Problem-based learning (PBL)
SoSe20, WiSe20/21
Hi :) 

- I am Maria
- I am a PhD student in Bioinformatics since 2017 and responsible for this PBL course
- If you have any (organizational questions), feel free to contact me at any time (on Piazza or via email: maria.littmann@tum.de)
Organization

- Officially: PBL takes place at MI 01.09.034, Monday 12:30-14:00
- However, due to the current situation, there won’t be any in-person meetings at least for the next couple of weeks
- → This course will be held online
- We use Piazza for communication, please sign up here: piazza.com/tum.de/spring2020/pbl
  - All course material will also be uploaded there
  - I will hold live Q&A sessions there to answer questions
- Each group will have one rolling slide deck
  - First slides: Topic description + material from your supervisor
  - Please insert all your presentation slides there
General outline

● Language of the course: English
● Students will work in groups of 2-3 on one topic
● One group will work on one prediction task, give all talks together, and write one final scientific report
  ○ Overall task: Predict a given feature for your dataset using (simple) Machine Learning
  ○ Each student will work on certain sub-tasks individually and will be assigned an individual grade based on these
  ○ Discuss possible sub-tasks with your supervisor
● Results and intermediate steps will be presented in 4 talks and one written report
● Requirements: Use Python and Scikit-learn
Overview over topics

More details can be found in the corresponding slide decks
Prediction of transmembrane helices

- Supervisor: Michael Bernhofer
- Task: Prediction of transmembrane proteins and topology
- Group: tbd (3 students)
- Slide Deck: See Piazza
Prediction of binding affinity

- Supervisor: Christian Dallago
- Task: Predict the binding affinity (regression) between a protein and a small molecule.
- Group: tbd (3 students)
- Slide Deck: See Piazza
Prediction of protein disorder

- Supervisor: Michael Heinzinger
- Task: Predict whether a residue in a protein is disordered or not using one-hot encoding (maybe using also other types of encoding, tbd). After predicting binary disorder, one student will focus on “functional disorder” (disorder induces function), while the other student will focus on “binding disorder” (disorder induces binding, e.g. to other proteins).
- Group: tbd (2 students)
- Slide Deck: See Piazza

https://weisgroup.ku.edu/intrinsically-disordered-proteins

Prediction of ligand binding

- Supervisor: Maria Littmann
- Task: Predict whether a protein binds DNA/RNA (nuclear), peptides, metal or small ligands (where proteins can bind multiple types of ligands)
- Group: tbd (3 students)
- Slide Deck: See Piazza

Prediction of nuclear localization

- Supervisor: Maria Littmann
- Task: Prediction of whether proteins are solely localized to the nucleus or not and prediction of nuclear sub-compartments
- Group: tbd (2 students)
- Slide Deck: See Piazza

https://www.genome.gov/genetics-glossary/Nucleus
Prediction of signal peptides

- Supervisor: Konstantin Weißenow
- Task: Per-residue prediction of signal peptide types and location of residues in the mature protein
- Group: tbd (3 students)
- Slide Deck: See Piazza
Detailed instructions
## Timeline - SoSe20

<table>
<thead>
<tr>
<th>Date</th>
<th>What</th>
<th>Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>27.04.2020</td>
<td>Kickoff</td>
<td>Maria Littmann</td>
</tr>
<tr>
<td>11.05.2020</td>
<td>How to give a good presentation (Lecture talk)</td>
<td>Maria Littmann</td>
</tr>
<tr>
<td>08.06.2020</td>
<td><strong>Introduction talks</strong> (groups tba)</td>
<td>Tba</td>
</tr>
<tr>
<td>15.06.2020</td>
<td><strong>Introduction talks</strong> (groups tba)</td>
<td>Tba</td>
</tr>
<tr>
<td>22.06.2020</td>
<td><strong>Introduction talks</strong> (groups tba)</td>
<td>Tba</td>
</tr>
<tr>
<td>29.06.2020</td>
<td>Introduction to Machine Learning &amp; Python (Lecture talk)</td>
<td>Maria Littmann</td>
</tr>
<tr>
<td>13.07.2020</td>
<td><strong>First milestone talk</strong> (Presentation of dataset)</td>
<td>Everyone</td>
</tr>
</tbody>
</table>
## Timeline - WiSe20/21

<table>
<thead>
<tr>
<th>Date</th>
<th>What</th>
<th>Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.10.2020</td>
<td><strong>Second milestone talk</strong> (ML setup)</td>
<td>Everyone</td>
</tr>
<tr>
<td>26.10.2020</td>
<td>How to evaluate ML &amp; how to write a scientific report (Lecture talk)</td>
<td>Maria Littmann</td>
</tr>
<tr>
<td>30.11.2020</td>
<td><strong>Final talks</strong> (groups tba)</td>
<td>Tba</td>
</tr>
<tr>
<td>7.12.2020</td>
<td><strong>Final talks</strong> (groups tba)</td>
<td>Tba</td>
</tr>
<tr>
<td>14.12.2020</td>
<td><strong>Final talks</strong> (groups tba)</td>
<td>Tba</td>
</tr>
<tr>
<td>31.01.2021</td>
<td>Deadline to submit final report</td>
<td>Everyone</td>
</tr>
</tbody>
</table>
Grading

- Grade = Practical work (incl. talks presenting the results of this work) + Final Scientific Report + Participation in Discussion/Online Assignments
  - Practical Work includes talks to present results
  - Not all parts are weighted equally
- While students will work in groups, each student will receive an individual grade
- For each talk and for the scientific report, we will provide a checklist with
  - The most basic requirements every group should meet
  - Which parts will be individually graded
  - Which parts will be graded the same for the entire group
- Not meeting the requirements on that checklist will worsen your grade!
Attendance & Participation

- Officially: Attendance is mandatory
- As long as this course takes place online: Attendance is mandatory only for student talks
  - Instead of attendance of the lecture talks, it is mandatory to work on the assignments/quizzes for the lecture talks
- Participation in Discussion/Online Assignments:
  - Each student has to ask at least one question after each introduction and final talk
  - Active participation in the discussion after the talks (incl. milestone talks) will be part of the final grade
  - Work on online assignments and quizzes for the lecture talks will be part of the final grade
Detailed information on different talks & final report
Introduction Talk

● **Should provide an introduction into the overall topic**
● **Duration:** 30 minutes + 15 minutes discussion
● **Using the literature given by your supervisor as starting point:**
  ○ Think about interesting aspects that should be covered in this introduction (e.g. biological background, public resources, state-of-the-art methods,...)
  ○ Decide on a meaningful split of these aspects into sub-topics so that
    ■ Each student can research, prepare, and present one sub-topic
    ■ Sub-topics don’t overlap
  ○ Research further literature relevant for your sub-topic
● **Discuss with your supervisor:**
  ○ The chosen sub-topics
  ○ The final slides (at least 2 weeks before your talk)
First milestone talk

- **Presentation of your dataset**
- **Duration**: 10 minutes + 5 minutes discussion
- **At this point, each group should be able to present:**
  - Simple statistics about their dataset (size, distribution of labels you want to predict)
  - The input features they are going to use for the prediction
  - A summary slide of the next steps
- **Less formal talk**
Second milestone talk

- **Present first Machine Learning pipeline**
- **Duration**: 10 minutes + 5 minutes discussion
- **At this point, each group should be able to present:**
  - Their feature extraction and preparation
  - Performance for baseline prediction
  - Cross-validation preparation (how many splits, stratified,...)
  - Training & validation for one set of cross-validation splits with default parameters (no parameter tuning needed at this point!)
  - A summary slide of the next steps including the different approaches followed by each student individually
- **Less formal talk**
Final talk

- Presentation of final results
- Duration: 30 minutes + 15 minutes discussion
- Short summary of previous results
  - Most important aspects from introduction talk
  - Dataset and dataset preparation
- Present final results
  - Thorough performance evaluation (what are good performance measurements for this problem, how to interpret them,...)
  - Error estimates
  - Comparison to baseline prediction
  - Problems/limitations
  - Clearly highlight individual contributions of each student
Lecture talks

- Information on topics relevant for this course
- Slides and reading material will be uploaded beforehand
- Students should read slides and reading material on their own
- There will be assignments/quizzes online for students to check if they understood the material
- On the date of a lecture talk, there will be an online meeting where
  - I will summarize the most important points
  - Students can ask questions about the material
Scientific report

- Written summary of final results
- Covers general parts of a scientific paper (abstract, introduction, materials & methods, results & discussion, conclusion, references, figures, tables)
- In addition to the results presented in the final talks, you should also include the points raised during the discussion of the final talk
- In general, everyone in the group is responsible for submitting a good report
- But: Individual grades will be assigned to the parts covering the aspects each student worked on individually
- Length: tba
- Final report should be submitted as PDF, you can work using Word or Latex
- Format: Use Bioinformatics template provided on the website and Piazza
Summary
Summary

- These slides should provide a general overview over the plan for PBL20/21
- More specific information and requirements for each talk will be published before the given dates
- **Next up: Lecture talk on “How to give a good presentation”**
  - Reading material and slides will be published before
  - Please read the material and check the slides
  - There will be an online session on May 11, 12:00-14:00 where
    - I will summarize the most important aspects
    - You can ask questions
- Don’t hesitate to contact your supervisor to discuss details of your topic or if you have any questions!
- Don’t hesitate to contact me or post on Piazza if you have any organizational questions!