Exam information

Organizational stuff

- You need to register for the exam
- Bring a passport to the exam to identify yourself
- This is a closed book exam. No electronic devices or other aides are allowed.
- Use an indelible pen
- Write your Student ID number and name on every page

Exam questions

- Exam structure:
  1. Multiple Choice questions
  2. Outline the pipeline/approach for a new molecule prediction task
  3. Explaining code
  4. Hands-on exercise on some of the techniques we have learned
  5. Describe the architecture and workflow of a model seen in the lecture

1. Multiple choice questions

- 4 block with each 4 True/False questions:
  - For each question you have to decide whether the statement is "`True"' or "`False”’.
  - You don't have to answer all questions.
  - For each block of 4 questions, the grading rule will be the following:
    - $Points = (Correct\ answers\ -\ wrong\ answers) \times 0.5.$
      - If there are more wrong than correct answers in a block, you will get 0 points.
  - Questions will ask knowledge from the slides/videos and exercises
    - Examples:
      - All proteins are responsible for catalyzing (speeding up) a chemical reaction
      - For k=2 there exist 40 different k-mers (or 2-mers) that can occur in protein sequences
      - In a gradient boosting decision tree model, the regularization coefficients lambda and alpha can lead to removing trees if the influence of these trees is too low
      - For the input of a transformer network: If we shuffle the order of the input tokens after we applied positional encoding to the input tokens, the Transformer Network cannot easily detect the correct original positions of the input tokens

2. Outline the pipeline/approach for a new molecule prediction task

- You will receive a description of a dataset and a prediction task that we have not discussed yet:
  - Example:
    - Prediction task: We want to predict if two proteins interact with each other (protein-protein interactions) by using the protein amino acid sequences of two proteins
    - A dataset with ~200k protein-protein pairs (with protein amino acid sequences) and with binary labels (interaction/no interaction)
- Your task:
  - Describe the data preprocessing
  - Describe a suitable model architecture that will likely result in high model performance
  - Describe the training process
  - …

3. Explaining code

- You will receive code of a model or of a technique you have seen in the lecture or worksheets
- You need to explain what the code is doing

4. Hands-on exercise on some of the techniques we have learned

- You will get some task, where you will actually need to execute/apply some methods/knowledge from the lecture
- Example:
  - Draw a molecular graph with at least 6 atoms with a chiral center
  - Draw a molecular graph with at least 6 atoms without a chiral center

5. Describe the architecture and workflow of a model seen in the lecture

- I will give you a name of a model from the lectures and you will need to provide a detailed explanation of the model
  - Input preprocessing
  - Model architecture
  - Training task
  - …