



UNIVERSITY OF
ABERDEEN

University of Aberdeen
South China Normal University
Aberdeen Institute of Data Science
& Artificial Intelligence.

BSc in Artificial Intelligence 2023 – 2024

*****Please read all the information below carefully*****

Assessment I Briefing Document – Individually Assessed (no teamwork)

Course: JC3509 – Machine Learning

Note: This part of assessment accounts for 30% of your total mark of the course.

Learning Outcomes

On successful completion of this component a student will have demonstrated competence in the following areas:

- Ability to identify, prepare, & manage appropriate datasets for analysis.
- Ability to appropriately present the results of data analysis.
- Ability to analyse the results of data analyses, and to evaluate the performance of analytic techniques in context.
- Knowledge and understanding of analytic techniques, and ability to appropriately apply them in context, making correct judgements about how this needs to be done.

Information for Plagiarism and Conduct: Your submitted report may be submitted for plagiarism check (e.g., Turnitin). Please refer to the slides available at MyAberdeen for more information about avoiding plagiarism before you start working on the assessment. Please also read the following information provided by the university: <https://www.abdn.ac.uk/sls/online-resources/avoiding-plagiarism/>

In addition, please familiarise yourselves with the following document “code of practice on student discipline (Academic)”: <https://tinyurl.com/y92xgkq6>

Report Guidance & Requirements

Your report must conform to the below structure and include the required content as outlined in each section. Each subtask has its own marks allocated. You must supply a written report, along with the corresponding code, containing all distinct sections/subtasks that provide a full critical and reflective account of the processes undertaken.

Overview

This assignment tasks you to undertake the full machine learning pipeline, including data handling and processing, model construction and training, and evaluation of the developed methods. You are tasked to create a neural network to classify data into 3 categories.

****Please read all the information below carefully****

The dataset needed to fulfil the requirements of this assessment can be found in MyAberdeen.

Data:

This data contains the chemical properties of food product produced by 3 different manufacturers. The purpose of this experiment is to explore the relationship between the chemical measures listed below and the manufacturer of the food product. The data has 177 records, where the first column “Producer” indicates which manufacturer produced the analyses sample. The features of the dataset are the following:

- Producer – Manufacturer of the product (TARGET).
- Amino_acid – The total percentage content of amino acid.
- Malic_acid – The percentage content of malic acid.
- Ash – The measure of ash present in the product.
- Alc – The alkalinity of ash present.
- Mg – The measure of magnesium.
- Phenols – The total measure of phenols.
- Flavanoids – The measure of flavonoid phenols in the product.
- Nonflavanoid_phenols – The measure of non-flavonoid phenols in the product.
- Proanth – Proanthocyanins measure.
- Colo_int – The color intensity.
- Hue – Hue of the color.
- OD – The protein content of the product.
- Proline – The measure of proline amino acids.

Objectives:

The main purpose of employing this data is the following:

1. Classification: to determine the origin (manufacturer) of the product given analytical measurements.
2. Analysis: to infer which analytical factors would potentially influence the classification of the product.

In order to achieve these objectives, we would like to accomplish the following subtasks using machine learning.

Submission

Please provide the follow:

1. A written report explaining the steps undertaken for each task, and the decisions behind each choice. You are expected to use machine learning principles to explain your results with graphs and/or tables.
2. A code submission, comprising of ONE commented python file with all code needed to replicate the findings in the written report.

****Please read all the information below carefully****

Task 1 – Data Preparation (10 Marks)

Subtasks:

1. Import the dataset: Please provide a short description of the data provided and import the data into your programming environment; provide snippets of code for these purposes.
2. Preprocess the data: If you did any preprocessing over the data, e.g., normalization, please explain it and the reasons why you did that preprocessing; if you did not do any preprocessing, also please explain.

Task 2 – Model Construction (50 Marks)

You are tasked to build simple fully connected artificial neural network from scratch to classify the records into 3 categories (1, 2, or 3).

You are not permitted to use any machine learning or statistical libraries, you are expected to construct the neural network from scratch, i.e. only using core Python and NumPy.

Subtasks:

1. Loss function: Select and implement an appropriate loss function, explain why you have selected that loss function in relation to the data and the problem.
2. Network Design: Construct a fully connected neural network with **at least one** hidden layer. Explain your architectural choice and demonstrate by code snippets, tests, and written explanation that your code operates as expected. To achieve this, you will need to implement both:
 - a. The Forward Pass.
 - b. The Backward Pass.
3. Gradient Descent: Update the weights by mini-batch stochastic gradient descent. Demonstrate by code snippets, tests, and written explanation that the weights are being updated. You can use advanced optimisation tricks if you wish i.e. momentum.

NOTE: If you are unable to complete the above tasks, you are permitted to use additional libraries (i.e. PyTorch) however, this will result in a deduction of 20 marks.

Task 3 – Model Training (15 Marks)

Take the model from the previous task and train it on the data you pre-processed in Task 1. Ensure that you train your model on a sub-set of the data, holding out a subset for validation.

Subtasks:

1. Model Training: Perform training and parameter selection on the training set.
2. Module Regularisation: Implement a regularisation method, briefly explain (Max 200 words) how it works in the context of your code, use code snippets to help.
3. Model inference: Validate the model by performing inference on the held-out validation data.

****Please read all the information below carefully****

Task 4 – Evaluation (25 Marks)

Evaluate the performance of your trained classifier and employ machine learning principles to explain your results with graphs and/or tables. In addition, perform some analyses on the trained model to better understand which analytical factors would potentially influence the classification of the product.

Subtasks:

1. **Present Results:** Present the results of your classifier via appropriate metrics for the problem statement.
2. **Plot:** Plot the loss curve for training and validation, answer the following questions:
 - a. What does your loss curve tell you?
 - b. Are you observing any overfitting or underfitting?
 - c. Does the addition of regularisation help?
3. **Explain Results:** Explain the results from the previous subtasks in context of the problem statement/setting.

Marking Criteria

- Depth and breadth of knowledge.
- Technical details of formalisation, implementation and pseudo-code.
- Communication skills (clear, technical contents and sound reasoning)
- Structure of document.

Submission Instructions

You should submit a PDF version of your report and the accompanying code to the Codio environment. For the deadline of this assessment, please check it on MyAberdeen. The name of the PDF file should have the form “JC3509_Assessment1_< your Surname>_<your first name>_<Your Student ID>”. For instance, “JC3509_Assessment1_Smith_John_4568985.pdf”, where 4568985 is your student ID.

Any questions pertaining to any aspects of this assessment, please address them to the course coordinator Aiden Durrant, aiden.durrant@abdn.ac.uk