Coursework Brief

Task To Be Done

You are given the ML Coursework Dataset (access it via the link) to build a multilayer perceptron model that can be used for forecasting the export value of crop products for a geographical region three years into the future (e.g. given historical data up until or at Year 2984, your model can predict the export value of crop products for Year 2987). Your model is allowed to be either a regression model or a classification model. For more information about the dataset, see the README file included in the dataset folder.

What To Submit

You are to submit THREE separate documents clarified in items A-C below. Do NOT submit them as zipped (or other compressed) files. Make sure the file name and content for each of the 3 files indicate your Candidate No. (and not your name).

A. (90 marks) **Report** - Prepare a report that answers ALL 4 questions below. Use the EXACT question themes highlighted in bold as section titles for your report, so that your report has 4 (titled) sections exactly. Submit the document as a single pdf file.

1. (15 marks) **Performance**
    - (8 marks max) Report the performance of your model, and also describe (using maths formulae if appropriate) how the reported performance was computed from the model outputs.
    - (7 marks max) State clearly the total number of instances used and then the number of instances in your training and test sets and how these sets were derived from the given data.

2. (20 marks) **MLP model**
    - (11 marks max) Describe your multilayer perceptron (MLP) model, making sure to specify (including maths formulae when appropriate) at the very least the activation function for its output layer, the loss function used to train it, and the number of units in its output layer.
    - (9 marks max) List and describe the steps that you took to prevent overfitting for your model (using maths formulae when appropriate).

3. (35 marks) **Features & Labels**

- (10 marks max) Describe (as a list) how the label(s) was/were derived from the given data (using maths formulae when appropriate).
- (25 marks max) Report the features that you used for your model, making sure to specify the features used and the total number used, clarify why these features were selected, and describe (as a list) how they were derived from the given data, if applicable.

4. (20 marks) **Preprocessing**
    - (20 marks max) List and describe the preprocessing that you did on your features for building your model (using maths formulae when appropriate).

B. (5 marks) **Code** - Prepare all the code used to complete all the tasks above and submit as a single pdf file.

C. (5 marks) **Model outputs for the test set(s)** - Prepare the prediction outputs for your model in a single csv file and using an appropriate header row. The file should have at least 3 columns: 1 column for the data instance id, 1 column for the corresponding true label, at least 1 column for the corresponding prediction(s) for your model. Submit the document as a single csv file.

Other Important Notice

1. You are not required to code machine learning algorithms from scratch. You can use standard machine learning libraries including Scikit-learn, PyTorch, TensorFlow. However, you must only use standard libraries.

2. You are also allowed to use any standard software libraries (e.g. Python pandas/numpy, Microsoft Excel, IBM SPSS, Stata, R libraries) for data extraction and preparation.

3. You are NOT permitted to use or submit someone else's code, output, or report as yours (you are permitted to use code snippets from the lab materials given by the teacher, from machine learning software library documentations as stated in #1 above, or from recommended textbooks).

4. You must NOT use generative artificial intelligence (AI) to generate any materials or content for your assessment submissions.
    Note that:
    - The baseline position at the University is Sussex is that the use of generative AI material in assessment submissions is prohibited, unless explicitly permitted by the module convenor.
    - Students registered with the Disability Advice team and in receipt of reasonable adjustments are still permitted to use other assistive technology as required.
    - If in any doubt about what is permissible, students should check with the module convenor.

Link to learning outcomes

This assessment is designed to evaluate how well you have achieved the learning outcomes of the module:

- To know the fundamentals of machine learning and understand the theory (maths) behind standard algorithms;
- To be able to build, optimize, and evaluate machine learning models appropriately with data;
- To be aware of ethical issues relevant to machine learning;

by how you are able to systematically and creatively apply these knowledge and experience to a new problem.