



Assignment 6 (Bonus Sheet)

Deadline: Thursday, February 03, 7:59 p.m.

This problem set is worth 25 points. You can submit in groups of two people. Submit your solutions digitally by uploading to the [ILIAS page](#) (none of the other students can see the files you upload). Just upload a zipped folder containing all necessary files and name the folder by your last names. The folder should be named according to the following scheme:

[MDS][Assignment6]_lastname1_lastname2

Problem 1 (T, 15 Points)

Deep Learning and Viruses

- (a) (2P) What is represented in a 2-D contact map (in the context of proteins)?
- (b) (2P) How can amino acid sequences be encoded?
- (c) (2P) Explain how the "*L/5 top scored predicted contact pairs*" are used as a performance measure.
- (d) (2P) How can dependence between two random variables be measured?
- (e) (2P) Why is max-pooling used in Neural Networks? Is it a procedure with or without information loss?
- (f) (2P) What are the infection routes of viruses?
- (g) (3P) The immune system can be divided into innate and adaptive immune system. Describe both of them briefly with their individual components, i.e. explain which cells are involved in which way.

Problem 2 (P, 10 Points)

Download the protein sequence of the S protein (spike protein) of the omicron variant from genbank: https://www.ncbi.nlm.nih.gov/protein/7QO7_C and predict the 3D structure of the protein with RaptorX: <http://raptorx.uchicago.edu/ContactMap/>. The server only accepts 1000 amino acids or less if one wants to have a 3D structure, so you can leave out any amino acids in the end after the 1000th amino acid. Describe what you see in comparison to the experimental structure from PDB <https://www.rcsb.org/structure/7QO7>.