NENGA M25352

Coursework

Neural Networks and Genetic Algorithms

Submission: **Friday 12th May 2023** (**by 11pm** to Moodle)

This is an assessed piece of coursework (worth 40% of your final mark). It is, therefore, essential to be completed and submitted on time. If you are unclear about any aspect of the assignment, including the assessment criteria, please raise this at first opportunity. The usual regulations apply to a late submission of work. The submitted work must use Matlab Toolboxes. The submitted coursework should be your own (group) work. If it includes other people ideas and material, they must be properly referenced or acknowledged. Failing to do so intentionally or unintentionally constitutes plagiarism. The University treats plagiarism as a serious offence.

Your Task

Includes two parts: Applying Neural Networks (NN) for classification/mapping task and Genetic Algorithms (GA) for solving a simple global optimization problem.

The first part of the task is to analyse, design, train, validate and test artificial NN (using Matlab Neural Network Toolboxes), that can solve classification/mapping problem for given sets of data that you are provided with (from the UCI Repository of Machine Learning Databases: **https://archive-beta.ics.uci.edu/**).

You have to choose one of the two problems for which data sets are provided: *Servo* or *Iris* (see **Appendix A**), you can find them in Moodle. *Servo* and *Iris* are popular simple benchmark problems with continuous and discrete outputs respectively, used for testing and evaluation of shallow NN architectures and learning methods.

The purpose of this part of the assignment is to learn and demonstrate that you can: analyse, pre-process, and visualise the available data; design neural network topology and architecture; use training algorithm (Backpropagation); validate and test your NN, and post-process the results. You must use Matlab NN toolbox to design and train neural networks that can classify, predict, or map input patterns with expected outputs and subsequently test the NN ability to generalise.

The second part of the coursework is about using GA for solving Global optimization problem. You will have to use Matlab Toolboxes (GA) to find the global minimum for one of the two provided families of multimodal, multidimensional mathematical functions (see **Appendix B**). You should choose one function (with at least two variables), appropriate representation of the GA initial population, fitness function, selection method, reproductive operators, stopping criteria, etc. You will also have to set up adequate parameters (number of chromosomes in the initial population, stochastic parameters (e.g., probability of mutation), bounds and constraints for some of the parameters, etc.).

You will have to use Matlab Optimization Toolbox (GA) to visualize the evolution of the chosen population, the best, the mean and the standard deviation values of the fitness function at each generation and to analyse and discuss the results.

You can also implement different selection strategies (e.g., fitness-proportional, rank-based, tournament selection, etc.), and compare and critically discuss the results. Also, some discussion should be given on premature convergence, how can it be avoided and how the selection pressure can be kept relatively constant throughout the whole run. In your search strategy pay attention to the characteristics/parameters that can improve the balance between the exploration and exploitation abilities of your global search technique.

Deliverable

By **11pm, Friday 12.V.2023** you should submit (to Moodle) a .pdf file with your report (**one file per group only**, e.g., **Gr1-3.pdf**, or **Gr2-5.pdf**, etc.). The main body of the report should be (excluding your Appendixes) **maximum of six pages** (up to **2000 words**). It should be well structured and soundly written, presenting critical and comparative analysis of the adopted approaches and interpretation of the results. Typical content (but not limited to):

1. Abstract;
2. Introduction;
3. NN part:
   - Data Set - Analysis and Pre-Processing;
   - NN analysis and design;
   - Learning/Training analysis and results;
   - Test/Post-Processing/Generalisation results – comparison and analysis;
   - Conclusion and recommendations.
4. GA part:
   - Brief analysis of the chosen function to optimise;
   - Analysis and discussion of the adopted selection strategy, fitness function, and reproductive operators;
   - Analysis and brief discussion of the chosen parameters, stopping criteria, and convergence of the search;
   - Comparative and critical analysis of the results;
   - Conclusion and recommendations.
5. Reference list;
6. Appendices (any results, screen shots, diagrams and figures, and your Matlab code related to the coursework).

Assessing your assignment, I shall be looking for the following:

- Some thought and analysis of what is the complexity of designing, training, testing, validating and using a neural network for solving the problem;
- You have thought about coding the output and appropriate NN topology and architecture, adoption of adequate training technique (appropriate parameters, activation functions, etc.);
- You have thought about what data to use and for what purpose (subsets for training, testing, and validation), and how the subsets are organized (proportions of the dataset, internal distribution, etc.);
- You have thought about what kind of selection strategy to adopt for the GA optimization and what kind of operators to use for the reproduction (e.g., standard crossover and mutation, choice of parameters, e.g., probability of mutation, population number, population gap, etc.);
- Premature convergence and the ways of avoiding it are considered, choice of stopping conditions and strategy to keep the balance between exploration and exploitation of your global search, and also choice of parameters that could keep the selection pressure relatively constant throughout the whole run;
- Finally, you should demonstrate ability to critically analyse, discuss, and interpret the implemented approaches and obtained results, and to give some recommendations for further improvements.

Assessment Criteria

A - An excellent, well-written and structured report that makes an interesting read. NN part - you have the data set explored and analysed in a number of ways (e.g., pre-processing, normalisation, standardisation; how are the different subsets split; will a swap of the training/validation and testing subsets produce different results); appropriate coding, NN topology and architecture are chosen; the training, validation and testing are well analysed and discussed, and you have neural networks that perform well and can generalise.
GA part – you have analysed the function in hand and have adopted suitable selection strategy and reproductive operators. The choice of particular parameters and conditions is critically analysed and discussed.
You have analysed the complexity of the problems and the performance of the applied methods (different approaches may have been applied and compared), and have critically analysed, interpreted, and presented (and visualised) the results in an interesting and sound way.

B - A well-written report that is well structured into an interesting read. You have neural networks that perform well on the data set and you have analysed their performance. Your GA performance is critically analysed and discussed. You have exhibited some initiative in the approach taken and the results are clearly presented and critically discussed and compared.

C - A reasonable report that presents an account of the adopted approaches with critical analysis and discussion at places. The results are presented reasonably clear.

D - A report that presents the results mostly in descriptive way, but show reasonable understanding of the applied approaches.

F - Either no report submitted, or a report that presents very few results and little or no understanding of how these methods work.

Mark Scheme

Analysis and Pre-Processing (normalisation, standardisation) of the datasets and the multimodal function to be optimized.	**10**
NN and GA assumptions, Analysis and Design.	**20**
NN Training, Testing, Post-Processing, Generalisation, and Results. Adopted training, validation and testing strategy analysed and explained – e.g., why 'split-sample' or 'cross-validation' adopted?	**20**
GA analysis of the Run, Adopted Selection strategy, Reproductive operators and Results.	**20**
Report – structure, style, and presentation. Critical and comparative analysis and discussion demonstrating in-depth understanding and grasp of the investigated methods and approaches. Good visualisation and sound discussion and interpretation of the results.	**30**