Deep Learning, Homework 2

Out: Feb 26, 2024, Due: March 08, 2024, Total: 50

Note:

- This homework will carry 5 points towards your final score
- Homeworks are individual work, please do not collaborate with others inside or outside of the class.

Task

The data provided for this homework is DNA string of a collection of samples. Each sample belongs to a class of species. Your task is to predict the class label using DNA string as feature.

DNA string is a sequence of characters from the set: $\{A, C, G, T\}$. In your data some position is given as _, which means undetected, but for your purpose you can think that as the fifth character. One simple way to represent the data numerically is to use one-hot-encoding (one out of five) for each position.

In this assignment, you will be using a 1-D CNN model with two convolution layers for solving this classification task. You need to first convert the input data as a 5-channel input. In the output use a softmax layer along with cross-entropy loss for multi-class classification.

Deliverable: Please submit source code (python file or a python notebook file) and your prediction on the test dataset.