What’s next?
Problem based learning - July 2
Where you should stand now

- Have a basic understanding of the background of the task you are given to do
- Have a basic idea of how to use ProtVec
- Have a basic understanding of Machine Learning and how to apply it using sci-kit learn
What is next?

● There won’t be any more official meetings in this semester

● Next meetings: beginning of the next semester
  ○ probably: 22.10/29.10 12:00-14:00
  ○ we will let you know as soon as we know for sure

● What should you do until then?
  ○ get the dataset from your supervisor
  ○ familiarize yourself with the dataset and the labels
    ■ size of the dataset
    ■ distribution of labels in the dataset
    ■ any other important information?
  ○ prepare the data splits to perform k-fold stratified cross-validation
  ○ discuss your results with your supervisor
What is next?

● Present your results in the first 2 meetings next semester
  ○ same groups as before (since you share a data set)
  ○ max. 15 minutes per group
  ○ present your dataset
  ○ present your (first) “prediction task” in 2 slides
    ■ feature you want to predict?
    ■ how many states?
    ■ per-residue or per-protein prediction?
    ■ imbalanced data?
    ■ preparation of cross-validation splits

● We will publish the exact dates for all students on the course website soon