

Protein Prediction II

Winter 18/19

for Computer Scientists
for Bioinformaticians

Timeline

25.10.	Introduction
1.11.	No exercise
8.11.	Build a machine learning pipeline
15.11	Optimize Word2Vec parameters
22.11.	Discuss W2V optimizations
29.11.	Present 'final' W2V optimizations
6.12.	No exercise, Dies academicus
13.12.	Present current progress
20.12.	No exercise session

10.1.	Feedback about your progress (no presentation necessary)
17.1.	Q&A session
24.1.	Final presentations w/ Prof. Rost Deadline for method submission
31.1.	Q&A Session Method evaluation
7.2.	Exam

Performance (error estimation)

- Show error (bars) in tables and plots, not just average performance
 - Use bootstrapping
1. Randomly resample your test data (e.g. proteins)
 - With replacement (can sample the same datapoint multiple times)
 - Bootstrap sample has same size as original sample
 2. Calculate performance value(s) on the bootstrap sample
 3. Repeat steps 1) and 2) at least 1000 times
 4. Calculate standard deviation of those performance values
 - Use corrected sample standard deviation (i.e. $N - 1$)
 5. Report standard deviation as error of average performance

Tasks for 24.1.

- Prepare (good) presentations summarizing your work
 - Include even the 'obvious' things that are the same for every group so someone with no prior knowledge can understand what you did
 - Aim for 10 minutes + discussion
 - Not everyone in the team has to talk
- Submit your methods
 - Give us pre-trained models
 - Input and output formats are given in the previous slides
 - How you treat arguments is up to you, as long as it's clear to us (e.g. use argparse)
 - Submit at the latest on the 24th

How to give a good presentation

(this is not an example of one)

Slide layout I

- Names and date on the first slide
 - Slide numbers (readable!)
 - Outlines are almost always unnecessary
-
- Unless you are citing use bullet points, not sentences
 - If you can fit more, your font size is too small (aim for ~18pt)
 - More than 5 bullet points are probably too much (sic!)

Slide layout II

- Don't get too fancy with the layout, don't waste precious space on fancy headers etc.
- Don't use animations
- Don't jump back and forth between slides
- Don't jump over slides
- Don't spend too little time on a slide (especially if it's a plot)

Figures I

- Figures are usually better to grasp than a wall of text (again, sic!)
- The default output of whatever plotting system you use looks fine on your screen, probably not on a projector
- Increase the font size of axis labels, titles, legends...
- Probably increase line width
- Use colors that are easy to see even with little contrast from the projector
- (increase resolution)
- Don't waste whitespace on the slide and crop your plot tightly

Figures II

- Explain what's going on!
 - Remember the audience has never seen the plot before
 - Start with the intention and what's on the axes
 - Only then discuss individual data points or give an interpretation
- Make sure you actually know what's going on... (what is the underlying dataset? etc.)

Presenting

- Introduce yourselves (in addition to names on the first slide)
- Speak slowly and clearly
- Talk to the audience
- You shouldn't need notes
- Match the given timeframe
- Practice

References

- At least a list of full references at the end
- If you have many references, better have them on the slides they apply to
 - E.g. in a footer
 - A shorter format such as Author, Journal, Year is fine