EE4104 – Machine Learning and Artificial Intelligence

Assignment 3

Prediction of the Compressive Strength of Concrete using MLP Regression

Submission Deadline: 11:59PM, Wednesday, November 20th

Introduction

The UCI Machine Learning Repository contains the Concrete Compressive Strength dataset

https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength

which is suitable for use in an investigation into the capabilities of Multi-Layer Perceptron regression. The dataset contains over 1000 examples of 8 measurements related to the process of manufacturing batches of concrete with a 9th parameter being the measured compressive strength of the batch of concrete, Further insights into this dataset and the study completed by the original submitters of this dataset can be found in their paper from 1998:

https://www.sciencedirect.com/science/article/pii/S0008884698001653

The goal of this assignment is to complete a thorough investigation into the use of contemporary MLP regression to solve the problem of predicting the compressive strength of the concrete from the 8 input parameters. The study reported in the original 1998 paper was extremely limited due to the state of the art at that time. Your study should, amongst other things:

- Establish a performance figure using a non-neural network regression algorithm (trained and tested in a comparable manner to your MLP-regression model)
- Compare the performance of a baseline MLP-regression algorithm to that obtained above
- Complete an investigation to optimise the performance of the MLP-regression algorithm in terms of the hyper-parameters of the MLP-regression model

Some questions which are left up to you to answers in completing this assignment include:

- Is the dataset complete or are these examples which contain missing features?
- Should feature normalisation be completed?
- What would be an appropriate percentage of examples to be reserved for the test dataset?
- What are the best metrics to use when reporting your results for this study
- What hyper-parameters relating to MLP architecture and training should be investigated?

Additional comments

- It is recommended, if you have not previously done so in the previous assignment, that you install the UCI ML repository package using the steps below:
  - Open the Command Prompt:
    - type "cmd" in the Windows search box and select the "Command Prompt"

- Navigate to your previously created directory for example
    - cd c:\Users\Administrator\my_ml_project
- ***Install the repository (you only have to do this additional step once!!!)***
    - *pip install ucimlrepo*

- It is recommended that you also install and use the **tensorflow** package for implementations\investigations using MLP-regression
    - Complete **pip install tensorflow** in a similar manner to how the ucimlrepo package is installed
    - Make it clear in any prompts to ChatGTP that you want to use tensorflow (and keras which is now included as part of tensorflow) in the generated code
- The "Import in Python" button in the top right of the UCI ML dataset page provides code which will import this dataset in correctly formatted feature (X) and output value (y) arrays which can then be directly passed to the standard python **train_test_split** function used to generate the training and test datasets.

    from ucimlrepo import fetch_ucirepo

    # fetch dataset
    concrete_compressive_strength = fetch_ucirepo(id=165)

    # data (as pandas dataframes)
    X = concrete_compressive_strength.data.features
    y = concrete_compressive_strength.data.targets

Report Submission

On completion of this INDIVIDUAL assignment, you are required to submit a **report in PDF format ONLY which is no shorter than 2 pages and no longer than 4 pages** and which MUST use the following (usual) report template:

https://www.ieee.org/conferences/publishing/templates.html

In addition to submitting the PDF file, a submission of the Jupyter Notebook (file) (or file(s) containing your scripts if you are not using a Jupyter Notebook) containing your work must also be completed.

This report should include:

- Paper title (come up with your own catchy name for your paper!!) and your name
- Abstract which provides a short (200 word maximum) summary of your complete paper (i.e. what did you do and what were the "headline" results)
- Introduction Section where you briefly outline the purpose of your study and inform the reader what is in the other sections of your report
- Background Section which should introduce the dataset and the classifier methods that you used.
- Methodology Section where you detail how you completed the relevant investigation(s) undertaken. This should detail what steps you completed for each type of regression algorithm and how investigations into optimising the hyper-parameters was completed.

- Results Section where, using appropriate Tables and Figures, you detail the results obtained\observations made
- Conclusions Section where your detail your conclusions and provide some commentary on what you observed from the results
- A References list should be included if appropriate

Please note that Times New Roman font should be used throughout the report and a font size no smaller than 10 can be used.

Penalties will be applied as follows:

- Report length problem – 20% penalty per additional page or when under minimum page length
- Failure to use the required format/template – 20% penalty
- Failure to submit a PDF file – 20% penalty
- Failure to adhere to font related requirements – 20% penalty
- Failure to submit Jupyter Notebook file or script files if not using Jupyter notebook – 100% penalty
- Late report submission – 10% penalty per day or part of day late

All submissions will be reviewed by standard Plagiarism and Generative AI detectors and any submissions flagged by these detectors will be reviewed in-detail by me and may result in the authors being interviewed to determine whether the relevant College code of practice has been violated.