

COMP9414 23T2

Artificial Intelligence

Assignment 2 - Multi-layer neural networks

Due: Week 10, Monday, 31 July 2023, 11:55 PM

1 Problem description

We want to model the evolution of biomass in a bioprocess, which is periodically fed with a compound A, resulting in an increase in the mass of living matter (biomass) and a decrease - as it grows - in the existing substrate (food). A schematic of the bioprocess can be seen in Figure 1.

When the bioprocess receives compound A, it increases the mass of living matter and decreases the existing substrate as it grows. The dynamics of the variables can be observed in Figure 2.

The problem, then, consists of using a multi-layer neural network to approximate the bioprocess, where compound A and substrate are considered as inputs to the network and biomass as its output. To achieve this, the basic rules of data preprocessing, neural network design, training, and validation must be followed.

Three data files representing the inputs and the target are available, as shown in Table 1.

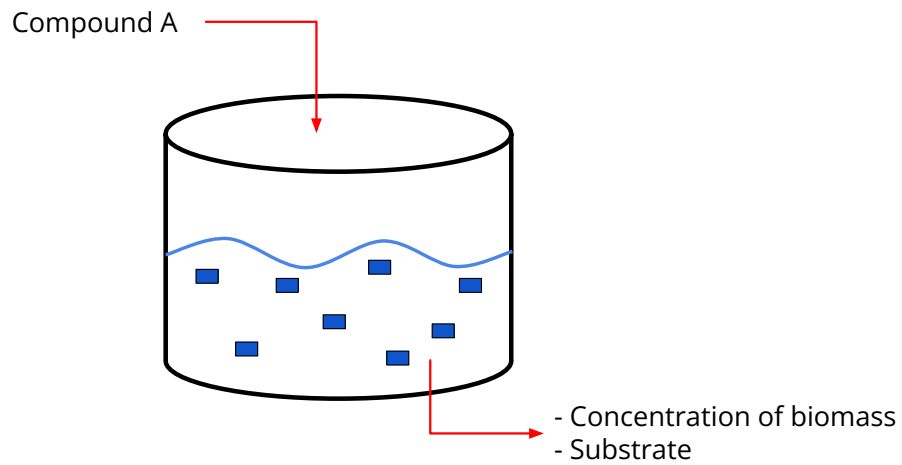


Figure 1: Bioprocess fed with compound A, which affects the concentration of biomass and the existing substrate.

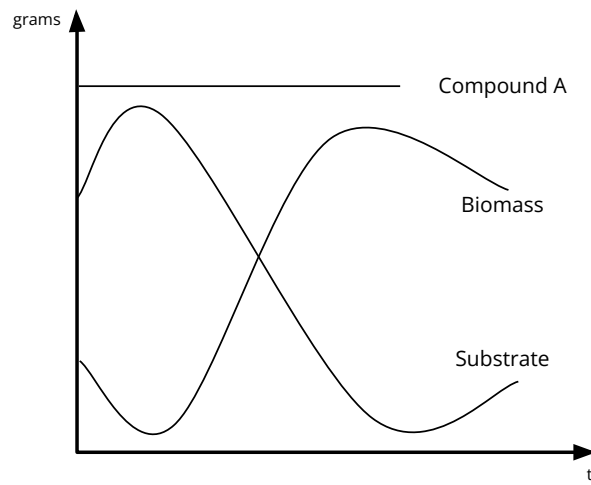


Figure 2: Dynamics of the process among the variables: compound A, substrate, and biomass.

Table 1: Files provided with the data.

File	Bioprocess's meaning	Neural model's meaning
compoundA.txt	Compound A	First input
substrate.txt	Substrate	Second input
biomass.txt	Biomass	Output

2 Activities

2.1 Data preprocessing

In this section, it is expected you analyse the provided data and perform any required preprocessing. Some of the tasks during preprocessing might include the ones shown below; however, not all of them are necessary and you should evaluate each of them against the results obtained.

- (a) Identify variation range for input and output variables.
- (b) Plot each variable to observe the overall behaviour of the bioprocess.
- (c) In case outliers are detected correct the data accordingly. For instance, as we are dealing with variables measured in grams, no value should be less than zero. A simple correction is to replace such values with a zero value.
- (d) Split the data into two subsets: training and test.

2.2 Design of the neural network

In this section, it is expected you select and design a neural architecture for biomass estimation. Consider the following steps.

- (a) Design the multi-layer neural network. Given the number of training samples, propose a neural architecture taking into account N_i – number of inputs, N_h – number of units in the hidden layer, and N_o – number of outputs.
- (b) Decide the number of layers and their respective activation functions.

- (c) Remember it's recommended your network accomplish the maximal number of parameters $Nw < (\text{number of samples})/10$.
- (d) Create the neural network using Keras and TensorFlow.

2.3 Training

In this section, you have to train your proposed neural network. Consider the following steps.

- (a) Decide the training parameters such as loss function, optimizer, batch size, learning rate, and episodes.
- (b) Train the neural model and verify the loss values during the process.
- (c) Verify possible overfitting problems.

2.4 Validating the neural model

- (a) Assess your results plotting the network response for the test inputs against the test targets. For instance, Figure 3 shows the results for 400 points.
- (b) Compute error indexes to complement the visual analysis. Use IA, RMS, and RSD.

$$IA = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (|o'_i| + |p'_i|)^2}$$

$$RMS = \sqrt{\frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n o_i^2}}$$

$$RSD = \sqrt{\frac{\sum_{i=1}^n (o_i - p_i)^2}{N}}$$

with:

- o_i : observed values.
- p_i : predicted values.

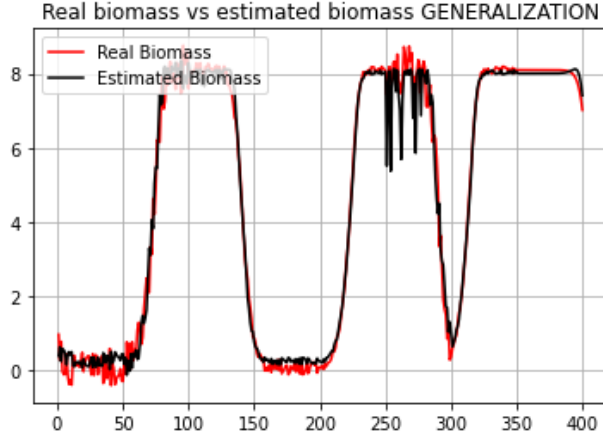


Figure 3: Real biomass and estimated biomass with a neural network.

- N : number of data samples.
- o_m : observations' mean value.
- $o'_i = o_i - o_m$
- $p'_i = p_i - o_m$

3 Testing and discussing your code

As in the previous assignment, your code will be tested by tutors along with you in a discussion session carried out in the last tutorial session in week 10.

You are expected to propose and build a multi-layer neural network to estimate biomass in a bioprocess. The minimal output we expect to see is a plot showing the real and estimated biomass for your test set (as shown in Fig. 3) and a plot with the loss of the training and test sets. You will receive **2 marks for each of these plots**. Additionally, you should compute the error indexes for the test set. However, it's fine if you want to include any other outcome to highlight particular aspects when testing and discussing your code with your tutor.

For marking your results, you should be prepared for simulating your neural model with a generalisation set we have saved apart for that purpose. Therefore, during the discussion, you should be ready to make some modifications to your code in order to run the test. These modifications to the

code might include loading new files, preparing the data as input for your proposed neural network, plotting the simulation results against targets, and computing the error indexes for the data provided.

As you need to compute IA, RMS, and RSD as output, we will test these indexes with the generalisation set. You will receive **2 marks for each index** as follows: $IA > 0.9$, $RMS < 0.1$, and $RSD < 0.3$.

4 Submitting your assignment

You can do the assignment either individually or working in a couple with a classmate. If you decide to work in a couple with a classmate in another tutorial section, you need the approval of one of these two tutors, who will conduct the discussion with you and your classmate. However, the other tutor still needs to be informed by the student.

Your submission should be done by Moodle and consist of only one .py file. If you work in a couple, only one person is required to submit the file. However, the file should indicate on top as a comment the **full name and zID** of the students. It is your responsibility to indicate the names, we will not add people to a work after the deadline if you forget to include the names.

You can submit as many times as you like before the deadline – later submissions overwrite earlier ones. After submitting your file a good practice is to take a screenshot of it for future reference.

Late submission penalty: UNSW has a standard late submission penalty of 5% per day from your mark, capped at five days from the assessment deadline, after that students cannot submit the assignment.

5 Deadline and questions

Deadline: Week 10, Monday 31st of July 2023, 11:55pm. Please use the forum on Moodle to ask questions related to the project. However, you should not share your code to avoid making it public and possible plagiarism.

Although we try to answer questions as quickly as possible, we might take up to 1 or 2 business days to reply, therefore, last-moment questions might not be answered timely.

6 Plagiarism policy

Your program must be entirely your own work. Plagiarism detection software might be used to compare submissions pairwise (including submissions for any similar projects from previous years) and serious penalties will be applied, particularly in the case of repeat offences.

Do not copy from others. Do not allow anyone to see your code. Please refer to the UNSW Policy on Academic Honesty and Plagiarism if you require further clarification on this matter.