

Natural Language Processing

Exercise – Summer term 2025

`tim.hagen@uni-kassel.de`

Organization

Communication

- ❑ Slides, announcements & materials on the course website [\[temir.org\]](https://temir.org)
- ❑ Questions and communication via Discord and email [\[discord.gg\]](https://discord.gg/temir)
(please use your full name as the server nickname)



Organization

Exercise

Exercise sheets every second week; programming exercises every other [\[temir.org\]](https://temir.org)

Exercise Sheets

- ❑ Mix of theory questions and toy examples
- ❑ The tasks are **highly relevant for the exam**
- ❑ Published Wednesday evening and due Monday evening
- ❑ Submit to me personally via Discord or email
- ❑ Required for the Studienleistung

Programming Exercises

- ❑ Published shortly before the exercise
- ❑ Jupyter Notebooks that let you work with the NLP hands on
- ❑ Don't need to be submitted

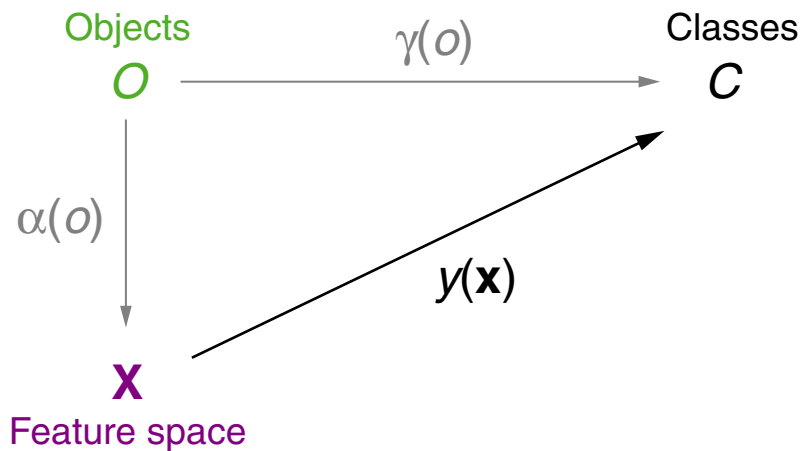
Getting to know you

Do you have prior experience with. . .

- ☐ Natural Language Processing
- ☐ Machine Learning
- ☐ Deep Learning
- ☐ Python
- ☐ NLTK
- ☐ spaCy
- ☐ PyTorch
- ☐ Hugging Face
- ☐ CNN, RNN, LSTM, Transformer

Classification

ML: Real World and Model World [model formation]



- What could be features for NLP?

Tokenization Schemes

- ❑ Character Level
- ❑ BPE, WordPiece, SentencePiece
- ❑ Morpheme Tokenization
- ❑ ...