitae metus a
elit bibendum malesuada
cras pulvinar. Quisque
pellentesque nibh in
sem. Curabitur ligula.
Suspendisse potenti.
Duis sit amet augue eu
arcu ultrices auctor.
Suspendisse elementum,
nunc ut molestie
elementum, neque augue
vulputate elit, eu
blandit enim velit
vitae nulla. Duis sed.



How to design machine learning techniques that consider these dependencies (and thus, hopefully, lead to better results)?

Contents of this Course

- Temporal & Sequential Data
 - Autoregressive models, HMM
 - Deep learning on sequences: RNN, LSTM, embeddings, Transformer
 - Temporal point processes
- Graphs
 - Laws & generative models
 - PageRank
 - Unsupervised representation learning: spectral clustering, embeddings
 - Supervised learning: label propagation, GNNs & their limitations
- Robustness
 - Adversarial examples
 - Attacks & defenses, certificates

Roadmap

- Chapter: Introduction
 - 1. What will you learn in this lecture?
 - 2. Organizational aspects + project tasks

Course Organization

- Lecturer
 - Prof. Dr. Stephan Günnemann
- Teaching assistants
 - Nicholas Gao, David Lüdke, Yan Scholten, Arthur Kosmala, Lukas Gosch
- 5 ECTS
- Language: English
- Ungraded exercise sheets
- Graded project tasks every 2-3 weeks
- Final exam + repeat exam
- Project tasks can grant a bonus of up to 0.3

Schedule

- In-person classroom (Timetable on Moodle)
 - Lecture and/or Exercise: Wednesday and Thursday 14:00 16:00
- Practice material and exercises uploaded to Moodle
 - Ungraded exercise sheets
- Project tasks via Artemis

Preliminary Timetable

Date		Lecture/Exercise	Projects
Mi	April 17, 2024	Introduction/AR MC	
Do	April 18, 2024	Autoregressive Models & Markov Chains	
Mi	April 24, 2024	Hidden Markov Models	
Do	April 25, 2024	Sequence Models	Project
Do	May 2, 2024	Neural Nets (Transformers)	
Mi	May 8, 2024	Exercise AR & MC & HMM	
Mi	May 15, 2024	Temporal Point Processes	Project
Do	May 16, 2024	Exercise Sequential	
Mi	May 22, 2024	Graphs - Laws, Patterns	
Do	May 23, 2024	Graphs - Clustering	
Mi	May 29, 2024	Exercise TPP	Project
Mi	<u>June 5, 2024</u>	Graphs - Embedding & Ranking	
Do	<u>June 6, 2024</u>	Graphs - Embedding & Ranking 2	
Mi	<u>June 12, 2024</u>	Exercise Graphs - Laws, Patterns	
Do	<u>June 13, 2024</u>	Graphs - Classifcation	Project
Mi	<u>June 19, 2024</u>	Exercise Graphs - Clustering, Embedding, Ranking	
Do	<u>June 20, 2024</u>	Graph Neural Networks 1	
Mi	<u>June 26, 2024</u>	Graph Neural Networks 2	
Do	<u>June 27, 2024</u>	Robustness Pt. 1	Project
Mi	July 3, 2024	Exercise Classification GNN all topics	
Do	<u>July 4, 2024</u>	Exercise Robustness 1	
Mi	July 10, 2024	Robustness Pt. 2 (GNN robustness)	
Do	July 11, 2024	Robustness Pt. 3 (GNN robustness)	Project
Mi	<u>July 17, 2024</u>	Exercise Robustness 2	
Do	July 18, 2024		

Prerequisites

 The course is designed for Master students of Computer Science (and specializations such as Data Engineering and Analytics, Games Engineering, etc.)

Prerequisites:

- Knowledge about the standard Machine Learning concepts (i.e. content of our lecture IN2064)
 - We assume the basic concepts are clear; no repetition!
 - We strongly recommend that you attend IN2064 first before taking this class
- Knowledge about:
 - Algorithms and Data structures
 - Programming
 - Mathematics: Linear Algebra, Statistics, Optimization

Course Material + Announcements

- All course materials (slides, exercises) will be uploaded to Moodle
 - Video recordings of previous years lectures are accessible via:

https://www.cs.cit.tum.de/daml/teaching/mlgs/

- Project submission via Artemis
- Use Piazza to ask questions! (please avoid sending e-mails)

http://piazza.com/tum.de/summer2024/in2323

Access Code: mlgs2024

Please read the <u>guidelines</u> for using Piazza

Exercises and Project Tasks

- Exercise sheets
 - Exam preparation
 - Solutions in the tutorials
 - Due to the high number of registrations, we are unable to provide corrections to your solutions
- Project tasks
 - Get hands-on experience with advanced machine learning methods
 - Improve your final grade! (details later)

Project Format

- Format of programming tasks
 - Tasks will be published via Artemis
 - artemis.in.tum.de
- How to solve programming tasks?
 - Clone template repository from Artemis exercise
 - Solve tasks described in the repository
 - Push repository with filled-in solutions
- Bonus regulations
 - 10 points for each programming sheet
 - A Bonus of 0.1, 0.2 and 0.3 grade points will be granted upon correct completion of 25%, 50% and 75% of all project points, respectively

Project topics

- Six project tasks on the following topics
 - Hidden Markov Chain
 - Word2vec
 - Temporal Point Processes
 - Clustering
 - Graph neural networks
 - Robustness
- The specific tasks and all details will be described in the corresponding exercise sheets

Exam & Grading Scheme

- Written final exam: 90 minutes
 - Date will be announced via TUMonline
 - We currently plan with an on-site exam
 - One handwritten two-sided A4 sheet with notes

```
def final_grade(exam_grade, project_grade):
    if exam_grade > 4.0:
        return exam_grade
    else:
        return max(1.0, exam_grade - bonus)
```

→ The project is voluntary and can only improve the final grade. The project bonus applies only if you passed, and you cannot improve beyond 1.0

Our Group's Focus

Reliable Machine Learning for Non-Independent Data



- Data corruptions, adversaries
 - Certificates

Non-independent data

- Temporal/sequence data
 - Graph data

- Interested? We offer:
 - Bachelor/Master theses, Guided Research projects, HiWi positions
- More details on specific topics closer to the end of the semester

References

Figures taken from

- Goodfellow et al. 2014, https://arxiv.org/abs/1412.6572
- Akbik et al. 2018, https://research.zalando.com/welcome/mission/research-projects/flair-nlp/
- Khan 2019, https://heartbeat.fritz.ai/stylegans-use-machine-learning-to-generate-and-customize-realistic-images-c943388dc672
- Liao et al. 2019, https://arxiv.org/abs/1910.00760